buku Bu Alif

by Lppm unugiri

METODE STATISTIKA

Alif Yuanita Kartini, M.Si
Denny Nurdiansyah, M.Si
Nita Cahyani, M.Stat

METODE STATISTIKA

ISBN:

Penulis:
Alif Yuanita Kartini, M.Si
Denny Nurdiansyah, M.Si
Nita Cahyani, M.Stat

Penelaah:
<< Nama Penelaah >>

Editor:
<< Nama Editor >>

Lay out
<< Nama Layouter >>

Desain sampul:
<< Nama Desainer >>

KATA PENGANTAR

PRAKATA

Puji Syukur kami panjatkan kehadirat Allah SWT atas limpahan rahmat dan karuniaNya sehingga kami dapat menyelesaikan penulisan buku ajar ini. Tak lupa sholawat serta salam selalu tercurahkan kepada Baginda Nabi Muhammad SAW yang kita nantikan syafaatnya di yaumul kiamah. Ucapan terima kasih kami sampaikan kepada semua pihak yang telah membantu dalam penulisan buku ajar ini yang tidak dapat disebutkan satu per satu.

Buku ajar yang kami tulis dengan judul Metode Statistika telah kami susun dengan baik dan semaksimal mungkin dengan harapan agar menjadi manfaat bagi pembaca yang membutuhkan informasi dan pengetahuan tentang statistika. Dalam buku ini tertulis bagaimana pentingnya statistika dan juga bagaimana pokok bahasan yang disajikan relevan dengan mata kuliah dasar statistika yang menjadi alternatif pegangan bagi mahasiswa dari Program Studi Statistika maupun mahasiswa dari luar Program Studi Statistika.

Kami menyadari, masih banyak kekurangan dalam penyusunan buku ajar ini. Oleh karena itu, kritik dan juga saran sangat kami butuhkan demi perbaikan untuk meningkatkan kualitas buku ajar ini. Demikian buku ajar ini kami buat, dengan harapan agar pembaca dapat memahami informasi dan mendapat wawasan tentang statistika.

DAFTAR ISI

BAB 1

Konsep Dasar Statistika

A. Tujuan Pembelajaran

Kemampuan akhir yang ingin dicapai setelah mempelajari bab ini yaitu mampu menerapkan dasar-dasar statistika:

1. Mendeskripsikan statistik dan statistika
2. Menggolongkan statistik
3. Mendeskripsikan data dan jenis-jenis data
4. Menjelaskan koefisien dan variabel
5. Mendeskripsikan skala data
6. Mendeskripsikan metode pengumpulan data
7. Mengidentifikasi populasi dan sampel
8. Mendeskripsikan bias sampel
9. Mendeskripsikan kesalahan sampel

B. Pendahuluan

Pengambilan Keputusan merupakan hal yang paling penting dalam sebuah penelitian. Pengambilan Keputusan didasarkan pada informasi yang diterima dan disesuaikan dengan aturan yang berlaku. Ketika dihadapkan pada Keputusan yang lebih rumit, informasi dan norma saja masih kurang memadai, sehingga diperlukan alat bantu yang lain. Disinilah ilmu statistika dapat digunakan untuk menentukan Keputusan yang paling baik diantara banyak pilihan Keputusan berdasarkan data-data statistik.

Dalam bidang ekonomi, penyajian data-data statistik sangat diperlukan oleh para pelaku ekonomi. Olehkarena itu di negara-negara di seluruh dunia terdapat badan tertentu yang mengurusi data-data statistik. Di Indonesia, data-data statistik ditangani oleh Badan Pusat Statistik (BPS) yang tugas pokoknya menyediakan data bagi pemerintah dan Masyarakat umum. Hampir semua lembaga resmi pemerintah maupun perusahaan swasta memiliki bagian tersendiri yang mengurusi masalah data. Apabila membahas tentang data, maka akan ditemukan uda terminology yang hampir sama tetapi memiliki arti yang berbeda yaitu statistik dan statistika.

C. Statistik dan Statistika

Menurut Kamus Besar Bahasa Indonesia (KBBI) **statistika** merupakan ilmu tentang cara mengumpulkan, menyajikan, menggolongkan, menganalisis dan mencari keterangan yang bermakna dari data yang berupa angka. Selain itu statistika merupakan pengetahuan yang berhubungan dengan pengumpulan data, penyelidikan dan kesimpulannya berdasarkan bukti yang berupa catatan bilangan.

Sementara itu **statistik** adalah catatan angka-angka (bilangan) dan perangkaan. Statistik juga dapat didefinisikan sebagai data yang berupa angka yang dikumpulkan, disajikan, dikelompokkan sehingga dapat memberikan informasi yang berarti mengenai suatu masalah atau gejala. Jadi dapat disimpulkan bahwa statistik merupakan hasil atau informasi yang berupa angka yang didapatkan dari metode statistika yang dilakukan mulai dari mengumpulkan data, pemrosesan data, dan diakhiri dengan penginterpretasian hasil dan penarikan kesimpulan (Walpole, 1995).

D. Jenis-Jenis Statistik

Statistika dibagi menjadi dua kelompok, yakni **statistika deskriptif** dan **statistika inferensia**. Pada awal perkembangan ilmu statistika, statistika hanya digunakan untuk menggambarkan suatu fenomena dan menyimpulkannya ke dalam angka-angka. Inilah yang disebut dengan statistika deskriptif.

Statistika deskriptif paling umum dan paling banyak digunakan. Penggambaran paling umum dari statistika deskriptif yaitu dalam bentuk tabel, gambar dan grafik yang digunakan sebagai pendukung suatu pendapat atau opini. Jadi statistika deskriptif adalah bagian dari ilmu statistika yang berisi metode-metode penyajian data dan menarik kesimpulan sehingga lebih mudah dimengerti.

Adakalanya menggambarkan data secara keseluruhan dengan menggunakan statistika deskriptif sangat sulit atau bahkan tidak mungkin dilakukan. Misalkan untuk data-data yang lingkupnya sangat besar, seperti suatu negara. Sebagai contoh apabila ingin mengetahui tinggi badan siswa kelas X di Indonesia, maka akan sangat sulit dilakukan dan membutuhkan waktu yang lama untuk mendapatkan keseluruhan data. Olehkarena itu yang dapat dilakukan yaitu dengan menggunakan sampel beberapa ratus siswa kelas X. **Statistika inferensia** memungkinkan untuk menarik kesimpulan dari keseluruhan populasi hanya dengan menggunakan sampel yang digunakan. Statistika inferensia dapat digunakan dalam pengambilan Keputusan dengan tetap memperhatikan kesalahan (*error*) yang akan muncul. Jadi statistika inferensia merupakan suatu metode statistik yang digunakan untuk menarik kesimpulan atau bahkan memperhatikan keseluruhan populasi dengan menggunakan informasi yang diambil dari sampel (Mendenhall et al., 2020).

E. Data dan Jenis-Jenis Data

Dalam kehidupan sehari-hari, statistik lebih dikenal dengan istilah data, meskipun sebenarnya antara statistik dengan data berbeda. Perbedaan yang paling mendasar adalah data tidak selalu dalam bentuk angka, sementara statistik selalu berbentuk data. Sebagai contoh data yang tidak berbentuk angka adalah data-data kependudukan yang ada di KTP seperti nama, jenis kelamin, alamat, agama, status perkawinan, pekerjaan dan kewarganegaraan yang semuanya berupa data, tetapi tidak berbentuk angka. Akan tetapi untuk mempermudah, setelah bagian ini akan digunakan istilah data untuk menggantikan istilah statistik.

Berdasarkan cara memperolehnya, data dikelompokkan menjadi dua yaitu data primer dan data sekunder. **Data primer** adalah data yang dikumpulkan secara langsung dari sumber aslinya. Data yang dikumpulkan dilaporkan sesuai dengan data aslinya dan tidak mengalami proses perhitungan. Untuk melakukan pengumpulan data primer melalui observasi, eksperimen, survey dan sensus. Sementara itu **data sekunder** adalah data yang dikumpulkan dari pihak lain yang telah mengalami pemrosesan dan perhitungan. Contohnya adalah data yang dipublikasikan oleh BPS, laporan keuangan Perusahaan, data yang disajikan di website, dan sebagainya.

Selain itu, berdasarkan bentuk datanya, data dibedakan menjadi data kualitatif dan data kuantitatif. **Data kuantitatif** adalah data yang berbentuk angka atau yang dapat diangkakan. Sebagai contoh adalah jarak, umur, tinggi badan, harga, IPK dan sebagainya. Sementara itu **data kualitatif** adalah data yang berbentuk uraian atau yang berbentuk selain angka. Contohnya adalah jenis kelamin, pekerjaan, warna kulit, Alamat domisili, dan sebagainya. Akan tetapi dalam beberapa kasus, terdapat data yang berbentuk angka, tetapi sebenarnya adalah data kualitatif, misalnya kode pos, nomer telepon, dan kode barang.

Dalam praktiknya, data kualitatif dapat diubah kedalam bentuk kode/angka yang menunjukkan kategori. Misalnya untuk jenis kelamin, angka 0 untuk kategori laki-laki dan angka 1 untuk kategori Perempuan. Akan tetapi angka 0 dan angka 1 tersebut tidak bisa dilakukan operasi matematika. Pengkategorian biasanya digunakan untuk mempermudah ketika akan mencari jumlah atau proporsi untuk setiap kategorinya.

Data kuantitatif dibagi menjadi dua yaitu data diskrit dan data kontinu. **Data diskrit** merupakan data yang berbentuk angka bulat dan bukan pecahan yang merupakan hasil hitung. Sementara **data kontinu** merupakan data yang memiliki nilai terus menerus yang dapat berbentuk pecahan yang merupakan hasil dari pengukuran (Nugroho, 2008).

F. Koefisien dan Variabel

Dalam statistika inferensia sering didengar istilah koefisien dan variabel. Statistika inferensia Sebagian besar dinyatakan dalam bentuk persamaan matematis. Perhatikan persamaan matematis berikut.

$$y = 8x_1 + 5x_2 + 2$$

Berdasarkan persamaan matematis tersebut, terdapat 3 komponen, yakni variabel, koefisien dan konstanta. **Variabel** biasanya dinyatakan dalam bentuk huruf. Dari persamaan diatas yang dinamakan variabel adalah y, x_1 dan x_2. Meskipun dinyatakan dalam bentuk huruf, variabel merupakan serangkaian angka yang bisa ditemukan pada data yang jumlahnya bervariasi sehingga dapat diwakilkan dengan menggunakan huruf.

Variabel dapat dibedakan menjadi dua kelompok, yakni variabel bebas atau yang sering disebut dengan variabel independen dan variabel terikat yang biasa disebut dengan variabel dependen. **Variabel terikat** didefinisikan sebagai variabel yang nilainya dipengaruhi oleh perubahan nilai variabel bebas. Sedangkan **variabel bebas** merupakan variabel yang mempengaruhi variabel terikat. Pengaruh terhadap variabel terikat inilah yang kemudian diobservasi. Lazimnya, dalam sebuah persamaan matematis, variabel terikat berada di sebelah kiri tanda sama dengan dan variabel bebas berada di sebelah kanan tanda sama dengan. Pada persamaan matematis diatas, yang dinamakan variabel terikat adalah variabel y dan yang dinamakan variabel bebas adalah variabel x_1 dan x_2.

Sementara itu **koefisien** merupakan nilai yang berada didepan variabel. Pada persamaan matematis diatas, yang dinamakan koefisien adalah 8, 5 dan 1. Angka 8 menunjukkan koefisien untuk variabel x_1 dan angka 5 menunjukkan koefisien untuk variabel x_2. Sedangkan untuk variabel yang tidak mempunyai koefisien di depannya (y), maka nilai koefisiennya adalah 1. Selain istilah koefisien dan variabel, dalam statistika inferensia juga dikenal istilah konstanta. **Konstanta** merupakan suatu nilai yang tidak memiliki variabel. Disebut dengan konstanta karena mempunyai nilai yang tidak berubah. Pada persamaan matematis diatas yang dinamakan konstanta adalah 2. Dalam contoh yang lain, biasanya konstanta dinyatakan dalam huruf c (Yitnosumarto, 1990).

G. Skala Pengukuran

Ilmu statistika erat kaitannya dengan data dan perhitungan, oleh karena itu diperlukan skala pengukuran. Skala pengukuran digunakan untuk mengkategorikan dan mengkuantifikasian suatu variabel. Penggunaan skala pengukuran yang tepat akan menghasilkan pengukuran yang akurat terhadap suatu permasalahan tertentu. Terdapat empat skala pengukuran yakni skala nominal, skala ordinal, skala interval dan skala rasio (Arifin, 2014).

1. Skala Nominal

Skala nominal merupakan pengelompokan objek penelitian dalam kategori tertentu yang telah ditetapkan. Data kualitatif masuk dalam kategori skala nominal. Contoh yang paling mudah untuk skala nominal adalah jenis kelamin. Misalkan kode 0 untuk jenis kelamin laki-laki dan kode 1 untuk jenis kelamin perempuan. Meskipun dilambangkan dalam bentuk angka, namun angka tersebut tidak menunjukkan nilai, hanya menunjukkan pembeda saja. Contoh yang lain misalkan dilakukan survey Kesehatan terhadap pengungsi banjir di daerah "X". Pertanyaan yang ditanyakan adalah keluhuan Kesehatan yang paling dirasakan selama berada dalam pengungsian. Adapun pilihan jawabannya adalah :

1) Batuk-batuk
2) Sesak nafas
3) Gatal-gatal
4) Diare
5) Demam
6) Sulit tidur
7) Depresi
8) Lainnya

Para pengungsi diminta untuk mengisi jawaban keluhan Kesehatan yang paling mengganggu. Penomoran 1 hingga 8 tidak menunjukkan nilai sama sekali dan juga tidak menunjukkan bahwa penyakit nomor 1 paling berat diantara penyakit yang lain. Akan tetapi penomoran tersebut hanya digunakan untuk mengkategorikan jenis penyakit. Sehingga dapat dikatakan bahwa skala nominal merupakan skala pengukuran yang paling lemah.

2. Skala Ordinal

Apabila skala nominal menunjukkan kategori, maka skala ordinal mempunyai tingkatan yang lebih tinggi. Dengan menggunakan skala ordinal dapat digunakan untuk melihat peringkat dari yang diamati. Sebagai contoh dilakukan survey terhadap pengguna jasa pesawat terbang. Para responden diminta untuk memberikan ranking terhadap maskapai-maskapai berikut sesuai dengan aspek pelayanan di kabin dengan angka 1 hingga angka 5, dimana 1 menyatakan maskapai yang paling tinggi pelayanannya dan 5 menyatakan maskapai yang paling rendah pelayanannya. Kuesioner yang diisi oleh salah satu responden menunjukkan hasil sebagaimana ditunjukkan pada Tabel 1.1 berikut.

Tabel 1.1. Hasil Pengisian Kuesioner

Ranking	Nama Maskapai
4	Aman Air
2	Sukses Air
1	Cendrawasih Air
3	Harimau Air
5	Nusantara Air

Pada contoh tersebut, responden dapat menyatakan pilihannya dalam rangking, yakni maskapai yang paling baik pelayanannya adalah cendrawasih air dan yang paling jelek pelayanannya adalah Nusantara air. Meskipun demikian dengan menggunakan skala ordinal, tidak dapat digunakan untuk mengkuantifikasi perbedaan Tingkat kenyamanan pelayanan satu maskapai dengan maskapai lain. Yang dapat dilakukan hanya menyatakan bahwa orang akan merasa lebih puas apabila memilih sukses air apabila dibandingkan dengan memilih aman air (sukses air memiliki ranking 2 sementara itu aman air memiliki ranking 4). Namun dapat dinyatakan bahwa orang yang memilih sukses air akan merasa dua kali lipat lebih puas daripada jika memilih aman air.

3. **Skala Interval**

Skala interval memiliki tingkatan yang lebih tinggi daripada skala nominal dan ordinal. Selain dapat digunakan untuk menyatakan ranking, skala interval juga dapat digunakan untuk menyatakan jarak absolut dari dua pengukuran. Skala interval telah menggunakan data kuantitatif sebagai pengukurnya. Namun, sebagaimana skala ordinal, besarnya jarak antara satu pengukuran dengan pengukuran yang lain tidak menunjukkan arti apa-apa. Sebagai contoh untuk mengetahui kemampuan berbahasa inggris calon mahasiswa baru, Universitas X melakukan tes Bahasa inggris berupa tes TOEFL. Dari puluhan calon mahasiswa baru, berikut ini merupakan hasil yang didapatkan Amir, Budi, Candra dan Dewi.

Amir : 500
Budi : 600

Candra : 500
Dewi : 400
Berdasarkan contoh tersebut didapatkan informasi bahwa skor yang didapatkan Budi 100 poin lebih tinggi dibandingkan dengan skor yang didapatkan Amir. Selain itu dapat pula dinyatakan bahwa skor yang diperoleh Amir 100 poin lebih tinggi dibandingkan dengan skor yang didapatkan Dewi. Olehkarena itu jarak absolut Budi dan Amir serta Amir dan Dewi adalah sama-sama bernilai 100. Berdasarkan contoh tersebut juga dapat disimpulkan bahwa Budi merupakan calon mahasiswa dengan kemampuan Bahasa Inggris paling baik karena mempunyai nilai skor 600. Sementara itu Dewi memiliki kemampuan Bahasa Inggris paling jelek diantara yang lain karena memiliki skor paling rendah yakni 400. Besarnya skor Budi adalah 50% lebih tinggi dibandingkan dengan skor yang didapat oleh Dewi. Namun hal ini tidak menunjukkan bahwa Budi 50% lebih pintar dibandingkan dengan Dewi dalam hal kemampuan Bahasa Inggris.

4. **Skala Rasio**

Skala Rasio merupakan skala data yang mempunyai tingkatan paling tinggi. Skala Rasio tidak hanya menyatakan ranking dan perbedaan absolut dari data, namun juga menyatakan rasio dari data. Skala Rasio selalu dinyatakan dalam data kuantitatif dan tidak bisa bernilai kurang dari nol. Sebagai contoh Amir dan Budi sama-sama bekerja di PT X, Amir memperoleh gaji bulanan sebesar Rp. 5.000.000,00 dan Budi memperoleh gaji bulanan sebesar Rp. 10.000.000,00. Berdasarkan contoh tersebut dapat disimpulkan bahwa Budi memiliki gaji bulanan yang lebih tinggi daripada Amir dengan selisih absolut sebesar Rp. 5.000.000,00 yang juga berarti Budi memiliki gaji dua kali lipat lebih besar dibandingkan dengan Amir. Banyak contoh yang dapat digunakan untuk menyatakan skala rasio, diantaranya Panjang jalan, jumlah orang, umur, berat badan, harga barang dan sebagainya.

1

Perhatikan kuesioner kepuasan pelanggan Restoran X dibawah ini. Masuk dalam kategori manakah skala pengukuran ini?

Kuesioner Kepuasan Pelanggan Restoran X

Jawablah pertanyaan di bawah ini dengan memberikan tanda centang (√) pada aspek kepuasan menurut pengalaman Anda

Nilai Kepuasan :

1 : sangat tidak puas, 2 : tidak puas, 3 : ragu-ragu, 4 : puas, 5 : sangat puas

No	Aspek Kepuasan	1	2	3	4	5
1	Pilihan menu					
2	Rasa masakan					
3	Penyajian/tampilan makanan					
4	Kebersihan					
5	Pelayanan					
6	Harga					

Skala pengukuran pada contoh kuesioner tersebut adalah **skala likert**. Skala likert sering dijumpai dalam survey yang menggunakan kuesioner. Kemudian masuk kedalam kategori manakah skala likert? Belum ada pendapat yang sama dari para ahli statistik mengenai skala likert. Beberapa ahli memasukkan skala likert dalam kategori skala ordinal, namun mayoritas ahli statistik memasukkan skala likert dalam skala interval, karena persepsi orang terhadap jarak kategori kepuasan adalah sama. Sebagai contoh misalkan persepsi antara jarak sangat tidak puas dan tidak puas adalah sama dengan persepsi antara jarak puas dengan sangat puas

H. Metode Pengumpulan Data

Berdasarkan cara memperolehnya, data dibedakan menjadi dua yakni data primer dan data sekunder. Data sekunder adalah data yang dikumpulkan dari suatu sumber data yang telah diolah dan dipublikasikan oleh pihak lain. Pada suatu titik tertentu, semua data berasal dari sumber pertama (primer) yaitu data dikumpulkan dari sumber pertama dan kemudian diolah oleh pengumpul data.

Sesuai dengan lingkupnya, data dikumpulkan dengan dua macam cara, yakni dengan cara sensus dan sampling (Sudjana, 2001).

1. **Sensus** yaitu metode pengumpulan data dengan melibatkan seluruh populasi yang akan diteliti. Sensus biasanya dilakukan oleh negara karena memakan banyak biaya dan sumber daya yang besar. Contohnya sensus yang dilakukan secara regular oleh BPS seperti sensus penduduk, sensus ekonomi, sensus pertania dan sebagainya
2. **Sampling** yaitu metode pengumpulan data menggunakan Sebagian kecil dari keseluruhan populasi yang diamati. Pada umumnya peneliti tidak memiliki sumber daya yang cukup untuk melakukan pengambilan data secara sensus sehingga perlu dilakukan Teknik sampling.

Berikut ini merupakan beberapa metode pengumpulan data primer yang sering dilakukan

1. **Observasi** adalah cara pengumpulan data dengan pengamatan yaitu pengumpul data atau peneliti tidak bisa mempengaruhi nilai atau respons yang diberikan oleh responden. Pada umumnya, penelitian-penelitian di bidang ekonomi yang menggunakan cara observasi ini meskipun saat ini sudah berkembang penelitian ekonomi eksperimental
2. **Percobaan (eksperimen)** adalah cara mengumpulkan data Ketika para responden diberikan beberapa perlakuan yang berbeda dan kemudian diamati pengaruh perbedaan perlakuan ini terhadap hasil yang diperoleh. Teknik eskperimen ini biasanya digunakan pada penelitian bidang eksakta, seperti penelitian di bidang kedokteran, biologi, fisika dan sebagainya
3. **Wawancara** adalah tanya jawab langsung yang dilakukan oleh pengumpul data dan sumber data (narasumber). Wawancara bisa dilakukan secara langsung (berhadapan) maupun tidak langsung melalui percakapan telepon atau video conference. Pemilihan antara wawancara langsung dan tidak langsung tentunya sangat terkait dengan jarak antara pewawancara dan narasumber serta ketersediaan biaya. Wawancara secara langsung mempunyai kelebihan yakni informasi dapat diperoleh secara lengkap, karena dengan wawancara langsung si pewawancara bisa menangkap tidak hanya Bahasa lisan, tetapi juga Bahasa tubuh dari narasumber. Selain itu apabila informasi tidak jelas langsung dapat diklarifikasi sehingga tidak menimbulkan kesalahan interpretasi. Sementara itu kelemahan dari wawancara langsung yakni biaya yang mahal,

memakan waktu yang lama serta kemungkinan akan terjadi bias jawaban dari narasumber karena pertanyaan yang diarahkan atau narasumber merasa terintimidasi apabila jawaban yang diberikan tidak sesuai dengan pendapat yang berlaku umum

4. **Kuesioner** adalah sekumpulan pertanyaan penelitian yang dibuat secara tertulis dan membutuhkan respons secara tertulis. Ada dua macam pertanyaan yang diajukan dalam kuesioner yaitu pertanyaan tertutup dan pertanyaan terbuka. Pertanyaan tertutup adalah pertanyaan yang sudah memiliki alternatif jawaban sehingga para responden hanya diminta untuk memilih jawaban. Sementara itu pertanyaan terbuka adalah pertanyaan yang tidak memiliki pilihan jawaban sehingga responden dibebaskan menjawab apa saja sesuai dengan yang ingin disampaikan. Kelebihan dari Teknik ini adalah mengurangi kemungkinan kesalahan yang dilakukan oleh pewawancara dan yang diwawancarai. Sedangkan kelemahannya adalah tidak adanya control dari peneliti apakah kuesioner dijawab oleh responden yang bersangkutan, Tingkat respon balik yang rendah terutama apabila kuesioner dikirim lewat pos serta peneliti tidak bisa mengklarifikasi jawaban responden yang kurang jelas.

I. Populasi dan Sampel

Dalam statistik juga dikenal istilah populasi dan sampel. Misalnya seorang peneliti akan melakukan penelitian tentang "Pengaruh Pemberian Beasiswa terhadap Prestasi Belajar Mahasiswa Tidak Mampu di Indonesia". Yang menjadi populasi dari penelitian ini adalah seluruh mahasiswa tidak mampu di Indonesia yang menerima beasiswa. Namun karena keterbatasan sumber daya, maka penelitian hanya akan dilakukan di lima kota besar dengan jumlah sampel di masing-masing kota dibatasi hanya 100 mahasiswa saja atau secara keseluruhan menajdi 500 mahasiswa saja. Dari penjelasan tersebut, dapat disimpulkan bahwa **populasi** adalah sekelompok unit atau objek penelitian yang memiliki karakteristik tertentu yang akan dipelajari. Sementara itu **sampel** adalah bagian dari populasi yang akan dipelajari yang dianggap bisa mewakilkan keseluruhan populasi. Teknik pengambilan sanpel ini disebut dengan istilah sampling.

Pengambilan data populasi ini sangat mungkin dilakukan untuk penelitian dengan lingkup kecil atau penelitian yang melibatkan objek penelitian yang jumlahnya kecil. Namun, untuk penelitian dengan lingkup yang sangat besar, penggunaan Teknik sampling sangat umum dilakukan. Akan tetapi harus tetap memperhatikan kemungkinan kesalahan (sampling error) yang harus tetap diperhitungkan dalam pengambilan Keputusan. Beberapa alasan mengapa Teknik sampling sangat umum dilakukan adalah sebagai berikut (Mendenhall et al., 2020):

1. **Murah**

   Pengambilan data dengan menggunakan metode sampel akan jauh menurunkan biaya penelitian

2. **Cepat**

   Disamping biaya yang lebih murah, Teknik ini juga akan mempercepat waktu pengambilan data

3. **Akurat**

   Pada beberapa kasus dengan menggunakan sampel, data yang didapatkan akan menjadi lebih akurat. Untuk penelitian yang membutuhkan partisipasi berulang dari responden, kemungkinan kesalahan akan lebih besar pada jumlah observasi yang besar.

4. **Efek samping**

   Penelitian dengan kemungkinan efek samping yang buruk tidak mungkin dilakukan dengan menggunakan keseluruhan populasi, contohnya adalah penelitian daya tahan mobil dari tabrakan dan penelitian tentang pengaruh obat baru terhadap manusia

5. **Jumlah populasi yang tidak jelas**

   Adakalanya jumlah responden dapat terus bertambah atau berkurang sebelum masa penelitian berakhir. Contohnya penelitian mengenai beasiswa pada contoh sebelumnya. Pada awal penelitian dirancang, mungkin jumlah populasi mahasiswa adalah sebanyak n orang. Pada perjalanannya jumlah populasi mahasiswa dapat bertambah atau berkurang karena faktor-faktor yang tidak bisa dikontrol oleh peneliti.

J. Latihan Soal

1. Cocokkan jawaban dibawah ini dengan menarik garis pada jawaban yang sesuai

Jarak Jakarta ke Yogyakarta Warna Pakaian Jumlah pakaian dalam lemari	Jarak Jakarta ke Yogyakarta Warna Pakaian Jumlah pakaian dalam lemari
Pengaruh pemberian ASI terhadap bayi prematur Karakteristik vegetasi tepi pantai	Statistika deskriptif Statistika Inferensia
Jarak Jakarta ke Yogyakarta Jawaban benar dan salah Urutan prefelensi konsumen pada merk deterjen Perbedaan nilai IQ pada individu A dan B	Skala nominal Skala ordinal Skala interval Skala rasio

2. Apakah penelitian ini lebih tepat dilakukan dengan Teknik sensus atau sampling? Jelaskan alasannya!
   a. Lama belajar siswa kelas 4 di SD X
   b. Persepsi ibu-ibu rumah tangga terhadap tiga merk sabun pencuci piring
3. Teknik apa yang paling cocok digunakan untuk mengumpulkan data berikut ini? Berikan alasannya!
   a. Pengaruh imunisasi X terhadap daya tahan tubuh anak usia dibawah 5 tahun

b. Survei kepuasan mahasiswa terhadap tutorial online
c. Pendapat ahli olahraga terhadap jalannya pertandingan sepak bola antara negara A dan negara B

DAFTAR PUSTAKA

Arifin, M. H. (2014). Konsep-konsep Dasar statistika. *Jakarta: Universitas Terbuka*.

Bluman, A. G. (2018). *Elementary Statistics: A Step By Step Approach* (Tenth Edit). McGraw-Hill Education.

Janah, M., & Kartini, A. Y. (2022). Penerapan Metode Regresi Linier Berganda Pada Kasus Balita Gizi Buruk Di Kabupaten Bojonegoro. *Jurnal Statistika Dan Komputasi*, *1*(2), 74–82.

Mendenhall, W. (2013). *Introduction to Probability and Statistics, 3rd*. Nelson Educatio.

Mendenhall, W., Beaver, R. J., & Beaver, B. M. (2012). *Introduction to probability and statistics*. Cengage Learning.

Mendenhall, W., Beaver, R. J., & Beaver, B. M. (2020). *Introduction to probability and statistics* (Fifteenth). Cengage Learning, Inc.

Nugroho, S. (2007). *Dasar-dasar Metode Statistika*. Grasindo.

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

Santoso, S. (2016). *Panduan Lengkap SPSS Versi 23*. Gramedia Direct.

Steel, R. G. D., & Torrie, J. H. (1976). Introduction to statistics. *(No Title)*.

Sudjana. (2001). *Metode Statistika*. Tarsito.

Sudjana. (2015). *Metode Statistika*. PT Tarsito Bandung.

Walpole, R. E. (1995). *Pengantar statistika*.

Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2016). Probability & Statistics for Engineers & Scientists. In *Journal of the American Statistical Association* (Ninth Edit). Pearson Education Limited.

Yitnosumarto, S. (1990). Dasar-Dasar Statistika. *C. V Rajawali: Jakarta*.

BAB 2
Penyajian Data

A. Tujuan Pembelajaran

Kemampuan akhir yang ingin dicapai setelah mempelajari bab ini adalah secara umum mahasiswa diharapkan mampu menyajikan data dalam bentuk tabel maupun diagram serta mampu menginterpretasikannya. Selain itu, secara khusus, mahasiswa diharapkan mampu :

1. Menjelaskan benang merah dari data yang disajikan
2. Membuat tabel angka distribusi frekuensi untuk data Tunggal
3. Membuat tabel distribusi frekuensi untuk data berkelompok
4. Menggambar diagram lingkaran
5. Menggambar histogram
6. Menggambar grafik garis

B. Pendahuluan

Data yang telah dikumpulkan kemudian diolah untuk diterjemahkan ke dalam sebuah laporan atau bacaan yang dapat dipahami oleh banyak orang. Data merupakan sekumpulan informasi berbentuk angka yang dapat mendukung suatu informasi tentang suatu hal. Walaupun sangat penting, penyajian data secara mentah akan membingungkan atau bahkan akan menyebabkan interpretasi yang berbeda-beda. Untuk itulah penyajian data secara benar dan efektif menjadi hal yang penting.

C. Mencari Benang Merah dari Data

Dalam kehidupan sehari-hari, secara sengaja maupun tidak sengaja tentu banyak berhubungan dengan data. Data banyak dijumpai di buku-buku teks, artikel ilmiah, artikel surat kabar, artikel majalah atau pemberitaan di banyak media lainnya. Data dalam bentuk angka lebih mudah dimengerti dan kemungkinan besar tidak memiliki interpretasi ganda. Misalnya Ketika ingin menjelaskan jarak antara Jakarta dan Yogyakarta, akan lebih jelas dinyatakan dalam bentuk "Jarak antara Jakarta dan Yogyakarta adalah kurang lebih 600 km" dibandingkan jika dinyatakan dalam bentuk "Jarak antara Jakarta dan Yogyakarta jauh". Kata "jauh" akan menimbulkan interpretasi yang macam-macam, terutama bagi orang-orang yang belum pernah bepergian dari Jakarta ke Yogyakarta begitupun sebaliknya. Sementara itu apabila dinyatakan dalam "600 km", maka 600 km merupakan sebuah ukuran yang absolut.

Namun, apabila sudah menyangkut data yang kompleks dan banyak jumlahnya, maka penyajian data secara mentah menjadi kurang efektif. Misalkan dalam satu penelitian, peneliti ingin satu per satu menggambarkan umur dari 50 responden seperti dibawah ini

35	27	37	28	39	46	41	23	19	18
17	25	23	19	20	27	33	32	33	54
50	47	45	36	49	44	43	34	49	23
25	28	40	50	38	34	27	35	37	37
28	29	34	21	22	24	27	41	36	41

Penyajian data seperti diatas, walaupun tepat, akan tetapi kurang jelas dan kurang menarik karena pembaca dipaksa untuk membaca satu per satu angka yang disajikan dan letaknya pun tidak berurutan, kemudian menarik kesimpulannya sendiri. Bagaimana misalnya jumlah respondennya adalah 2000 orang? Bagaimana bisa mengakomodasi data 2000 orang dalam bentuk teks? Data diatas akan lebih mudah dipahami apabila diubah dalam bentuk informasi sebagaimana ditunjukkan pada Tabel 2.1.

Dari tabel 2.1 tersebut dapat ditarik kesimpulan misalnya 64 % dari respon berusia antara 20 sampai 39 tahun. Keunggulan penyajian data dengan menggunakan tabel dan grafik adalah lebih memudahkan pembaca menarik benang merah atau kesimpulan dari data yang disajikan. Keunggulan lainnya adalah penyajian data menggunakan tabel dan grafik akan membuat tampilan data menjadi lebih menarik. Akan tetapi, penyajian data seperti pada Tabel 2.1 juga memiliki kelemahan yakni menghilangkan perincian penting yang mungkin diperlukan. Misalnya diantara kelompok umur 20 – 29 tahun, berapa yang berumur 20, 25 atau 28 tahun?

Tabel 2.1 Umur Responden

Kelompok Umur	Proporsi (%)
> 20	8
20 – 29	34
30 – 39	30
40 – 49	22
> 50	6

Terdapat beberapa bentuk penyajian data yang lazim digunakan dalam suatu laporan atau buku teks yakni tabel, diagram dan grafik. Setiap bentuk penyajian data memiliki fungsi yang berbeda meskipun bentuk yang dipilih adalah sesuai dengan keinginan dari penyampai data.

D. Menyimpan Data dengan Tabel

Setelah data terkumpul, langkah berikutnya yaitu menyimpan atau mendokumentasikan data. Data dapat disimpan secara manual dengan cara menuliskan data-data ke dalam kolom-kolom di kertas atau buku catatan atau dapat pula didokumentasikan dengan menggunakan computer. Cara terakhir ini yang dianggap lebih aman dan lebih efektif karena dapat menyimpan data dalam jumlah yang tak terhingga dan dapat dibuka Kembali apabila dibutuhkan. Perangkat lunak seperti Microsoft Excel paling dapat diandalkan dan paling mudah untuk mendokumentasikan data. Selain itu menyimpan data dengan excel akan memungkinkan untuk mengekspor data ke dalam program komputer yang lain.

Untuk menyimpan data baik secara manual maupun digital, dianjurkan untuk mendokumentasikannya dalam bentuk tabel. Pembuatan tabel secara manual hanya dimungkinkan untuk data-data sederhana yang jumlahnya tidak terlalu besar. Meskipun demikian, pengetahuan mengenai penyusunan data secara manual juga tetap penting untuk dipelajari.

E. Tabel Distribusi Frekuensi

Tabel berisi informasi berupa angka-angka yang disajikan dalam baris dan kolom. Tabel merupakan penyajian data yang paling sering digunakan, karena dapat memuat banyak informasi dan juga karena jumlah baris maupun kolomnya tidak terbatas. Pada dasarnya, segala jenis angka dapat disajikan dalam bentuk tabel, apakah itu jumlah, proporsi dan sebagainya. Namun, menyajikan terlalu banyak angka dalam satu tabel juga akan membingungkan dan menjadi kurang efektif. Untuk itu perlu dilakukan pemilihan data apa yang dapat atau tidak dapat disajikan dalam bentuk tabel.

Pada dasarnya semua data berawal dari data mentah. Penyajian data mentah sesuai dengan bentuk aslinya tentu sangat membingungkan. Untuk itu diperlukan penyajian dalam bentuk tabel. Hal ini dikarenakan ada kemungkinan bahwa suatu nilai yang sama dapat muncul lebih dari satu kali. Salah satu bentuk tabel yakni tabel distribusi frekuensi.

Teknik yang paling umum digunakan untuk Menyusun tabel distribusi frekuensi yaitu dengan sistem turus. Turus sangat umum digunakan untuk menghitung frekuensi dan biasanya digunakan dalam pencatatan hasil voting, misalnya pemilihan ketua kelas, ketua RT bahkan pencatatan hasil pemilu di Tingkat TPS. Berikut Langkah-langkah dalam pembuatan tabel distribusi frekuensi

1. Membuat tabel, memberikan label, menghitung turus, menghitung frekuensi serta jumlah baris yang disesuaikan dengan rentangan data dari yang terkecil hingga terbesar
2. Memasukkan semua nilai dengan cara diurutkan dari yang terkecil hingga yang terbesar pada kolom pertama
3. Menghitung frekuensi dengan cara membuat turus pada kolom kedua
4. Menghitung jumlah turus pada setiap baarisnya dan menuliskannya di kolom ketiga

Perhatikan contoh data umur responden di bawah ini

35	27	37	28	39	46	41	23	19	18
17	25	23	19	20	27	33	32	33	54
50	47	45	36	49	44	43	34	49	23
25	28	40	50	38	34	27	35	37	37
28	29	34	21	22	24	27	41	36	41

Dengan menggunakan turus didapatkan tabel distribusi frekuensi sebagaimana ditunjukkan pada Tabel 2.2 berikut.

Tabel 2.2 Menyusun Tabel Distribusi Frekuensi dengan Turus

Umur	**Turus**	**Frekuensi**
25	II	2
26	I	1
27	~~IIII~~	5
28	IIII	4
29	II	2
30		0
31	II	2
32	III	3
33	II	2
34	IIII	4
35	II	2
36	III	3
37	IIII	4
38	I	1
39	I	1
40	III	3
41	III	3
42	II	2
43	II	2
44	I	1
45	III	3
Total		50

F. Tabel Distribusi Frekuensi Relatif dan Distribusi Frekuensi Kumulatif

Distribusi frekuensi relative menunjukkan proporsi relative munculnya angka tertentu terhadap jumlah frekuensi secara keseluruhan. Cara menghitungnya yaitu dengan membagi jumlah frekuensi pada setiap baris dengan jumlah keseluruhan data. Jumlah keseluruhan dari frekuensi relative ini akan sama dengan 1. Selain dapat dituliskan dalam bentuk proporssi, distribusi relative juga dapat dituliskan dalam bentuk persentase. Caranya dengan mengalikan setiap proporsi frekuensi dengan 100%. Hasil penjumlahan persentase distribusi relative akan sama dengan 100%. Distribusi frekuensi kumulatif diperoleh dengan cara menjumlahkan frekuensi di setiap baris dengan baris sebelumnya. Untuk mengecek apakah frekuensi kumulatif yang telah dihitung benar atau salah yaitu dengan membandingkan dengan angka kumulatif di baris terakhir dengan jumlah datanya. Apabila jumlahnya sama maka perhitungan yang dilakukan adalah benar. Sebagai tambahan, juga dapat ditampilkan persentase frekuensi kumulatifnya. Apa kegunaan dari tabel distribusi frekuensi? Tabel distribusi frekuensi dapat digunakan untuk membantu menarik kesimpulan dari data yang disajikan.

Berdasarkan Tabel 2.2 dapat dicari distribusi relative, persentase distribusi relative dan juga distribusi kumulatifnya sebagaimana ditunjukkan pada Tabel 2.3 berikut.

Tabel 2.3 Tabel Distribusi Frekuensi Relatif dan Distribusi Frekuensi Kumulatif

Umur	**Frekuensi**	**Frekuensi Relatif**	**% Frekuensi Relatif**	**Frekuensi Kumulatif**	**% Frekuensi Kumulatif**
25	2	$\frac{2}{50} = 0{,}04$	$0{,}04 \times 100\% = 4\%$	2	4 %
26	1	$\frac{1}{50} = 0{,}02$	$0{,}02 \times 100\% = 2\%$	1 + 2 = 3	2 % + 4 % = 6 %
27	5	0,10	10 %	5 + 3 = 8	10% + 6%

Umur	Frekuensi	Frekuensi Relatif	% Frekuensi Relatif	Frekuensi Kumulatif	% Frekuensi Kumulatif
					= 16 %
28	4	0,08	8 %	4 + 8 = 12	24 %
29	2	0,04	4 %	14	28 %
30	0	0,00	0 %	14	28 %
31	2	0,04	4 %	16	32 %
32	3	0,06	6 %	19	38 %
33	2	0,04	4 %	21	42 %
34	4	0,08	8 %	25	50 %
35	2	0,04	4 %	27	54 %
36	3	0,06	6 %	30	60 %
37	4	0,08	8 %	34	68 %
38	1	0,02	2 %	35	70 %
39	1	0,02	2 %	36	72 %
40	3	0,06	6 %	39	78 %
41	3	0,06	6 %	42	84 %
42	2	0,04	4 %	44	88 %
43	2	0,04	4 %	46	92 %
44	1	0,02	2 %	47	94 %
45	3	0,06	6 %	50	100 %
Jumlah	50	1,00	100 %		

Berdasarkan Tabel 2.3 dapat diambil beberapa kesimpulan diantaranya adalah sebagai berikut:

1. Sebanyak 28% responden berusia ≤ 30 tahun
2. Mayoritas responden berusia antara 30 sampai 40 tahun dengan persentase sebesar 50 %
3. Sebanyak 10 % dari responden (dan merupakan persentase terbanyak) berumur 27 tahun
4. Dan sebagainya

Menarik kesimpulan dengan menggunakan tabel akan lebih mudah dilakukan dibandingkan hanya dengan bergantung pada data mentah

G. Tabel Distribusi Frekuensi untuk Data Berkelompok

Pengelompokan data dapat dilakukan untuk data-data yang rentang nilainya besar atau data yang perbedaan nilai terkecil dan terbesarnya besar. Pada dasarnya besarnya rentangan kelompok-kelompok data merupakan Keputusan dari si penyusun data. Namun, perlu diperhatikan apabila interval kelas terlalu kecil, hal ini akan mengurangi kepraktisan karena akan menghasilkan terlalu banyak kelas. Apabila interval kelasnya terlalu lebar, akan menyebabkan hilangnya detail-detail data.

Teknik yang dapat digunakan untuk Menyusun tabel distribusi frekuensi untuk data berkelompok adalah Teknik *steam and leaf* (dahan dan daun). Perhatikan Kembali contoh data umur responden di bawah ini

35	27	37	28	39	46	41	23	19	18
17	25	23	19	20	27	33	32	33	54
50	47	45	36	49	44	43	34	49	23
25	28	40	50	38	34	27	35	37	37
28	29	34	21	22	24	27	41	36	41

Berdasarkan data diatas mempunyai rentangan data yang lebih lebar, yakni 17 sampai 54. Penggunaan data berkelompok untuk data-data dengan rentangan besar akan mempermudah dalam menyajikan dalam bentuk tabel dan juga mempermudah dalam penarikan kesimpulan. Namun kelemahannya adalah pengelompokan data dapat menghilangkan detail-detail dari data.

Untuk melakukan Teknik *steam and leaf*, lakukan Langkah-langkah sebagai berikut (Santoso, 2016):

1. Sediakan dua kolom yakni satu untuk kolom *steam* dan satu lagi untuk kolom *leaf*. Beri label keduanya. Jumlah baris mengikuti urutan angka digit pertama pada data, misal untuk contoh diatas, dimulai dari 1 (17) dan diakhiri dengan 5 (54)
2. Urutkan angka digit pertama dari baris-baris yang telah dibuat
3. Masukkan angka digit kedua pada kolom kedua
4. Rapikan angka digit kedua dengan mengurutkannya

Sehingga didapatkan steam and leaf sebagaimana ditunjukkan pada Tabel 2.3 berikut.

Tabel 2.3 *Steam and Leaf*

Steam	***Leaf***
1	9 8 7 9
2	7 8 3 5 3 0 7 3 5 8 7 8 9 1 2 4 7
3	5 7 9 3 2 3 6 4 8 4 5 7 7 4 6
4	6 1 7 5 9 4 3 9 0 1 1
5	4 0 0
Cara membacanya bila steam = 1, leaf = 9 maka datanya adalah 19	

Berikutnya ikuti Langkah keempat dengan mengurutkan bilangan digit kedua dan didapatkan hasil sebagaimana ditunjukkan pada Tabel 2.4 berikut.

Tabel 2.4 *Steam and Leaf* (setelah diurutkan)

Steam	***Leaf***
1	7 8 9 9
2	0 1 2 3 3 3 4 5 5 7 7 7 7 8 8 8 9
3	2 3 3 4 4 4 5 5 6 6 7 7 7 8 9
4	0 1 1 1 3 4 5 6 7 9 9
5	0 0 4
Perhatikan sekarang data terendah dimulai dari 17 dan berakhir dengan 54	

Berdasarkan pengelompokan pada Tabel 2.3 didapatkan lima interval untuk kelompok umur yang dapat diringkas sebagaimana ditunjukkan pada Tabel 2.5 berikut. Jarak interval kelas untuk semua kelompok adalah sama, yaitu 10. Sangat dianjurkan untuk membuat interval kelas dengan jarak yang sama. Apabila interval kelas yang dibuat terlalu lebar, maka dapat membagi lagi interval kelas ke dalam kelompok yang lebih kecil, misalnya interval kelas dengan jarak 5. Caranya yaitu dengan membuat steam and leaf Kembali dan membagi digit kedua ke dalam dua kelompok. Kelompok pertama memiliki angka 0 – 4 dan kelompok kedua memiliki angka 5 – 9 sebagaimana ditunjukkan pada Tabel 2.6.

Tabel 2.5 Distribusi Frekuensi dengan Kelompok Umur

Kelompok Umur	**Frekuensi**
10 – 19	4
20 – 29	17
30 – 39	15
40 – 49	11
50 – 59	3
Jumlah	**50**

Tabel 2.6 *Steam and Leaf* (dengan jarak interval 5)

Steam	***Leaf***
1*	7 8 9 9
2	0 1 2 3 3 3 4
2*	5 5 7 7 7 7 8 8 8 9
3	2 3 3 4 4 4
3*	5 5 6 6 7 7 7 8 9
4	0 1 1 1 3 4
4*	5 6 7 9 9
5	0 0 4
Tanda *) menunjukkan kelompok kedua	

Tabel *steam and leaf* pada Tabel 2.6 dapat diringkas ke dalam tabel distribusi frekuensi dengan kelompok umur sebagaimana ditunjukkan pada Tabel 2.7. keuntungan menggunakan interval kelas yang lebih kecil adalah tidak akan terlalu kehilangan detail dari data. Misalnya pada pembagian kelompok di Tabel 2.4, responden dengan usia termuda, yaitu 17 tahun, akan masuk ke dalam kategori 10 – 19, sedangkan tidak terdapat responden dengan usia antara 10 – 16 tahun. Demikian pula responden paling tua yaitu yang berusia 54 tahun. Pada Tabel 2.5 usia 54 tahun dimasukkan pada kelompok umur 50 – 59 tahun, padahal tidak ada responden yang berusia 55 – 59 tahun.

Tabel 2.7 Distribusi Frekuensi dengan Kelompok Umur

Kelompok Umur	Frekuensi
15 – 19	4
20 – 24	7
25 – 29	10
30 – 34	6
35 – 39	9
40 – 44	6
45 – 49	5
50 – 54	3
Jumlah	**50**

Beberapa hal yang perlu diperhatikan dalam membuat kelompok data adalah sebagai berikut (Yitnosumarto, 1990):

1. Setiap data hanya bisa menjadi anggota dari satu kelompok saja
2. Semua data harus memiliki kelompok
3. Untuk alasan kepraktisan, lazimnya lebar interval kelas adalah antara 5 sampai 15

H. Tabel Distribusi Frekuensi Relatif dan Distribusi Frekuensi Kumulatif untuk Data Berkelompok

Seperti halnya pada data tunggal, tabel distribusi frekuensi relatif dan distribusi frekuensi kumulatif dapat disusun untuk data berkelompok. Berdasarkan Tabel 2.7 dapat disusun menjadi tabel distribusi relatif dan distribusi frekuensi kumulatif untuk data berkelompok sebagaimana ditunjukkan pada Tabel 2.8 berikut.

Tabel 2.8 Tabel Distribusi Relatif dan Distribusi Frekuensi Kumulatif untuk Data Berkelompok

Umur	Frekuensi	Frekuensi Relatif	% Frekuensi Relatif	Frekuensi Kumulatif	% Frekuensi Kumulatif
15 – 19	4	0,08	8 %	4	8 %
20 – 24	7	0,14	14 %	11	22 %

25 – 29	10	0,20	20 %	21	42 %
30 – 34	6	0,12	12 %	27	54 %
35 – 39	9	0,18	18 %	36	72 %
40 – 44	6	0,12	12 %	42	84 %
45 – 49	5	0,10	10 %	47	94 %
50 – 54	3	0,06	6 %	50	100 %
Jumlah	50	1,00	100 %		

Berikut ini terdapat beberapa istilah penting yang digunakan untuk data berkelompok antara lain (Mendenhall et al., 2020):

1. Interval kelas adalah kelompok-kelompok data. Berdasarkan Tabel 2.8 interval kelasnya adalah sebagai berikut:
   a. 15 – 19 adalah interval kelas pertama
   b. 20 – 24 adalah interval kelas kedua
   c. 25 – 29 adalah interval kelas ketiga
   d. 30 – 34 adalah interval kelas keempat
   e. Dan seterusnya
2. Batas kelas adalah nilai terkecil dan nilai terbesar dari setiap kelompok. Nilai terkecil disebut batas bawah dan nilai terbesar disebut batas atas. Berdasarkan Tabel 2.8, yang merupakan batas bawah dari interval kelas 15-19 adalah 15 dan yang menajdi batas atasnya adalah 19
3. Tepi kelas adalah jarak nyata antar kelas. Cara mencari tepi kelas adalah sebagai berikut:
   a. Tepi bawah = batas bawah – 0,5
   b. Tepi atas = batas atas + 0,5

   Berdasarkan Tabel 2.8 yang menjadi tepi bawah untuk kelas interval 15 – 19 adalah 14,5 dan tepi atasnya adalah 19,5
4. Lebar kelas adalah selisih antara tepi kelas atau lebar kelas = tepi atas – tepi bawah. Berdasarkan Tabel 2.8 lebar kelasnya adalah 19,5 – 14,5 = 5.
5. Titik Tengah adalah nilai yang berada tepat di Tengah-tengah interval kelas. Untuk menghitungnya digunakan rumus :

   Titik Tengah = (batas atas + batas bawah) / 2

Berdasarkan Tabel 2.8 yang merupakan titik Tengah untuk interval 15 – 19 adalah 17

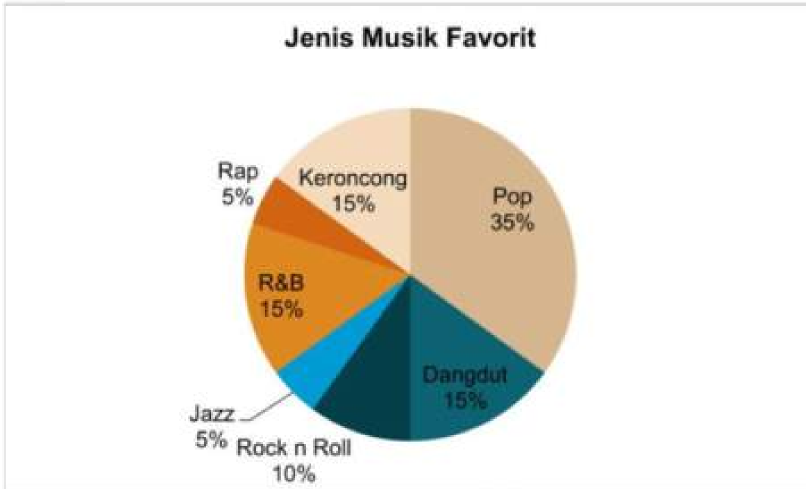
I. Diagram Lingkaran

Adakalanya menampilkan data dengan tabel akan membingungkan pembaca, terutama apabila data yang akan disajikan banyak jumlahnya sehingga membutuhkan banyak kolom dan baris. Olehkarena itu ad acara lain yang juga efektif dalam penyampaian data, yaitu dengan menggunakan diagram dan gambar. Mengapa diagram dan grafik? Diagram dan grafik atau bentuk visual dianggap efektif karena selain mudah dimengerti juga lebih menarik. Selain tampilan bentuk yang lebih menarik, juga memungkinkan bagi penyampai data untuk menyajikan data dalam pilihan-pilihan warna.

Diagram lingkaran (pie chart) digunakan untuk menunjukkan relative dari data. Diagram lingkaran mirip kue yang dipotong-potong. Potongan-potongan inilah yang merupakan ukuran relatif yang dinyatakan dalam persentase. Diagram lingkaran ini tidak memiliki dimensi waktu karena hanya digunakan untuk menunjukkan ukuran relatif dari beberapa kategori dalam satu waktu tertentu (Nugroho, 2008). Perhatikan contoh di bawah ini yang menggambarkan jenis musik yang menjadi favorit 20 orang responden.

Pop	Dangdut	Rock	Jazz	R&B	Rap	Keroncong
7	3	2	1	3	1	3

Berdasarkan data tersebut didapatkan diagram lingkaran sebagaimana ditunjukkan pada Gambar 2.1. Dengan menggunakan diagram lingkaran dapat dengan mudah melihat jenis lagu mana yang paling disukai dan yang paling tidak disukai 20 responden tersebut. Diagram lingkaran dapat digunakan untuk membantu dalam pengambilan kesimpulan secara singkat. Berdasarkan contoh tersebut, dapat disimpulkan bahwa jenis musik yang paling banyak disukai adalah pop dan yang paling tidak disukai adalah rap dan jazz



Gambar 2.1 Diagram Lingkaran

Bagaimana cara membuat diagram lingkaran? Diagram lingkaran dapat digambarkan baik secara manual maupun dengan bantuan komputer. Membuat diagram lingkaran dengan manual, meskipun mudah, namun memerlukan waktu yang lebih lama dan membutuhkan ketelitian. Untuk membuat diagram lingkaran dengan cara manual, dapat dilakukan dengan langkah-langkah sebagai berikut (Arifin, 2014).

1. Mengubah data ke dalam bentuk persentase, yaitu dengan membagi jumlah data pada setiap kategori dengan jumlah keseluruhan data dibagi 100%
2. Mengalikan persentase kategori yang diperoleh pada Langkah pertama dengan 3,6. Hal ini karena sebuah lingkaran memiliki 360 derajat maka setiap 1% akan diwakili oleh 3,6 derajat.
3. Menggambar lingkaran penuh dengan bantuan busur, kemudian membagi lingkaran sesuai dengan persentase setiap kategori yang didapatkan pada Langkah kedua
4. Menuliskan label pada setiap bagian dan juga memberikan judul

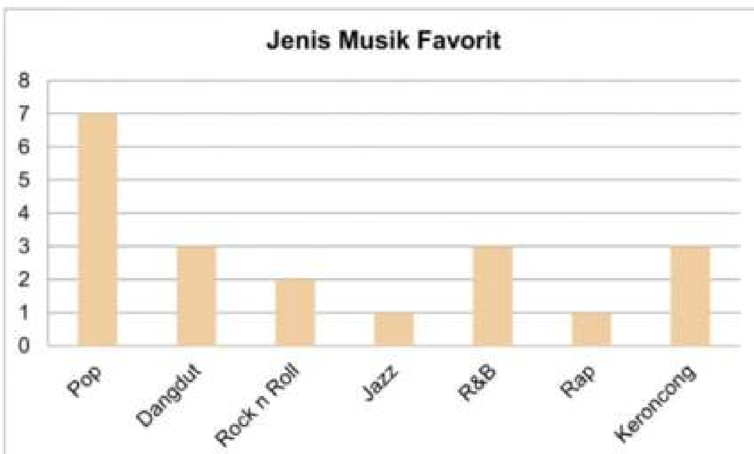
J. Diagram Batang

Diagram batang adalah penggambaran data secara grafis dengan menggunakan batang-batang vertikal. Batang-batang ini mempunyai tinggi yang berbeda-beda sesuai dengan nilai datanya. Diagram batang memiliki fungsi yang hampir sama dengan diagram lingkaran.

Perbedaannya adalah apabila di diagram lingkaran yang ditunjukkan adalah ukuran relative dari suatu kategori terhadap kategori lainnya, sedangkan pada diagram batang yang ditunjukkan adalah perbedaan ukuran absolutnya. Sama halnya dengan diagram lingkaran, diagram batang tidak memiliki dimensi waktu dan sangat cocok untuk membandingkan suatu kategori dengan kategori lainnya (Walpole et al., 2016). Sebagaimana contoh pada pembahasan sebelumnya yang menggambarkan jenis musik yang menjadi favorit 20 orang responden.

Pop	Dangdut	Rock	Jazz	R&B	Rap	Keroncong
7	3	2	1	3	1	3

Didapatkan diagram batang sebagaimana ditunjukkan pada Gambar 2.2 berikut.



Gambar 2.2 Diagram Batang

Sama halnya dengan diagram lingkaran, dengan diagram batang dapat dilihat jenis lagu mana yang paling disukai dan yang paling tidak disukai dengan membandingkan tinggi dari setiap batang kriteria. Dari contoh tersebut dapat disimpulkan bahwa jenis musik yang paling disukai adalah pop dan yang paling tidak disukai adalah rap dan jazz.

K. Histogram

Histogram memiliki bentuk yang hampir sama dengan diagram batang, yaitu menggambarkan batang-batang dengan tinggi yang berbeda-beda sesuai dengan data yang diwakilkannya. Namun histogram memiliki fungsi yang berbeda dari diagram batang. Apabila diagram batang digunakan untuk menunjukkan ukuran suatu kategori (diskrit), maka histogram digunakan untuk menunjukkan rentangan nilai dari beberapa observasi dalam kategori yang sama. Karena bersifat rentang nilai, histogram biasanya digunakan untuk data-data kontinu.

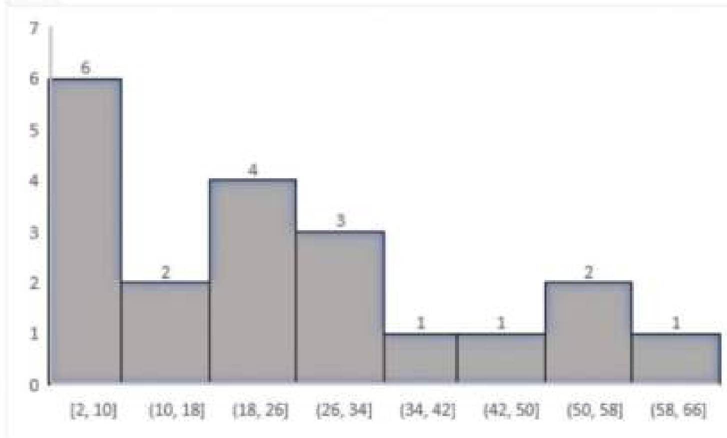
Secara bentuk fisik sangat berbeda antara diagram batang dan histogram. Pada diagram batang, batang-batang tidak saling menyentuh satu sama lain (terpisah), sedangkan pada histogram, karena datanya adalah kontinu, batang-batangnya saling bersentuhan (berhimpit). Diagram batang jarak (space) antara dua batang menunjukkan bahwa tidak ada nilai pada daerah tersebut. Misalnya apabila menunjukkan data jumlah kepemilikan mobil antara rumah tangga A (satu mobil) dan rumah tangga B (dua mobil), batangnya akan terpisah antara rumah tangga A dan rumah tangga B karena tidak mungkin aka nada rumah tangga C yang memiliki 1,5 mobil (Sudjana, 2001).

Sebaliknya pada histogram, karena datanya bersifat kontinu, ada kemungkinan setiap batang saling berhimpit. Contoh penggunaan histogram adalah menampilkan berat badan. Antara satu individu dengan individu yang lain dimungkinkan adanya nilai di tengahnya, misalnya A memiliki berat 60 kg dan B memiliki berat 65 kg, dimungkinkan ada individu lain yang memiliki berat antara berat tersebut, misalnya 62,5 kg atau 65,2 kg dan seterusnya. Selain data yang bersifat kontinu, histogram menampilkan data dalam kelompok-kelompok, atau bisa juga dikatakan kelas interval. Hal ini pula yang menyebabkan batang-batang pada histogram saling berhimpit. Perhatikan contoh berikut :

Data umur penerima vaksinasi pada puskesmas A

2	4	27	32	5	63	25	18	16	25
4	45	29	19	22	51	58	9	42	6

Berdasarkan data tersebut didapatkan distogram sebagaimana ditunjukkan pada Gambar 2.3 berikut.



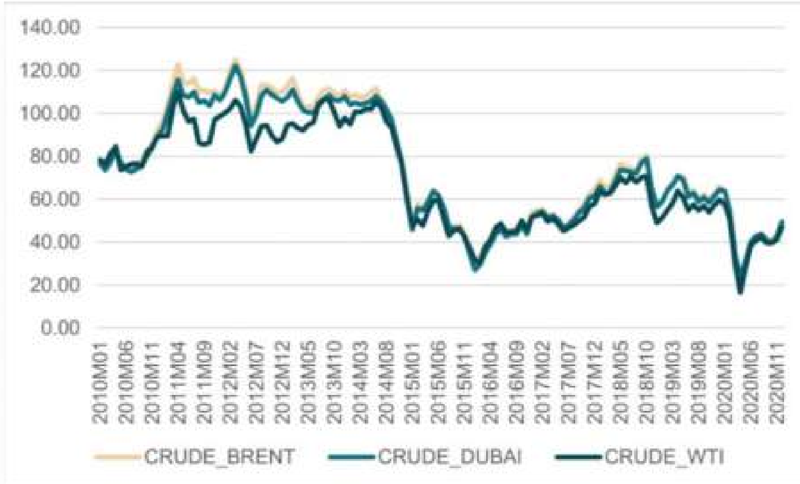
Gambar 2.3 Histogram

Berdasarkan Gambar 2.3, data dalam histogram ditampilkan dalam kelompok-kelompok umur. Kelompok 1 berisi individu yang berusia antara 2 – 10 tahun, kelompok 2 berusia 10 – 18 tahun dan seterusnya. Kelompok umur disajikan pada sumbu x, sedangkan frekuensi setiap kelompoknya disajikan pada sumbu y.

L. Diagram Garis (*Line Graph*)

Diagram garis adalah diagram yang digunakan untuk memberikan informasi runtut waktu (time series). Diagram garis pada dasarnya adalah garis yang menghubungkan titik-titik data pada waktu yang berbeda-beda. Diagram garis sangat cocok digunakan untuk menyajikan informasi historis suatu variabel dari waktu ke waktu, misalnya data pertumbuhan PDB, data pertumbuhan penduduk, data pertumbuhan keuntungan perusahaan, data perkembangan harga komoditas dan sebagainya (Steel & Torrie, 1976).

Diagram garis juga dapat digunakan untuk membandingkan dua atau lebih set data dengan variabel yang sama yaitu dengan cara menggambarkan dua atau lebih diagram garis pada diagram yang sama. Berikut adalah data perkembangan harga minyak dunia dalam US$/barrel. Perhatikan bahwa terdapat tiga garis yang menggambarkan tiga pasar minyak mentah yaitu Brent, Dubai dan WTI sebagaimana ditunjukkan pada Gambar 2.4 berikut.



Gambar 2.4 Diagram Garis

Berdasarkan Gambar 2.4 dapat dilihat perkembangan harga minyak mentah bulanan dunia mulai tahun 2010 sampai dengan 2020. Selain menunjukkan perkembangan harga minyak secara historis, juga dapat digunakan untuk membandingkan harga minyak mentah dari beberapa pasar yang berbeda. Secara umum, harga di setiap pasar memiliki pergerakan yang sama, tetapi memiliki harga yang sedikit berbeda.

M. Latihan Soal

Perhatikan data umur 20 penghuni panti jompo dibawah ini

75	72	85	80	90	84	76	78	75	81
78	87	89	91	79	82	91	85	89	90

1. Buatlah tabel distribusi frekuensi tunggalnya dengan menyertakan frekuensi relatif, frekuensi kumulatif, dan persentase frekuensi kumulatifnya!
2. Buatlah tabel distribusi frekuensi kelompoknya dengan menyertakan frekuensi relatif, frekuensi kumulatif, dan persentase frekuensi kumulatifnya!
3. Dari soal point (b) tentukan batas kelas bawah dan batas kelas atas, tepi kelas bawah dan tepi kelas atas, serta nilai tengahnya!

DAFTAR PUSTAKA

Arifin, M. H. (2014). Konsep-konsep Dasar statistika. *Jakarta: Universitas Terbuka*.

Bluman, A. G. (2018). *Elementary Statistics: A Step By Step Approach* (Tenth Edit). McGraw-Hill Education.

Janah, M., & Kartini, A. Y. (2022). Penerapan Metode Regresi Linier Berganda Pada Kasus Balita Gizi Buruk Di Kabupaten Bojonegoro. *Jurnal Statistika Dan Komputasi*, *1*(2), 74–82.

Mendenhall, W. (2013). *Introduction to Probability and Statistics, 3rd*. Nelson Educatio.

Mendenhall, W., Beaver, R. J., & Beaver, B. M. (2012). *Introduction to probability and statistics*. Cengage Learning.

Mendenhall, W., Beaver, R. J., & Beaver, B. M. (2020). *Introduction to probability and statistics* (Fifteenth). Cengage Learning, Inc.

Nugroho, S. (2007). *Dasar-dasar Metode Statistika*. Grasindo.

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

Santoso, S. (2016). *Panduan Lengkap SPSS Versi 23*. Gramedia Direct.

Steel, R. G. D., & Torrie, J. H. (1976). Introduction to statistics. *(No Title)*.

Sudjana. (2001). *Metode Statistika*. Tarsito.

Sudjana. (2015). *Metode Statistika*. PT Tarsito Bandung.

Walpole, R. E. (1995). *Pengantar statistika*.

Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2016). Probability & Statistics for Engineers & Scientists. In *Journal of the American Statistical Association* (Ninth Edit). Pearson Education Limited.

Yitnosumarto, S. (1990). Dasar-Dasar Statistika. *C. V Rajawali: Jakarta*.

BAB 3

Ukuran Tendensi Pusat

A. Tujuan Pembelajaran

Kemampuan akhir yang ingin dicapai setelah mempelajari bab ini yaitu secara umum mahasiswa diharapkan mampu menghitung nilai tendensi pusat. Adapun tujuan khusus yang ingin dicapai adalah mahasiswa diharapkan mampu :

1. Menghitung rata-rata (mean)
2. Menghitung nilai Tengah (median)
3. Menentukan nilai yang paling sering muncul (modus)
4. Menentukan nilai tendensi pusat mana yang paling cocok digunakan untuk data tertentu
5. Menghitung rata-rata (mean) pada data berkelompok
6. Menentukan nilai Tengah (median) pada data berkelompok
7. Menentukan nilai yang paling sering muncul (modus) pada data berkelompok

B. Pendahuluan

Ukuran pemusatan data secara umum digunakan untuk merangkum data dan masuk dalam kategori statistika deskriptif. Beberapa istilah dalam ukuran pemusatan data yang lazim digunakan yakni rata-rata, nilai tengah dan juga modus. Disini akan dibahas Langkah demi Langkah bagaimana menghitung nilai-nilai pemusatan data yaitu nilai rata-rata (selanjutnya akan digunakan istilah **mean**), nilai Tengah (selanjutnya akan digunakan istilah **median**), dan nilai yang paling sering muncul (selanjutnya akan digunakan istilah **modus**).

C. Mean

Sesuai dengan Namanya mean adalah nilai rata-rata dari data yang dihitung secara keseluruhan (Sudjana, 2001). Istilah lain yang sering digunakan untuk menggambarkan rata-rata adalah average. Mean dapat dihitung dengan menggunakan persamaan sebagai berikut.

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

Keterangan $\bar{x}$ menunjukkan nilai rata-rata, $\sum x_i$ menunjukkan penjumlahan nilai x dan n adalah jumlah observasi yang akan dihitung nilainya.

Perhatikan contoh tinggi badan 10 orang siswa (dalam sentimeter) yang menjadi sampel suatu penelitian sebagai berikut.

Contoh 3.1

Hitunglah rata-rata dari data tinggi badan siswa berikut ini.

152 165 168 158 160 171 158 165 160 155

Berdasarkan contoh tersebut didapatkan tinggi rata-rata 10 orang siswa (dalam sentimeter) tersebut adalah sebagai berikut.

$$\bar{x} = \frac{\sum x_i}{n} = \frac{152 + 165 + \cdots + 155}{10} = 161{,}2$$

Untuk data yang banyak jumlahnya, tentu akan lebih rumit jika harus dijumlahkan satu per satu. Rata-rata juga dapat dihitung dari tabel distribusi frekuensi. Sehingga rumus rata-rata dapat disesuaikan menjadi sebagai berikut.

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

Untuk memudahkan perhitungan, data tinggi badan pada Contoh 3.1 dimasukkan dalam tabel distribusi frekuensi, kemudian akan didapatkan hasil sebagaimana ditunjukkan pada Tabel 3.1 berikut.

Tabel 3.1 Tabel Distribusi Frekuensi

x_i	f_i
152	1
155	1
158	2
160	2
165	2
168	1
171	1

Berdasarkan Tabel 3.1 didapatkan nilai rata-rata sebagai berikut :

$$\bar{x} = \frac{\sum x_i f_i}{n} = \frac{152(1) + 155(1) + \cdots + 171(1)}{10} = 161{,}2$$

Dapat pula disajikan dalam bentuk tabel 3.2 berikut.

Tabel 3.2 Tabel bantu mencari rata-rata

x_i	f_i	$x_i f_i$
152	1	152
155	1	155
158	2	316
160	2	320
165	2	330
168	1	168
171	1	171
Jumlah		1612

Berdasarkan tabel 3.2 didapatkan nilai rata-rata sebesar

$$\bar{x} = \frac{\sum x_i f_i}{n} = \frac{1612}{10} = 161{,}2$$

Pada tabel 3.2 nilai x_1=152, hanya ada satu siswa yang memiliki tinggi badan 152 cm, maka f_1=1. Untuk siswa dengan tinggi badan 158 cm, karena ada 2 siswa yang memiliki tinggi badan 158 cm, maka nilai f_3=2 dan seterusnya. Jika kolom f_i dijumlahkan ke bawah, nilainya akan sama dengan 10, sama dengan nilai n atau jumlah data. Penjumlahan ke bawah ini salah satunya bisa dilakukan untuk melihat apakah tabel distribusi yang telah dibuat sudah sesuai atau belum.

Rumus yang ditunjukkan diatas merupakan rumus yang digunakan untuk menghitung rata-rata sampel. Untuk menghitung nilai rata-rata dari populasi pada dasarnya adalah sama, tetapi menggunakan notasi yang sedikit berbeda. Secara umum, rumus untuk menyatakan nilai rata-rata dari suatu sampel adalah sebagai berikut.

$$\mu = \frac{\sum x_i}{N}$$

Dimana μ adalah nilai rata-rata dari populasi dan N adalah jumlah populasi. Untuk data yang memiliki frekuensi lebih dari satu, maka rumusnya juga dapat disesuaikan menjadi seperti berikut :

$$\mu = \frac{\sum x_i f_i}{N}$$

D. Median

Median adalah nilai yang terletak di bagian Tengah dari data yang diobservasi. Yang perlu diperhatikan dan perlu diingat dalam mencari nilai Tengah suatu data adalah data harus diurutkan dari yang terkecil hingga terbesar. Untuk jumlah data yang ganjil, nilai tengahnya akan langsung dapat ditemukan. Sementara untuk data yang jumlahnya genap, nilai tengahnya adalah nilai rata-rata dari dua data yang berada di Tengah (Nugroho, 2008).

Untuk mengetahui letak median, dapat dihitung dengan rumus sebagai berikut :

$$\text{Letak median} = \frac{n}{2}$$

Contoh 3.2

Temukan mediannya!

5 7 6 5 4 7 8 4 7

Jawab

Setelah terlebih dahulu diurutkan, maka datanya menjadi sebagai berikut :

4	4	5	5	6	7	7	7	8

Letak median $= \frac{9}{2} = 4{,}5$, sehingga nilai tengahnya adalah data setelah data ke-empat, yaitu data ke-lima, sehingga median = 6.

Contoh berikut menggambarkan cara menghitung median pada jumlah data yang genap.

Contoh 3.3

Temukan mediannya!

5	7	6	5	4	7	8	4	7	7

Jawab

Setelah terlebih dahulu datanya diurutkan, didapatkan data sebagai berikut.

4	4	5	5	6	7	7	7	7	8

Letak median $= \frac{10}{2} = 5$, sehingga nilai tengahnya adalah data setelah data antara data ke-lima dan data ke-enam, sehingga median = $\frac{6+7}{2} = 6{,}5$.

Nilai Tengah juga bisa ditemukan dengan menggunakan distribusi frekuensi. Perhatikan contoh dibawah ini.

Contoh 3.4

Tentukan median dari data di bawah ini :

x_i	f_i
4	2
5	2
6	1 (median)
7	3
8	1

Jawab

Letak median $= \frac{9}{2} = 4{,}5$, sehingga nilai tengahnya adalah data setelah data ke-empat yaitu data kelima. Apabila dilihat dari tabel data kelima berada pada baris ketiga yang berisi *x=6*, sehingga median = 6.

E. Modus

Sebagaimana namanya, modus adalah nilai yang memiliki frekuensi paling banyak (Arifin, 2014).

Contoh 3.5

Tentukan modus dari data di bawah ini

5 7 6 5 4 7 8 4 7 9

Jawab

Untuk lebih memudahkan, buatlah tabel distribusi frekuensi seperti di bawah ini

x_i	f_i
4	2
5	2
6	1
7	3
8	1
9	1

Dari tabel tersebut, nilai yang paling sering muncul adalah nilai yang muncul tiga kali, yaitu 7. Jadi modus = 7

Tidak seperti rata-rata dan nilai Tengah yang pasti hanya memiliki satu nilai, modus bisa tidak ada atau lebih dari satu. Perhatikan contoh berikut ini.

Contoh 3.6

Tentukan modus dari data di bawah ini

5 7 6 5 4 7 8 4 6 8

Jawab

Untuk lebih memudahkan, dapat dibuat tabel distribusi frekuensi seperti di bawah ini

x_i	f_i
4	2
5	2

6	2
7	2
8	2

Karena seluruh data memiliki frekuensi muncul yang sama, yaitu 2 kali, maka pada data diatas tidak memiliki modus

Contoh 3.7

Tentukan modus dari data di bawah ini

5 7 6 5 5 7 7 4 6 8

Jawab

Untuk lebih mudah, dapat disusun tabel distribusi frekuensi seperti di bawah ini

x_i	f_i
4	1
5	3
6	2
7	3
8	1

Karena angka 5 dan 7 sama-sama muncul tiga kali, sedangkan angka-angka yang lain muncul kurang dari itu, maka modusnya adalah 5 dan 7.

F. Perbandingan antara Mean, Median dan Modus

Sebagaimana yang telah disajikan pada sebelumnya, ketiga nilai mean, median dan modus mengacu pada ukuran tendensi pusat. Namun nilai mana yang paling baik merefleksikan ukuran tendensi pusat? Jawabannya tentu akan sangat tergantung pada sebaran datanya.

Dalam kehidupan sehari-hari mungkin lebih terbiasa dengan istilah mean dibandingkan dengan median dan modus. Akan tetapi mean akan menyebabkan interpretasi yang tidak terlalu tepat apabila terdapat data ekstrim, baik itu ekstrim kecil maupun ekstrim besar. Hal ini karena rata-rata sangat dipengaruhi oleh nilai masing-masing data termasuk nilai ekstrim (Walpole et al., 2016). Perhatikan contoh berikut ini.

Contoh 3.8

Berikut ini adalah harga sepuluh unit rumah yang ada di suatu daerah tertentu dalam jutaan rupiah

250	350	350	400	450
500	600	620	700	1200

Hitung nilai mean, median dan modusnya

Jawab

Cara mencari nilai mean, median dan modus telah dijelaskan sebelumnya. Berikut adalah hasilnya

Mean = 572

Median = 475

Modus = 350

Berdasarkan hasil di atas, jelas mean memiliki nilai yang jauh lebih tinggi dibandingkan median. Hal ini disebabkan adanya nilai ekstrem yaitu 1200 yang jauh melebihi harga-harga rumah di sekitarnya. Mean sangat dipengaruhi oleh nilai masing-masing data yang diobservasi, sedangkan median adalah nilai Tengah dari data sehingga nilainya tidak dipengaruhi oleh satu atau dua nilai ekstrem. Bandingkan dengan contoh berikut ini.

Contoh 3.9

Berikut ini adalah harga 10 unit rumah yang ada di suatu daerah tertentu dalam jutaan rupiah

250	350	350	400	450
500	600	620	700	800

Hitung nilai mean, median dan modusnya

Jawab

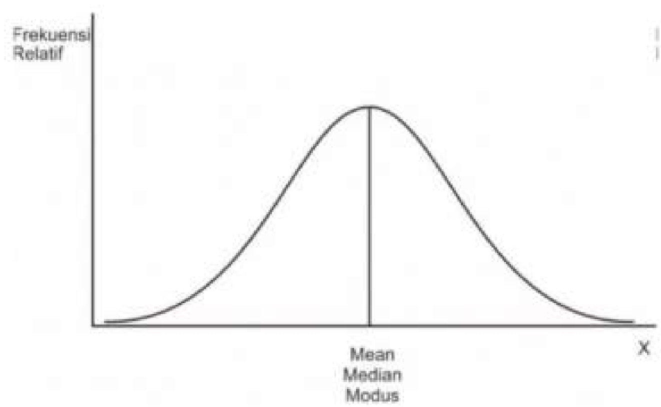
Cara mencari nilai mean, median dan modus telah dijelaskan sebelumnya. Berikut adalah hasilnya

Mean = 497

Median = 475

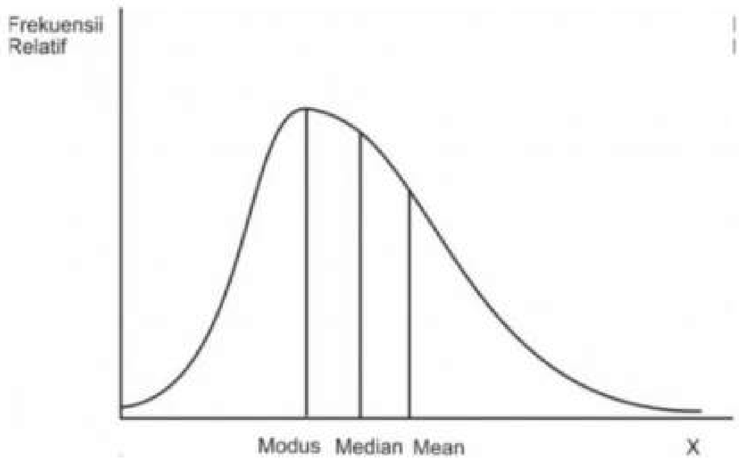
Modus = 350

Apabila data ekstrem berubah dari 1200 menjadi 800, hal ini akan mempengaruhi mean. Hanya dengan mengubah satu data ekstrem saja telah mengubah nilai mean dari 572 menjadi 497, sedangkan mediannya tidak berubah. Pada keadaan Ketika data terdistribusi secara simetris dan membentuk kurva distribusi normal yang sempurna, maka ketiga nilai mean, median dan modus akan memiliki nilai yang tepat sama (identik) sebagaimana ditunjukkan pada Gambar 3.1 berikut.



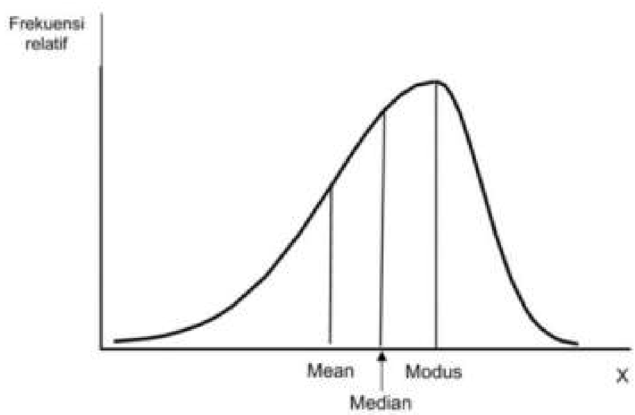
Gambar 3.1 Kurva Distribusi Normal yang Simetris

Namun adakalanya data cenderung mengelompok ke satu sisi tertentu yang menyebabkan kurva distribusi data berbentuk menceng (skewed). Pada keadaan kurva menceng ini, mean, median dan modus tentu tidak akan memiliki nilai yang identik seperti Gambar 3.1. Ada dua kemungkinan bentuk kemencengan kurva distribusi, yaitu menceng ke kanan (*skewed to the right*) dan menceng ke kiri (*skewed to the left*). Kurva distribusi disebut menceng ke kanan apabila sebagian besar data terkonsentrasi di sebelah kiri dan terdapat beberapa data ekstrem besar di sebelah kanan yang menyebabkan kurva memiliki ekor di sebelah kanan. Seperti yang telah dibahas pada Contoh 3.8 pada kasus data dengan nilai ekstrem, meannya akan lebih tinggi dari median. Modus, karena berada pada data dengan frekuensi tertinggi, akan memiliki nilai paling rendah karena data mengelompok pada nilai yang kecil. Ilustrasi distribusi menceng ke kanan ditunjukkan pada Gambar 3.2 berikut.



Gambar 3.2 Distribusi Normal yang Menceng ke Kanan

Sebaliknya apabila data terkonsentrasi di sebelah kanan dan terdapat beberapa data ekstrem rendah (di sebelah kiri), kurva distribusi akan menceng ke kiri (skewed to the left). Pada kasus ini, mean akan lebih rendah dari median dan median akan lebih rendah dari modus sebagaimana ditunjukkan pada Gambar 3.3.



Gambar 3.3 Kurva Distribusi Normal yang Menceng ke Kiri

Kembali ke pertanyaan awal mengenai ukuran tendensi pusat mana yang terbaik? Secara umum, mean atau rata-rata adalah ukuran yang baik dan sangat umum digunakan untuk mengambil kesimpulan tentang data. Namun, ketika data yang dimiliki cenderung terkonsentrasi di suatu titik tertentu, akan sangat diajurkan untuk menggunakan median. Modus bukan ukuran yang baik untuk menggambarkan data, melainkan saat-saat tertentu ketika dibutuhkan pengambilan keputusan secara cepat, modus dapat digunakan.

G. Data Berkelompok

Data adalah sekelompok informasi yang memiliki ciri-ciri tertentu. Data bisa disajikan secara sendiri-sendiri dan dapat juga disajikan secara berkelompok di dalam kelas-kelas interval. Pengelompokan data dilakukan dengan mempertimbangkan kesamaan property (Yitnosumarto, 1990). Misalkan jika akan mengelompokkan data kependudukan ada berbagai macam pengelompokan yang biasanya digunakan. Apabila tujuannya adalah kepentingan perhitungan Angkatan kerja, pengelompokan yang digunakan adalah kelompok umur dan status pekerjaan. Apabila misalnya untuk kepentingan perhitungan indeks ketimpangan pendapatan, pengelompokan yang digunakan adalah berdasarkan besarnya penghasilan.

Untuk setiap kelas interval, terdapat jarak kelas interval yang menggambarkan jarak antara nilai terkecil (batas bawah) dan nilai terbesar (batas atas) dari kelas interval. Kelas interval dilambangkan dengan notasi *i*. Bagaimana cara menghitung kelas interval? Cara paling mudah adalah mengurutkan anggota dari setiap kelas, lalu menghitung jumlah anggotanya.

H. Mean untuk Data Berkelompok

Rumus yang digunakan untuk mean untuk data berkelompok tidak berbeda dengan yang digunakan pada data Tunggal. Bedanya adalah apabila pada data berkelompok nilai x yang dipakai adalah nilai Tengah kelas interval. Untuk mengingat Kembali, rumus yang digunakan untuk mencari mean pada data Tunggal adalah sebagai berikut:

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

Keterangan $\bar{x}$ menunjukkan nilai rata-rata, $\sum x_i f_i$ menunjukkan penjumlahan nilai x , f dan n adalah jumlah observasi yang akan dihitung nilainya.

Untuk data berkelompok notasi x diganti dengan X (kapital x), rumus yang digunakan untuk mencari mean adalah sebagai berikut (Nugroho, 2007):

$$\bar{x} = \frac{\sum X_i f_i}{n}$$

Keterangan $\bar{x}$ menunjukkan nilai rata-rata, $\sum X_i f_i$ menunjukkan penjumlahan dari nilai Tengah dari kelas interval dikalikan frekuensi kelasnya , dan n adalah jumlah observasi yang akan dihitung nilainya.

Lalu, bagaimana cara mencari nilai Tengah kelas interval? Nilai Tengah kelas interval adalah nilai yang letaknya paling Tengah dari kelas interval. Untuk kelas interval yang jumlahnya ganjil, nilai Tengah intervalnya bisa langsung ditemukan, sedangkan untuk kelas interval yang jumlahnya genap, nilai Tengah kelas intervalnya adalah nilai rata-rata dari dua nilai yang berada di Tengah.

Contoh 3.10

Berikut ini adalah data berat badan dari 30 siswa SD. Tentukan meannya!

Kelompok berat badan	**Frekuensi**
20 – 24	7
25 – 29	15
30 – 34	8

Jawab

Kelompok berat badan	X_i **(Nilai Tengah)**	f_i **(Frekuensi)**	$X_i f_i$
20 – 24	$\frac{20+24}{2} = 22$	7	154
25 – 29	27	15	405
30 – 34	32	8	256
Jumlah			815

$$\bar{x} = \frac{\sum X_i f_i}{n} = \frac{815}{30} = 27{,}17$$

Sehingga, mean untuk data berkelompok diatas adalah sebesar 27,17

I. Median untuk Data Berkelompok

Median adalah nilai yang terletak di posisi tengah data. Pada data berkelompok, median dapat dinyatakan dengan menyebutkan kelas intervalnya atau nilai Tengah kelas interval yang berada di tengah-tengah data (Mendenhall et al., 2012). Perhatikan contoh berikut :

Contoh 3.11

Berikut ini adalah berat badan dari 30 siswa SD. Tentukan mediannya!

Kelompok berat badan	Frekuensi
20 – 24	7
25 – 29	15
30 – 34	8

Jawab

Kelompok berat badan	Frekuensi (f_i)	Nilai Tengah (X_i)
20 – 24	7	22
25 – 29	15	27
30 – 34	8	32

Karena jumlah data adalah 30, maka median akan berada pada data ke-15 dan 16. Sehingga didapatkan nilai median = 25 – 29 atau median = 27.

J. Modus untuk Data Berkelompok

Modus adalah nilai yang paling sering muncul atau paling tinggi frekuensinya. Seperti halnya median, untuk data berkelompok, modus dapat dinyatakan dalam dua bentuk yaitu dengan menyebutkan kelas intervalnya atau nilai Tengah dari kelas interval yang paling besar frekuensinya (Mendenhall et al., 2020). Perhatikan contoh di bawah ini.

Contoh 3.12

Berikut ini adalah data berat badan dari 30 siswa SD

Kelompok berat badan	Frekuensi
20 – 24	7

25 – 29	15
30 – 34	8

Jawab
Kelompok atau kelas interval yang paling sering muncul adalah kelompok kedua, yaitu kelas interval siswa dengan berat badan 25 – 29 kilogram, sehingga modus berada di kelas interval tersebut.

Kelompok berat badan	Frekuensi (f_i)	Nilai Tengah (X_i)
20 – 24	7	22
25 – 29	15	27
30 – 34	8	32

Untuk data di atas, karena kelas interval yang paling tinggi frekuensinya adalah kelas interval 25 – 29, maka modusnya adalah sebagai berikut:
Modus = 25 – 29 atau modus = 27.

K. Rata-rata Terbobot

Rata-rata terboboti digunakan apabila ingin memberikan bobot yang berbeda terhadap beberapa hal yang diobservasi (Steel & Torrie, 1976). Rata-rata terbobot dapat dihitung dengan rumus sebagai berikut:

$$\bar{x}_w = \frac{\sum x_i w_i}{\sum w_i}$$

Keterangan $\bar{x}_w$ menunjukkan rata-rata terbobot, x_i menunjukkan data yang akan dicari rata-ratanya, dan w_i adalah bobot dari masing-masing data dimana $\sum w_i = 1$

Contohnya adalah Ketika dosen memberikan bobot yang berbeda terhadap nilai akhir mahasiswa. Biasanya ujian akhir semester dibobot lebih tinggi dibandingkan ujian Tengah semester dan tugas. Perhatikan contoh berikut.

Contoh 3.13
Seorang dosen menentukan bobot tugas 20%, ujian Tengah semester 30%, dan ujian akhir semester 50%. Berapa nilai rata-rata yang diperoleh mahasiswa tersebut apabila nilai yang diperolehnya sebagai berikut :

Tugas = 80 ; UTS = 90 ; dan UAS = 70

Jawab

$$\bar{x}_w = \frac{\sum x_i w_i}{\sum w_i}$$

$$= \frac{(80 \times 0{,}2) + (90 \times 0{,}3) + (70 \times 0{,}5)}{0{,}2 + 0{,}3 + 0{,}5}$$

$$= \frac{16 + 27 + 35}{1}$$

$$= 78$$

L. Latihan Soal

1. Perhatikan data umur 10 penghuni panti jompo di bawah ini
   75 63 85 80 90 84 69 78 75 81
   Hitunglah :
   a. Mean
   b. Median
   c. Modus
2. Berikut ini adalah nilai ulangan harian Bahasa Inggris siswa kelas IX di SMP X

Nilai	Jumlah Siswa
41 – 50	3
51 – 60	4
61 – 70	7
71 – 80	8
81 – 90	5
91 – 100	3

   Hitunglah :
   a. Mean
   b. Median
   c. Modus

DAFTAR PUSTAKA

Arifin, M. H. (2014). Konsep-konsep Dasar statistika. *Jakarta: Universitas Terbuka*.

Bluman, A. G. (2018). *Elementary Statistics: A Step By Step Approach* (Tenth Edit). McGraw-Hill Education.

Janah, M., & Kartini, A. Y. (2022). Penerapan Metode Regresi Linier Berganda Pada Kasus Balita Gizi Buruk Di Kabupaten Bojonegoro. *Jurnal Statistika Dan Komputasi*, *1*(2), 74–82.

Mendenhall, W. (2013). *Introduction to Probability and Statistics, 3rd*. Nelson Educatio.

Mendenhall, W., Beaver, R. J., & Beaver, B. M. (2012). *Introduction to probability and statistics*. Cengage Learning.

Mendenhall, W., Beaver, R. J., & Beaver, B. M. (2020). *Introduction to probability and statistics* (Fifteenth). Cengage Learning, Inc.

Nugroho, S. (2007). *Dasar-dasar Metode Statistika*. Grasindo.

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

Santoso, S. (2016). *Panduan Lengkap SPSS Versi 23*. Gramedia Direct.

Steel, R. G. D., & Torrie, J. H. (1976). Introduction to statistics. *(No Title)*.

Sudjana. (2001). *Metode Statistika*. Tarsito.

Sudjana. (2015). *Metode Statistika*. PT Tarsito Bandung.

Walpole, R. E. (1995). *Pengantar statistika*.

Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2016). Probability & Statistics for Engineers & Scientists. In *Journal of the American Statistical Association* (Ninth Edit). Pearson Education Limited.

Yitnosumarto, S. (1990). Dasar-Dasar Statistika. *C. V Rajawali: Jakarta*.

BAB 4

Simpangan Data

A. Tujuan Pembelajaran

Kemampuan akhir yang ingin dicapai setelah mempelajari bab ini yaitu Mahasiswa mampu menghitung dan menerapkan simpangan data. Adapun tujuan khusus yang ingin dicapai adalah mahasiswa mampu menghitung hal-hal berikut :

1. Simpangan baku populasi
2. Simpangan baku sampel
3. Koefisien variasi

B. Pendahuluan

Secara umum, nilai tendensi pusat dapa tmenggambarkan secara baik nilai Tengah dari sekumpulan data. Akan tetapi, penggunaan nilai tendensi pusat saja tidak memberikan informasi tentang bagaimana sebaran dari data tersebut dibandingkan dengan nilai rata-ratanya. ukuran yang digunakan untuk mengukur jarak antara sebaran dari data dan nilai rata-ratanya dikenal dengan simpangan.

Ada beberapa metode yang dapat digunakan untuk menghitung simpangan dan yang paling sederhana adalah menghitung rentang nilai (range). Pada bab ini akan membantu mahasiswa untuk mempelajari simpangan. Beberapa materi yang akan disajikan antara lain simpangan baku baik populasi maupun sampel, koefisien variasi, kuantil dan persentil. Perhatikan contoh berikut.

Contoh 4.1

Penjahit A dan B sama-sama menghasilkan rata-rata potong baju setiap harinya. Penjahit A menjahit dua potong baju per hari atau jika digambarkan dari Senin hingga Minggu menghasilkan 2, 2, 2, 2, 2, 2, 2 potong baju, sedangkan penjahit B menghasilkan 3, 5, 2, 1, 1, 2, 0.

Apabila dijumlahkan, keduanya sama-sama menghasilkan 14 potong baju setiap minggunya atau dua potong baju per hari. Namun, kedua orang penjahit ini memiliki Tingkat variabilitas yang sangat berbeda dalam menghasilkan baju setiap harinya. Penjahit A menghasilkan dua potong baju setiap harinya dan penjahit B menghasilkan antara 0 sampai 5 potong baju per hari.

C. Nilai Rentang (*Range*)

Cara paling sederhana untuk mengukur variabilitas suatu data adalah menggunakan rentang nilai (range). Rentang adalah selisih antara nilai data terbesar dan yang terkecil (Walpole et al., 2016). Walaupun sangat mudah dihitung, rentang dinilai tidak akurat dalam menghitung variabilitas data. Hal ini karena rentang hanya memperhatikan perbedaan antara nilai terkecil dan terbesar sehingga perubahan yang terjadi diantara kedua nilai tersebut tidak akan berpengaruh apa-apa terhadp variabilitas. Adapun rumus untuk menghitung rentang adalah
sebagai berikut.

Rentang = nilai tertinggi – nilai terendah

Contoh 4.2

Perhatikan Kembali contoh 4.1 di atas. Hitunglah rentang dari penjahit A dan B!

Jawab

Rentang penjahit A = 2 – 2 =0

Rentang penjahit B = 5 – 0 = 5

Pada contoh 4.2 Penjahit B memiliki nilai rentang 5 karena selisih antara nilai terbesar dan terkecilnya sama dengan 5. Apabila data distribusi hasil jahitan adalah 5, 4, 3, 2, 0, 0, 0, nilai range-nya tetap 5 dan rata-ratanya tetap 2, tetapi variasi hasil jahitan per harinya berbeda.

D. Simpangan Baku Populasi

Ukuran lain yang dapat digunakan untuk menggambarkan variabilitas suatu data adalah menggunakan simpangan baku (standar deviasi). Simpangan baku adalah jarak deviasi atau simpangan dari nilai asli ke nilai rata-rata datanya (Arifin, 2014). Dari contoh 4.1, karena penjahit A menghasilkan jumlah jahitan yang sama setiap harinya dibandingkan dengan nilai rata-ratanya, maka standar deviasinya adalah 0. Penjahit B, karena menghasilkan jumlah jahitan yang berbeda-beda setiap harinya, maka akan menghasilkan nilai simpangan baku yang tentunya lebih besar dari 0.

Rumus yang digunakan untuk menghitung simpangan baku suatu populasi adalah sebagai berikut.

$$\sigma^2 = \frac{\sum(x_i - \mu)^2}{N}$$

Keterangan σ^2 merupakan nilai simpangan dari rata-rata, μ adalah nilai rata-rata dari data dan N adalah jumlah populasi

Contoh 4.3

Perhatikan Kembali contoh 4.1 di atas. Langkah-langkah berikut dapat dilakukan untuk menghitung simpangan baku penjahit B (simpangan baku penjahit A tidak perlu dihitung)

1. Menghitung nilai rata-ratanya (μ)
2. Menghitung $(x_i - \mu)$
3. Menghitung $(x_i - \mu)^2$
4. Menghitung $\sum(x_i - \mu)^2$
5. Menghitung σ^2
6. Menghitung σ

Sehingga didapatkan hasil sebagaimana ditunjukkan pada tabel berikut.

x_i	$(x_i - \mu)$	$(x_i - \mu)^2$
3	1	1
5	3	9
2	0	0
1	-1	1

1	-1	1
2	0	0
0	-2	4
Jumlah		16

$$\sigma^2 = \frac{\sum(x_i - \mu)^2}{N} = \frac{16}{7} = 2{,}29$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{2{,}29} = 1{,}51$$

E. Simpangan Baku Sampel

Apabila data yang digunakan adalah data sampel, maka rumus simpangan baku sampel adalah sebagai berikut.

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

Keterangan s^2 merupakan simpangan baku sampel, $\bar{x}$ adalah rata-rata sampel dan n adalah jumlah sampel. Jika diperhatikan, rumus simpangan baku sampel berbeda dengan rumus simpangan baku populasi. Bukan hanya dari notasi akan tetapi juga pembaginya. Untuk simpangan baku sampel pembaginya adalah jumlah populasi (N), sementara untuk simpangan baku sampel pembaginya adalah jumlah sampel dikurangi 1 $(n-1)$. Pembagi $(n-1)$ digunakan pada perhitungan simpangan baku sampel karena sampel dianggap lebih tidak variative dibandingkan dengan populasi sehingga diharapkan hasilnya tidak berlebihan (*overestimate*) atau kurang (*underestimate*) dibandingkan dengan simpangan baku populasi (Yitnosumarto, 1990).

Contoh 4.4

Anggaplah bahwa contoh 4.2 menggambarkan nilai sampel dari penjahit A dan B dalam satu tahun. Dari contoh diatas, dapat digunakan untuk menghitung simpangan baku sampel. Gunakan rumus untuk menghitung simpangan baku sampel

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
3	1	1
5	3	9

2	0	0
1	-1	1
1	-1	1
2	0	0
0	-2	4
Jumlah		16

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{16}{6} = 2{,}67$$

$$s = \sqrt{s^2} = \sqrt{2{,}67} = 1{,}63$$

F. Membandingkan Dua Simpangan Baku dan Apa Artinya

Simpangan memberikan informasi tentang bagaimana variasi (jarak sebaran) data terhadap nilai rata-ratanya. Nilai simpangan baku yang lebih kecil menunjukkan bahwa data menyebar lebih dekat dengan nilai rata-ratanya. Sementara itu, simpangan baku yang lebih besar berarti bahwa data menyebar lebih jauh ke rata-ratanya. Karena simpangan baku pada dasarnya adalah jarak nilai dengan rata-ratanya, secara teoritis simpangan baku tidak bia bernilai negatif. Akan tetapi pada kasus yang sanga tjarang terjadi, simpangan dapat bernilai nol. Simpangan baku akan bernilai nol apabila semua data nilainya sama (Mendenhall et al., 2020). Kembali ke contoh 4.1, terdapat dua orang penjahit A dan B. Keduanya memproduksi jumlah baju yang sama, yaitu dua potong, sedangkan penjahit B menjahit jumlah baju yang berbeda-beda maka penjahit A memiliki simpangan baku sama dengan nol. Karena simpangan mengukur jarak antara data dan rata-ratanya, maka secara umum peneliti menganggap bahwa data dengan simpangan yang besar sebagai sesuatu yang "jelek" dan berusaha untuk dihindari dengan cara mengontrol data.

Dua atau lebih data dapat dibandingkan melalui simpangan baku apabila persoalan yang diukur adalah sama. Misalnya membandingkan simpangan baku antara nilai rata-rata ujian matematika antara kelas A, B, dan C jika diberikan soal yang sama atau paling tidak berasal dari materi yang sama. Tentu saja tidak bisa membandingkan simpangan baku dari penjualan sepeda motor dengan penjualan mobil karena harga keduanya sangat jauh berbeda walaupun sama-sama merupakan kendaraan bermotor. Perhatikan contoh berikut.

Contoh 4.5

Rata-rata nilai ujian Tengah semester mata Pelajaran statistik 10 siswa kelas 12-A dan 12-B di SMA X menunjukkan nilai rata-rata yang sama, yaitu 60. Sebaran nilai ujian dapat dilihat dari tabel berikut.

	12-A	12-B	$(x_{iA}-\mu)^2$	$(x_{iB}-\mu)^2$
1	35	53	625	49
2	40	54	400	36
3	45	64	225	16
4	86	75	676	225
5	25	47	1225	169
6	92	63	1024	9
7	62	48	4	144
8	52	65	64	25
9	88	80	784	400
10	75	51	225	81
Jumlah	**600**	**600**	**5252**	**1154**

Dari tabel diatas dapat dihitung simpangan baku dari masing-masing kelas sebagai berikut.

$$\sigma_A = \sqrt{\frac{5252}{10}} = 22{,}92$$

$$\sigma_B = \sqrt{\frac{1154}{10}} = 10{,}74$$

Dari kedua kelas tersebut didapatkan bahwa simpangan baku pada kelas 12-A lebih besar daripada kelas 12-B. hal ini berarti bahwa nilai ujian Tengah semester siswa di kelas 12-B lebih merata (kurang bervariasi) dibandingkan dengan siswa kelas 12-A.

G. Koefisien Variasi

Kelemahan dari pengukuran variasi menggunakan simpangan baku adalah tidak dapat membandingkan simpangan baku dari dua atau lebih set data yang mempunyai satuan yang berbeda. Koefisien variasi menghilangkan kelemahan ini karena unit yang digunakan adalah persentase sehingga dua atau lebih set data dapat dibandingkan dengan menggunakan koefisien variasi.

Koefisien variasi adalah perbandingan antara simpangan baku dengan rata-ratanya dalam bentuk persentase sehingga secara matematis dapat dirumuskan sebagai berikut (Walpole, 1995).

Koefisien variasi untuk data populasi

$$CV_P = \frac{\sigma}{\mu} \times 100\%$$

Koefisien variasi untuk data sampel

$$CV_S = \frac{s}{\bar{x}} \times 100\%$$

Contoh 4.6

Perhatikan Kembali data pada contoh 4.3, sehingga didapatkan nilai koefisien variasi untuk data populasi sebagai berikut.

$$CV_P = \frac{\sigma}{\mu} \times 100\% = \frac{1{,}51}{2} \times 100\% = 75{,}5\%$$

H. Latihan Soal

1. Berikut ini adalah data penjualan sepeda di Toko Sepeda Makmur pada tahun 2022

Bulan	**Penjualan**	$(x_i - \mu)^2$
Januari	25	
Februari	20	
Maret	12	
April	10	
Mei	12	
Juni	18	

Juli	28	
Agustus	30	
September	18	
Oktober	19	
November	23	
Desember	32	

Isilah kolom $(x_i - \mu)^2$ dan hitunglah simpangan bakunya (gunakan rumus simpangan baku populasi)

DAFTAR PUSTAKA

Arifin, M. H. (2014). Konsep-konsep Dasar statistika. *Jakarta: Universitas Terbuka*.

Bluman, A. G. (2018). *Elementary Statistics: A Step By Step Approach* (Tenth Edit). McGraw-Hill Education.

Janah, M., & Kartini, A. Y. (2022). Penerapan Metode Regresi Linier Berganda Pada Kasus Balita Gizi Buruk Di Kabupaten Bojonegoro. *Jurnal Statistika Dan Komputasi*, *1*(2), 74–82.

Mendenhall, W. (2013). *Introduction to Probability and Statistics, 3rd*. Nelson Educatio.

Mendenhall, W., Beaver, R. J., & Beaver, B. M. (2012). *Introduction to probability and statistics*. Cengage Learning.

Mendenhall, W., Beaver, R. J., & Beaver, B. M. (2020). *Introduction to probability and statistics* (Fifteenth). Cengage Learning, Inc.

Nugroho, S. (2007). *Dasar-dasar Metode Statistika*. Grasindo.

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

Santoso, S. (2016). *Panduan Lengkap SPSS Versi 23*. Gramedia Direct.

Steel, R. G. D., & Torrie, J. H. (1976). Introduction to statistics. *(No Title)*.

Sudjana. (2001). *Metode Statistika*. Tarsito.

Sudjana. (2015). *Metode Statistika*. PT Tarsito Bandung.

Walpole, R. E. (1995). *Pengantar statistika*.

Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2016). Probability & Statistics for Engineers & Scientists. In *Journal of the American Statistical Association* (Ninth Edit). Pearson Education Limited.

Yitnosumarto, S. (1990). Dasar-Dasar Statistika. *C. V Rajawali: Jakarta*.

BAB 5

Konsep Dasar Peubah Acak, Peluang dan Sebaran Peluang

A. Tujuan Pembelajaran

Kemampuan akhir yang ingin dicapai setelah mempelajari bab ini yaitu Mahasiswa mampu menerapkan konsep teori peubah acak, peluang dan sebaran peluang:

1. Menjelaskan peubah acak, peluang dan sebaran peluang dari distribusi binomial.
2. Menerapkan peubah acak, peluang dan seberan peluang dari distribusi binomial dalam studi kasus.
3. Menjelaskan distribusi normal dan daerah dibawah kurva normal.
4. Menerapkan distribusi normal dalam kehidupan sehari-hari.

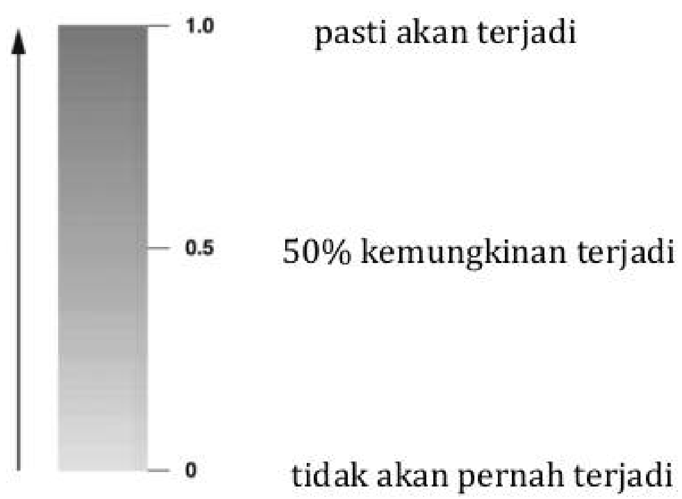
B. Pendahuluan

Dalam kehidupan sehari-hari sering terdengar istilah "kemungkinan" untuk kejadian-kejadian yang akan terjadi walaupun informasi lain mengenai kejadian tersebut tidak 100 persen bisa dipastikan. Dalam kehidupan sehari-hari sering meramalkan klub sepak bola mana yang akan menang dalam liga Inggris, pembalap mana yang akan menang dalam Formula Satu, hari ini akan hujan atau tidak, dan sebagainya. Ilmu statistik bisa digunakan untuk menghitung kemungkinan-kemungkinan satu atau lebih kejadian yang akan terjadi.

C. Probabillitas

"Saya mungkin akan pulang lebih awal", "Hari ini mungkin akan turun hujan", "Peluang kami untuk mendapatkan kontrak sangat kecil", "Seberapa besar peluang pasien untuk bertahan hidup selama enam bulan ke depan", "Seberapa besar kemungkinan kami dapat mendapatkan tempat duduk di penerbangan berikutnya". Kalimat seperti ini biasa digunakan dalam kehidupan nyata. Probabilitas atau kemungkinan atau peluang terjadinya digunakan dalam kasus di mana terdapat ketidakpastian. Jika sebuah bola dilempar dari ketinggian dan tidak ada yang bisa menghentikan bola ditengah jalan, pasti bola tersebut akan jatuh ke tanah. Tidak ada ketidakpastian di sini. Tapi bayangkan sebuah anak panah yang dilempar ke papan. Ini mungkin menyerang dimana saja. Pemain yang berpengalaman lebih mungkin mencapai target dibandingkan dengan pemain yang tidak berpengalaman. Dalam kasus anak panah yang dilempar ke papan ada sebuah kemungkinan, sedangkan dalam kasus pelemparan bola yang ada hanyalah kepastian.

Dengan demikian Probabilitas dapat didefinisikan sebagai sebuah konsep yang digunakan untuk situasi yang tidak memiliki hasil tertentu. Hal ini bisa berupa melempar koin, mencapai tujuan tepat waktu, lulus ujian, mendapatkan pekerjaan, menulis kode dengan benar, dan seterusnya. Probabilitas adalah teknik matematika untuk memprediksi hasil yang tidak pasti serta memprediksi seberapa besar kemungkinan peristiwa tertentu akan terjadi (Walpole, 1995). Probabilitas diukur pada skala 0 hingga 1,0 seperti yang ditunjukkan pada Gambar 4.1.



Gambar 4.1 Skala probabilitas.

Misalnya, ketika seseorang melempar koin secara adil, terdapat peluang 50% untuk memperoleh gambar. Perhatikan bahwa probabilitas biasanya dinyatakan dalam format desimal – 50% menjadi 0,5, 10% menjadi 0,1, dan 5% menjadi 0,05, misalnya. Oleh karena itu, peluang munculnya gambar pada saat sebuah mata uang logam dilempar secara adil adalah 0,5. Probabilitasnya tidak boleh lebih dari 1,0 dan tidak boleh negatif. Ada berbagai metode untuk menghitung probabilitas untuk berbagai situasi.

C.1. Menghitung Probabilitas (P) Terjadinya Kejadian Tunggal (A)

Misalnya, untuk mencari peluang pelemparan angka enam pada satu pelemparan sebuah dadu yang tidak bias dengan rumus sebagai berikut :

$$\mathrm{P(A)} = \frac{\text{jumlah kejadian yang mungkin terjadi}}{\text{jumlah kemungkinan hasil yang sama kemungkinannya}}$$

$$\mathrm{P(A)} = \frac{\text{jumlah angka enam pada dadu}}{\text{jumlah sisi pada dadu}}$$

$$\mathrm{P(A)} = \frac{1}{6} = 0.167(\text{atau } 16{,}7\%)$$

C.2. Menghitung Probabilitas Kejadian (A) Dan Kejadian (B) Terjadi (Peristiwa Independen)

Misalnya, jika terdapat dua paket kartu yang identik (paket A dan paket B), berapa peluang terambilnya kartu as sekop dari kedua paket tersebut?

Dengan menggunakan rumus : P(A) x P(B)

Didapatkan :

P (paket A) = 1 kartu, dari paket berisi 52 kartu = 1/52 = 0.0192

P (paket B) = 1 kartu, dari paket berisi 52 kartu = 1/52 = 0.0192

$\mathrm{P(A)} \times \mathrm{P(B)} = 0.0192 \times 0.0192 = 0{,}00037$

Ini disebut aturan perkalian atau hukum perkalian probabilitas. Dalam contoh ini, kejadian A dan B tidak bergantung satu sama lain. Artinya peristiwa yang satu terjadi tanpa menghiraukan peristiwa yang lain, dan tidak ada hasil yang berhubungan dengan peristiwa yang lain. Terkadang probabilitas bersifat kondisional, artinya satu probabilitas bergantung pada kejadian lain.

C.3. Menghitung Probabilitas Kejadian (A) Dan Kejadian (B) Terjadi (Peristiwa Kondisi)

Berapakah peluang terambilnya kartu as sekop dan ratu klab secara berurutan dari satu pak kartu?

Rumus : $P(A) \times P(B|A)$

Dimana $P(B|A)$ adalah peluang kejadian B dengan syarat A

Telah diketahui bahwa peluang terambilnya as sekop dari sekumpulan 52 kartu adalah 1/52 = 0,0192, jadi P(A) = 0,0192.

Peluang untuk mendapatkan ratu klub sekarang sedikit lebih tinggi, karena jumlah kartu yang tersisa di paket berkurang satu, jadi probabilitas $P(B|A)$ sekarang adalah 1/51 = 0,0196.

$$P(A) \times P(B|A) = \left({}^{1}/_{52}\right) \times \left({}^{1}/_{51}\right) = 0.0192 \times 0.0196 = 0.0004$$

Peristiwa bisa saling eksklusif. Artinya suatu peristiwa menghalangi terjadinya peristiwa lain. Misalnya, melempar dadu satu kali akan menghasilkan angka satu, atau dua, atau tiga, atau empat, atau lima, atau enam, namun hanya satu angka yang dapat diperoleh. Oleh karena itu, membuang lima aturan keluar dari nomor lainnya. Dalam kasus seperti ini, aturan penjumlahan digunakan untuk mencari peluang terjadinya suatu peristiwa.

C.4. Menghitung Probabilitas Terjadinya Kejadian (A) Atau Kejadian (B) Terjadi (Dimana Kejadian Saling Eksklusif)

Misalnya, berapa peluang pelemparan dadu berjumlah enam atau lima?

Rumus: $P(A) + P(B)$

$P(A) = 0.1667$

$P(B) = 0.1667$

$P(A) + P(B) = 0.1667 + 0.1667 = 0.333(\text{atau } 33.3\%)$

Ini disebut aturan penjumlahan atau hukum penjumlahan probabilitas.

C.5. Menghitung Probabilitas Terjadinya Kejadian (A) Atau Kejadian (B) Terjadi (Dimana Kejadian Tidak Saling Eksklusif)

Misalkan penelitian lokal menemukan bahwa 90% orang berusia di atas 60 tahun di Epitown menderita setidaknya satu kali flu biasa selama periode 1 tahun, dan 20% menderita sakit maag setidaknya satu kali. Tentu saja, kedua kejadian tersebut tidak dapat dipisahkan satu sama lain, sehingga satu orang mungkin menderita kedua penyakit tersebut. Berapa probabilitas seseorang yang berusia di atas 60 tahun akan menderita flu biasa, sakit maag, atau kedua-duanya? Asumsikan bahwa flu biasa dan sakit maag terjadi secara independen.

Menggunakan aturan penjumlahan menghasilkan probabilitas 0.9 + 0.2, yaitu sama dengan 1.1. Ini tidak mungkin benar, karena diketahui bahwa probabilitasnya tidak akan pernah lebih dari 1,0.

Dalam situasi ini, digunakan rumus yang berbeda:

$\mathrm{P(A) + P(B) - P\ (keduanya)}$

$\mathrm{P(A) = 0.9\ (flu\ biasa)}$

$\mathrm{P(B) = 0.2\ (maag)}$

$\mathrm{P(keduanya) = 0.9 \times 0.2 = 0.18}$

(dengan asumsi independen)

Sehingga, $\mathrm{P(A) + P(B) - P\ (keduanya) = \ (0.9 + 0.2) - 0.18}$

$= 1.1 - 0.18$

$= 0.92 (\mathrm{atau}\ 92\%)$

Jadi, dalam contoh ini, terdapat kemungkinan sebesar 0.92 (atau 92%) bahwa setiap orang berusia di atas 60 tahun di Epitown akan menderita flu biasa atau sakit maag atau keduanya selama periode 1 tahun.

D. Permutasi dan Kombinasi

Banyak permasalahan probabilitas kompleks yang dapat diselesaikan hanya dengan memahami konsep permutasi dan kombinasi. Dalam konsep kombinasi, terdapat n cara memilih m satuan. Permutasi juga serupa tetapi urutan unit yang dipilih paling penting.

Aturan penghitungan permutasi: Banyaknya cara menyusun n benda berbeda agar diambil r sekaligus adalah $^{n}P_{r} = \frac{n!}{(n-r)!}$ dimana n dan r bilangan bulat angka dan n ≥ r.

Aturan penghitungan kombinasi: Banyaknya kombinasi n benda yang diambil r sekaligus adalah ${}^{n}C_{r} = \frac{n!}{r!(n-r)!}$ dimana n dan r adalah bilangan bulat dan n ≥ r. Cara menentukan banyaknya hasil suatu percobaan adalah sebagai berikut (Mendenhall, 2013).

1. Gunakan aturan perkalian atau diagram pohon, jika percobaan terdiri dari serangkaian tahapan dengan hasil yang berbeda-beda.
2. Gunakan aturan permutasi, ${}^{n}P_{r} = \frac{n!}{(n-r)!}$ jika hasilnya terdiri dari subkelompok terurut yang terdiri dari r item yang diambil dari sekelompok n item.
3. Gunakan aturan kombinasi ${}^{n}C_{r} = \frac{n!}{r!(n-r)!}$ jika hasilnya terdiri dari subgrup tak berurutan yang terdiri dari r item yang diambil dari grup yang terdiri dari n item.

D.1. Teorema Bayes

Teorema ini sangat penting dan penerapannya ada di mana-mana di hampir semua bidang studi. Hal ini karena ia menetapkan dasar matematis untuk inferensi probabilistik. Namanya diambil dari nama ahli matematika Inggris abad kedelapan belas, Thomas Bayes.

E. Random Variable (Variabel Acak)

E.1. *Random Variable* (Variabel Acak)

Variabel acak adalah fungsi yang memberikan bilangan real pada setiap titik dalam ruang sampel S. Variabel tersebut dapat dikategorikan sebagai variabel diskrit dengan nilai yang tepat atau variabel kontinu yang dapat mengambil nilai apa pun dalam suatu interval. Variabel acak dilambangkan dengan huruf kapital seperti X dan Y. Variabel acak adalah proses pemberian nilai pada kasus yang berbeda. Variabel acak mempunyai distribusi probabilitas baik diskrit maupun kontinu. Sementara itu distribusi probabilitas adalah penetapan probabilitas pada nilai spesifik dari variabel acak atau rentang nilai variabel acak. Adapun fitur distribusi probabilitas variabel acak diskrit adalah sebagai berikut :

1. Distribusi probabilitas memiliki probabilitas yang ditetapkan untuk setiap nilai variabel acak.
2. Jumlah semua probabilitas yang ditetapkan harus 1.

Rata-rata dan simpangan baku dari distribusi probabilitas populasi diskrit diperoleh dengan menggunakan rumus berikut:

$$\mu = \sum x \times P(x)$$

Dimana, μ (rata-rata populasi) disebut nilai yang diharapkan dari x (sampel). $\sigma = \sqrt{\sum P(x) \times (x - \mu)^2}$ disebut deviasi standar x (deviasi standar populasi yang mendasari) dimana x adalah nilai variabel acak, P(x) adalah probabilitas variabel tersebut, dan jumlah Σ diambil untuk semua nilai variabel acak.

Tabel 4.1 Menghitung probabilitas munculnya kepala

Jumlah kepala (X)	Jumlah hasil	Probabilitas
0	1	1/10=0.1
1	3	3/10=0.3
2	4	4/10=0.4
3	2	2/10=0.2
Total	10	1

E.2. Variabel Acak Diskrit

Variabel yang hanya dapat mengasumsikan sejumlah nilai riil yang dapat dihitung dan nilai variabel tersebut bergantung pada peluang disebut variabel acak diskrit. Dengan kata lain, fungsi bernilai riil yang didefinisikan pada ruang sampel diskrit disebut variabel acak diskrit. Nama lainnya adalah variabel stokastik diskrit atau variabel peluang diskrit.

Tabel 4.2. Perbedaan antara Variabel Acak Diskrit dan Kontinu

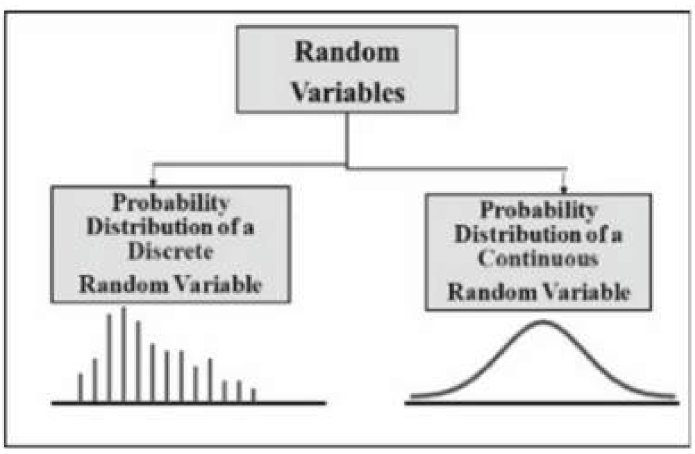
	Diskrit	**Kontinu**
Tipe	Dapat dihitung	Tidak dapat dihitung
Definisi	Probabilitas fungsi massa	Fungsi kepadatan probabilitas (*Probability density function*)
Distribusi	Binomial dan distribusi	Gaussian dan distribusi

	Poisson	exponensial

Contoh 5 jumlah kecelakaan dalam sebulan, jumlah panggilan telepon per satuan waktu, jumlah keberhasilan dalam uji coba “n”.

E.3. Variabel Acak Kontinu

Suatu variabel acak X dikatakan kontinu jika ia dapat mengambil semua nilai yang mungkin (integral maupun pecahan) di antara batas-batas tertentu yang telah ditentukan. Dengan kata lain, suatu variabel acak dikatakan kontinu bila nilai-nilainya yang berbeda tidak dapat disamakan 1-1 dengan himpunan bilangan bulat positif, misalnya umur, tinggi badan, berat badan remaja, dan lain-lain. (Gbr. 3.6) .
Contoh 6 Apakah variabel acak berikut ini diskrit atau kontinu?
Skenario yang berbeda
Skenario 1. Siswa memerlukan waktu untuk membayar denda apabila ia terlambat mengembalikan buku
ke perpustakaan.
Variabel: Variabel acak kontinu.
Alasan: Waktu adalah variabel acak kontinu karena dapat mengambil nilai berapa pun.



Gambar 4.2 Diskrit dan variabel acak kontinu

Expected Value dan Variance dari Random Variable
Umumnya, variabel acak dapat dikarakterisasi dan ditangani secara efektif untuk tujuan praktis dengan mempertimbangkan kuantitas yang

disebut ekspektasinya. Ekspektasi matematis (expected value) adalah nilai rata-rata dari suatu fenomena acak.

Nilai yang diharapkan dari variabel acak diskrit X dengan fungsi massa probabilitas (p.m.f) sebesar $f(x)$ diberikan oleh:$E(X) = \sum_x x \times f(x)$, asalkan deret sebelah kanan benar-benar konvergen (Tabel 3.5).

Nilai yang diharapkan dari variabel acak kontinu X dengan fungsi kepadatan probabilitas (pdf) $f(x)$ diberikan oleh: $E(X) = \int_{-\infty}^{\infty} x \times f(x).dx$, asalkan integral kanannya mutlak konvergen.

Varians dari variabel acak X diberikan oleh $V(x) = E(x^2) - [e(x)]^2$

Contoh Sebuah perusahaan pemasaran komersial swasta ingin menyelidiki apakah pelanggan dipengaruhi untuk membeli produk dengan menonton iklan TV.

Jawab

Kita dapat memperlakukan informasi yang ditampilkan sebagai perkiraan distribusi probabilitas karena peristiwa-peristiwa tersebut saling eksklusif dan jumlah persentasenya adalah 100%. Mari kita cari tahu nilai rata-rata/ekspektasi dan varians distribusinya (Tabel 4.4).

Tabel 4.3 Distribusi dan jenis variabel acak

Jumlah orang yang terpengaruh	1	2	3	4	5
Persentase pembeli	31	30	25	9	5

Tabel 4.4 Rata-rata dan varians pengaruhnya terhadap pembeli

Jumlah orang yang terpengaruh	Persentase pembeli (%)	Expected value	variance
1	31	1×0.31=0.31	1^2×0.31=0.31
2	30	2×0.30=0.6	2^2×0.30=2.4
3	25	3×0.25=0.75	3^2×0.75=6.75
4	9	4×0.09=0.36	4^2×0.36=5.76
5	5	5×0.05=0.25	5^2×0.05=6.25

F. Distribusi Teoritis

Distribusi Teoritis

Distribusi teoretis adalah distribusi yang didasarkan pada rumus matematika dan logika, bukan observasi empiris (skor mentah). Mereka digunakan dalam statistik untuk menentukan probabilitas. Di sini kita mengalokasikan probabilitas ke titik frekuensi/interval kelas yang berbeda, dari total frekuensi. Pengambilan keputusan menjadi tantangan ketika kita menghadapi fakta yang tidak pasti dan acak. Di sinilah konsep distribusi probabilitas memandu kita. Distribusi probabilitas dapat disebut sebagai distribusi logis karena distribusi ini menggabungkan beberapa fakta yang diketahui dan beberapa perkiraan yang masuk akal untuk membangun model matematika yang menghasilkan beberapa solusi realistis di tengah ketidakpastian (Gambar 3.7 dan Tabel 3.8).

Distribusi Binomial

Awalan "bi" mengacu pada dua atau dua kali. Bahkan dalam distribusi ini, ada dua kemungkinan hasil dan merupakan kelanjutan dari Bernoulli dengan banyak percobaan. Artinya n > 1. Suatu variabel acak diskrit X dikatakan mengikuti distribusi binomial jika fungsi massa probabilitasnya (p.m.f) berbentuk,

$$p(x) = P[X = x]$$

$$= \begin{cases} n_x^c\, p^x q^{(n-x)}, x = 0,1,2, \dots, n \; dan \; 0 < p < 1 \; dan \; p + q = 1 \\ \qquad\qquad\qquad\qquad 0, lainnya \end{cases}$$

Model binomial digunakan untuk menghitung

- Jumlah orang yang melanggar peraturan lalu lintas dalam sehari.
- Berapa kali danau meluap karena kelebihan hujan.
- Banyaknya perkara hukum yang tertunda peradilannya karena kendala waktu.
- Berapa kali AC diperbaiki di hotel pada musim panas, dll.

Karakteristik

- Jumlah percobaan n terbatas.
- Uji coba ini bersifat dikotomis.
- Uji coba ini bersifat independen.

• Probabilitas keberhasilan p tetap sama untuk semua percobaan.

$Mean = \mu = E(x) = np$

$Variance = \sigma^2 = np(1-p)$

$Standar\ deviation = \sigma = \sqrt{np(1-p)}$

$dimana\ n = jumlah\ percobaan,$

$p = peluang\ sukses, 1 - p = peluang\ gagal$

Contoh

Manitoba Northern Pike adalah ikan yang kuat dan tangguh! dengan menggunakan umpan buatan dengan kait treble berduri, ditemukan angka kematian akibat kait hanya sekitar 5%. Artinya, hanya 5% tombak yang ditangkap dan dilepasliarkan yang mati (Diadaptasi dari prosiding National Symposium on Catch and Release Fishing yang disponsori oleh Humboldt State University). Misalkan sekelompok pemancing menangkap dan melepaskan 15 tombak utara Manitoba. Berapa kemungkinannya,

A. Tidak ada ikan yang mati?
B. Kurang dari 3 ikan mati?
C. Semua ikan itu hidup?
D. Lebih dari 13 ikan hidup?

Studi Kasus: Distribusi binomial.
Alasan:
• Jumlah percobaan n = 15 terbatas.
• Uji coba ini bersifat dikotomis—mati/selamat.
• Uji coba tersebut bersifat independen—satu uji coba tidak bergantung pada uji coba lainnya.
• Probabilitas keberhasilan p = 0,05 tetap sama untuk semua percobaan.
Jawab:
Uji cobanya adalah menangkap dan melepaskan tombak. Berhasil = tombak mati. Kegagalan = tombak hidup.
n = 15; p = 0,05; q = 0,95.
Misalkan X adalah peubah acak yang menunjukkan jumlah ikan yang mati.

A. P (x = 0) = 0,463
B. P (x > 3) = 1 − P(X ≤ 3) = 0,964
C. P (x = 0) = 0,463 (semua nyawa sama dengan tidak ada yang mati).
D. Ubah kesuksesan menjadi hidup. p = 0,95, P (x > 13) = 0,828

Pengenalan Distribusi Normal

Mari kita mulai dengan Unknown X lagi! (Yang kita lihat di awal bab pertama). Jika kita perlu memprediksi X, maka kita perlu mengetahui beberapa informasi latar belakang tentang X. Agar prediksi kita tidak berdasar. Dalam istilah statistik, kami menyebutnya sebagai "Memahami distribusi yang mendasarinya".

Ketika kita memahami sebaran data, kita mendapatkan persepsi umum tentang bagaimana data dimodelkan sehingga mudah untuk memperkirakan X yang tidak diketahui. Salah satu distribusi yang paling terkenal dalam statistik adalah distribusi Gaussian atau normal. Hal ini karena sebagian besar kumpulan data yang tersedia dapat dimodelkan ke dalam distribusi normal. Distribusi ini seperti kurva berbentuk lonceng simetris dengan ekor asimtotik di ujungnya.

Kurva lonceng seperti "Satu ukuran cocok untuk semua model"! Ini berarti bahwa ketika n (ukuran sampel) bertambah, sebagian besar kumpulan data cenderung masuk ke dalam kurva lonceng, inilah gagasan di balik teorema terkenal yang disebut "teorema batas pusat" yang didefinisikan sebagai,

Teorema batas pusat (CLT) menyatakan bahwa distribusi variabel sampel mendekati distribusi normal (yaitu 'kurva lonceng') ketika ukuran sampel menjadi lebih besar, dengan asumsi bahwa semua sampel memiliki ukuran yang identik, dan terlepas dari distribusi populasi yang sebenarnya. membentuk.

Untuk lebih memahami CLT, lihat Gambar 4.3 baris pertama menunjukkan bentuk beberapa model data seperti distribusi seragam, kurva normal miring, dan pola data acak umum. Pada baris kedua, bentuk kumpulan data ini digambarkan ketika jumlah titik datanya hanya dua unit. Pada baris ketiga, kita dapat melihat bahwa seiring bertambahnya ukuran sampel (n = 30) semua tipe data cenderung berbentuk kurva lonceng. Inilah konsep di balik teorema limit pusat. Ini adalah teorema penting yang dapat diterapkan di semua bidang analitik.

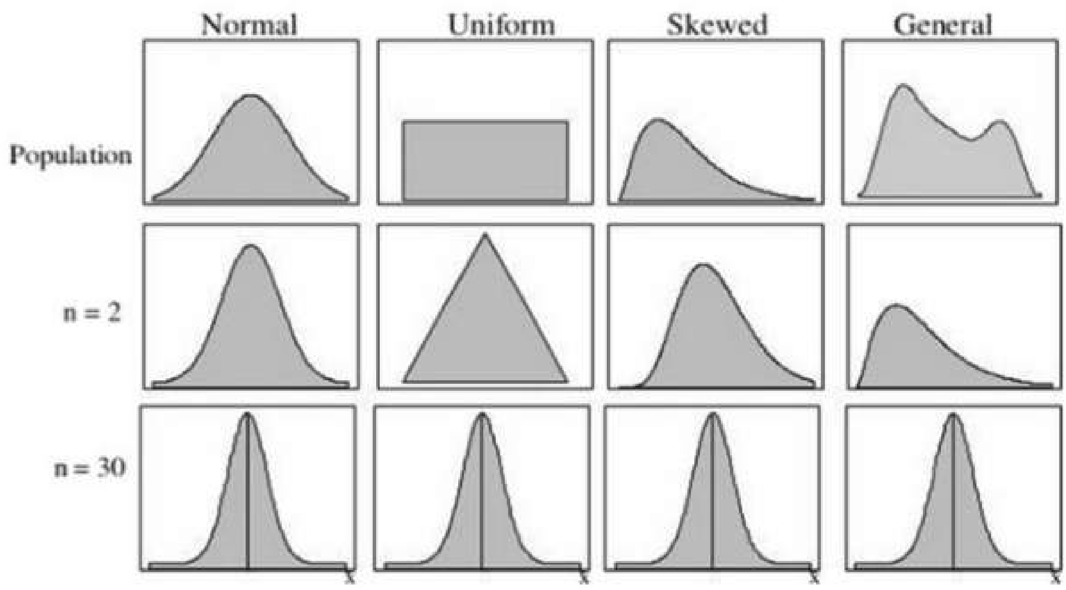
Grafik Distribusi Normal:

Mari kita mengingat kembali dan menyusun semua yang telah kita pelajari tentang distribusi normal.

1. Kurvanya simetris sempurna terhadap rata-rata distribusi. Hasilnya, mean distribusinya sama dengan modus dan mediannya.

2. Kurva mendekati sumbu horizontal tetapi tidak pernah menyentuh atau memotongnya.

3. Titik transisi antara bekam ke atas dan ke bawah terjadi di atas dan di bawah rentang interval ini, $\mu + \sigma$ dan $\mu - \sigma$.



Gambar 4.3 Penjelasan diagram (C.L.T)

G. Latihan Soal

Tabel di bawah ini menunjukkan probabilitas perkiraan cuaca pada hari ini

Hujan lebat	0,2
Hujan gerimis	0,5
Cerah	0,1
Angin ribut	0,2

Berapakah probabilitas bahwa hari ini :

a. Cerah
b. Tidak cerah
c. Tidak hujan
d. Hujan lebat atau hujan gerimis
e. Hujan lebat dan hujan gerimis

DAFTAR PUSTAKA

Arifin, M. H. (2014). Konsep-konsep Dasar statistika. *Jakarta: Universitas Terbuka*.

Bluman, A. G. (2018). *Elementary Statistics: A Step By Step Approach* (Tenth Edit). McGraw-Hill Education.

Janah, M., & Kartini, A. Y. (2022). Penerapan Metode Regresi Linier Berganda Pada Kasus Balita Gizi Buruk Di Kabupaten Bojonegoro. *Jurnal Statistika Dan Komputasi*, *1*(2), 74–82.

Mendenhall, W. (2013). *Introduction to Probability and Statistics, 3rd*. Nelson Educatio.

Mendenhall, W., Beaver, R. J., & Beaver, B. M. (2012). *Introduction to probability and statistics*. Cengage Learning.

Mendenhall, W., Beaver, R. J., & Beaver, B. M. (2020). *Introduction to probability and statistics* (Fifteenth). Cengage Learning, Inc.

Nugroho, S. (2007). *Dasar-dasar Metode Statistika*. Grasindo.

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

Santoso, S. (2016). *Panduan Lengkap SPSS Versi 23*. Gramedia Direct.

Steel, R. G. D., & Torrie, J. H. (1976). Introduction to statistics. *(No Title)*.

Sudjana. (2001). *Metode Statistika*. Tarsito.

Sudjana. (2015). *Metode Statistika*. PT Tarsito Bandung.

Walpole, R. E. (1995). *Pengantar statistika*.

Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2016). Probability & Statistics for Engineers & Scientists. In *Journal of the American Statistical Association* (Ninth Edit). Pearson Education Limited.

Yitnosumarto, S. (1990). Dasar-Dasar Statistika. *C. V Rajawali: Jakarta*.

BAB 6

Distribusi Sampling dan Keterkaitannya dengan Populasi Serta Dalil Limit Pusat

A. Tujuan Pembelajaran

Diharapkan kemampuan mahasiswa untuk mengidentifikasi dan merumuskan permasalahan yang konkret atau nyata yang dapat dipecahkan melalui penggunaan distribusi sampling. Ini berkaitan dengan keahlian dalam mengenali dan mendefinisikan masalah yang memerlukan pendekatan melalui distribusi sampling. Mahasiswa diharapkan dapat menggunakan dan menerapkan berbagai metode statistik yang relevan dalam konteks distribusi sampling. Ini termasuk pemahaman tentang cara menggunakan teknik sampling untuk mengumpulkan data dan menjalankan analisis statistik yang diperlukan.

Kemampuan untuk menyampaikan hasil analisis distribusi sampling secara jelas dan efektif, baik melalui presentasi lisan maupun laporan tertulis. Hal ini penting untuk memastikan bahwa informasi hasil analisis dapat dipahami dengan baik oleh berbagai pihak yang terlibat. Diharapkan juga kemampuan mahasiswa dalam mengerti dan mengaplikasikan konsep dasar tentang teori probabilitas, variabel acak, dan konsep distribusi dalam konteks teknik sampling. Ini termasuk pemahaman tentang bagaimana probabilitas dan distribusi digunakan dalam pengambilan sampel yang representatif dari populasi.

B. Pendahuluan

Distribusi sampling adalah konsep penting dalam statistika yang berkaitan dengan pengambilan sampel dari suatu populasi. Dalam materi distribusi sampling akan mencakup beberapa poin penting:

1. Distribusi Sampling Rata-Rata.
2. Distribusi Sampling Proporsi.
3. Distribusi Sampling Yang Lain.

4. Keterkaitan Distribusi Sampling Dari Rata-Rata Sampel Dengan Distribusi Normal.

Ketika pemilihan sampel atau sampling acak dari suatu populasi, ukuran deskriptif numerik yang dihitung dari sampel disebut statistik. Statistik ini bervariasi atau berubah untuk setiap sampel acak yang berbeda yang dipilih, demikian ini disebuat sebagai variabel acak. Distribusi probabilitas untuk statistik disebut distribusi sampling karena pada proses pengambilan sampel secara berulang sekaligus memberitahu informasi terkait nilai-nilai statistik apa saja yang dapat muncul, dan seberapa sering setiap nilai muncul. Dengan kata lain, distribusi sampling dari suatu statistik adalah distribusi probabilitas untuk kemungkinan nilai statistik yang dihasilkan ketika sampel acak berukuran n diambil secara berulang-ulang dari populasi (Mendenhall et al., 2020).

C. Distribusi Sampling Rata-Rata

Salah satu dalil statistik penting yang menjelaskan distribusi sampling statistik yang adalah jumlah atau rata-rata adalah Dalil Limit Pusat. Jika sampel acak dari n pengamatan diambil dari populasi tidak Normal dengan rata-rata berhingga μ dan simpangan baku σ, maka sampling distribusi sampel dari rata-rata sampel x ketika n besar kira-kira berdistribusi Normal, dengan rata-rata μ dan simpangan baku $\sigma/\sqrt{n}$ (Mendenhall et al., 2020). Jika sampel acak berukuran n dipilih dari populasi dengan rata-rata μ dan simpangan baku σ, distribusi sampling dari rata-rata sampel $\bar{x}$ akan memiliki rata-rata μ dan simpangan baku $\sigma/\sqrt{n}$. Jika populasi berdistribusi Normal, distribusi sampling $\bar{x}$ akan berdistribusi Normal, berapa pun ukuran sampelnya (n). Jika distribusi populasi tidak Normal, distribusi sampling dari $\bar{x}$ akan berdistribusi Normal untuk sampel yang besar (berdasarkan Dalil Limit Pusat). Secara konservatif, dibutuhkan ukuran $n \geq 30$. Simpangan baku dari statistik yang digunakan sebagai dugaan parameter populasi adalah *standard error* (SE) dugaan karena mengacu pada ketepatan dugaan. Oleh karena itu, simpangan baku dari $\bar{x}$ yang (diberikan oleh $\sigma/\sqrt{n}$) disebut sebagai *standard error* dari rata-rata, $SE(\bar{x})$.

Dalam kasus khusus pada distribusi sampling rata-rata, ada dua prinsip penting terkait pengambilan sampel (Walpole et al., 2016), sebagai berikut:

1. Jika dilakukan pengambilan sampel dengan pengembalian dari populasi berhingga, maka distribusi sampling dari rata-rata sampel $\bar{x}$ akan memiliki rata-rata μ dan simpangan baku $\sigma/\sqrt{n}$.
2. Jika dilakukan pengambilan sampel tanpa pengembalian dari populasi berhingga, maka distribusi sampling dari rata-rata sampel $\bar{x}$ akan memiliki rata-rata μ dan simpangan baku $\frac{\sigma}{\sqrt{n}}\left(\frac{N-n}{N-1}\right)$.

Ketika $\bar{x}$ merupakan rata-rata sampel dari suatu sampel acak berukuran n yang diambil dari populasi dengan rata-rata μ dan simpangan baku σ, maka bentuk formulasi pembatas distribusi dari Z adalah $Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ dengan syarat n mendekati ∞ dan kemudian Z mengikuti distribusi Normal baku yang memiliki rata-rata 0 dan simpangan baku 1 (Walpole et al., 2016).

Diberikan contoh soal tentang waktu tempuh perjalanan bus kampus. Waktu tempuh bus tersebut memiliki rata-rata dalam bulan lalu adalah 28 menit dengan simpangan baku 5 menit. Pada bulan ini bus tersebut digunakan sebanyak 40 kali. Berapa probabilitas rata-rata waktu tempuh lebih dari 30 menit? ; Untuk menjawabnya, diberikan penyelesaian dengan mengetahui informasi data bahwa μ = 28 menit dan σ = 5 menit, serta ukuran sampel n = 40 dan $\bar{x}$ = 30.

$$P(\bar{X} > 30) = \cdots ?$$

$$P(\bar{X} > \bar{x}) = P\left(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} > \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)$$

$$P(\bar{X} > 30) = P\left(Z > \frac{30-28}{5/\sqrt{40}}\right)$$

$$P(\bar{X} > 30) = P(Z > 2{,}53)$$

$$P(\bar{X} > 30) = 1 - P(Z \leq 2{,}53)$$

Dengan menggunakan tabel statistik yaitu Tabel Normal Baku, probabilitas $P(Z \leq 2{,}53)$ dapat diperoleh.

$$P(Z \leq z) = F_Z(z) \text{ dimana } z = 2{,}53$$

$$P(Z \leq 2{,}53) = \cdots ?$$

Tabel 6.1 Tabel statistik Normal baku untuk penerapan pertama.

z	0.00	...	0.03	...
:			↓	
2.5	-----	→	0.9911	
:				

sehingga $P(Z \leq 2{,}53) = 0{,}9911$,

maka $P(\bar{X} > 30) = 1 - P(Z \leq 2{,}53)$

$P(\bar{X} > 30) = 1 - 0{,}9911 = 0{,}0089$.

Hasil probabilitas rata-rata waktu tempuh lebih dari 30 menit bernilai sangat kecil, sehingga dapat dilakukan dugaan bahwa rata-rata waktu tempuh bus tersebut tidak lebih dari 30 menit.

D. Distribusi Sampling Proporsi

Jika sampel acak sebanyak n pengamatan dipilih dari populasi Binomial dengan parameter p, maka distribusi sampling dari proporsi sampel $\hat{p} = \frac{x}{n}$ yang mana dihasilkan dari populasi dengan mean $\mu = np$ dan varian terhingga $\sigma^2 = npq$ dengan $q = 1 - p$ maka bentuk distribusi limit dari $Z = \frac{\hat{p}-p}{\sqrt{pq/n}}$ untuk n mendekati ∞ dan kemudian Z mengikuti distribusi Normal baku (Mendenhall et al., 2020).

Diberikan contoh soal misalkan seseorang mengklaim bahwa proporsi p dari populasi orang tua adalah 0,55. Berapa probabilitas untuk mengamati proporsi sampel yang sama atau lebih besar dari nilai yang diamati yaitu 0,60 dengan ukuran sampel n = 500 ? ; Untuk menjawabnya, diberikan penyelesaian dengan mengetahui informasi data bahwa p = 0,55 , serta ukuran sampel n = 500 dan dan $\hat{p} = 0{,}60$.

$P(\hat{p} \geq 0{,}60)$?

$$P(\hat{p} \geq \hat{p}_x) = P\left(Z > \frac{\hat{p}-p}{\sqrt{pq/n}}\right)$$

$$P(\hat{p} \geq 0{,}60) = P\left(Z > \frac{0{,}60-0{,}55}{\sqrt{\frac{(0{,}55)(0{,}45)}{500}}}\right)$$

$$P(\hat{p} \geq 0{,}60) = P(Z > 2{,}25) = 1 - P(Z \leq 2{,}25)$$

Dengan menggunakan tabel statistik yaitu Tabel Normal Baku, probabilitas $P(Z \leq 2{,}25)$ dapat diperoleh.

$$P(Z \leq z) = F_Z(z)\;; \quad z = 2{,}25$$

$$P(Z \leq 2{,}25) = \ldots ?$$

Tabel 6.2 Tabel statistik Normal baku untuk penerapan kedua.

z	0.00	...	0.05	...
:			↓	
2.2	------	→	0.9878	
:				

sehingga $P(Z \leq 2{,}25) = 0{,}9878$,

maka $P(\hat{p} > 0{,}60) = 1 - P(Z \leq 2{,}25)$

$$P(\hat{p} > 0{,}60) = 1 - 0{,}9878 = 0{,}0122$$

Hasil probabilitas proporsi sampel bernilai sangat kecil, sehingga dapat dilakukan dugaan bahwa proporsi sampel tidak lebih besar dari 0,60.

E. Distribusi Sampling Yang Lain

Jika simpangan baku dari populasi tidak diketahui (σ), maka simpangan baku tersebut akan diduga dengan simpangan baku dari sampel yaitu s. Demikian hingga, digunakan distribusi *t-student* dengan bentuk formulasi pembatas distribusi dari t adalah $T = \frac{\bar{x}-\mu}{s/\sqrt{n}}$ dengan derajat bebas $v = n - 1$ (Walpole et al., 2016). Sebagai catatan, distribusi t dapat digunakan ketika ukuran sampel kecil atau besar di mana ukuran sampel kecil yaitu $n \leq 30$. Ketika ukuran sampel sangat besar yaitu n mendekati ∞, bentuk kurva distribusi t hampir sama dengan distribusi Normal baku (Z).

Diberikan contoh soal tentang waktu tempuh perjalanan bus kampus. Waktu tempuh bus tersebut memiliki rata-rata dalam bulan lalu adalah 28 menit. Dari sampel 40 pengamatan, diperoleh simpangan baku dari sampel sebesar 5,5 menit. Berapa probabilitas rata-rata waktu tempuh lebih dari 30 menit? ; Untuk menjawabnya, diberikan penyelesaian dengan mengetahui informasi data bahwa μ = 28 menit, serta ukuran sampel n = 40, $v = n - 1 = 39$, s = 5,5, dan $\bar{x}$ = 30.

$$P(\bar{X} > 30) = \cdots ?$$

$$P(\bar{X} > \bar{x}) = P\left(\frac{\bar{X}-\mu}{s/\sqrt{n}} > \frac{\bar{x}-\mu}{s/\sqrt{n}}\right)$$

$$P(\bar{X} > 30) = P\left(T > \frac{30-28}{5{,}5/\sqrt{40}}\right)$$

$$P(\bar{X} > 30) = P(T > 2{,}30)$$

Dengan menggunakan tabel statistik yaitu Tabel *t-student*, probabilitas $P(T > 2{,}30)$ dapat diperoleh.

$P(T > t)$ dimana t = 2,30

$P(T > 2{,}30) = \cdots ?$ dan v = 39

	A	B	C	D
1	*t*	*v*	P(T > *t*)	
2	2.30	39	0.0134	
3				

Gambar 6.1 Tabel statistik *t-student* untuk penerapan pertama.

Digunakan software MS-Excel untuk menyelessaikan masalah ini dengan prosedur sebagai berikut:

1. Memasukan nilai t sebesar **2.30** pada kotak **cell A2** dan kemudian masukkan nilai v sebesar **39** pada kotak **cell B2**.
2. Memasukkan Formula perhitungan tabel statistik untuk distribusi *t-student* pada kotak **cell C2**, yaitu **=TDIST(A2,B2,1)** .

sehingga $P(T > 2{,}30) = 0{,}0134$,

Hasil probabilitas rata-rata waktu tempuh lebih dari 30 menit bernilai sangat kecil, sehingga dapat dilakukan dugaan bahwa rata-rata waktu tempuh bus tersebut tidak lebih dari 30 menit meskipun simpangan baku dari populasi tidak diketahui.

F. Keterkaitan Distribusi Sampling Dari Rata-Rata Sampel Dengan Distribusi Normal

Dalam statistik, konsep distribusi sampling dari rata-rata sampel terkait erat dengan distribusi Normal, terutama melalui Dalil Limit

Pusat atau Central Limit Theorem (CLT). CLT menyatakan bahwa jika diambil sampel acak dari suatu populasi dengan ukuran sampel yang cukup besar, maka distribusi dari rata-rata sampel tersebut akan mendekati distribusi Normal, terlepas dari bentuk distribusi asli dari populasi tersebut. Artinya, bahkan jika populasi tidak terdistribusi secara Normal, rata-rata dari sampel-sampel yang diambil dari populasi tersebut akan memiliki distribusi yang mendekati Normal seiring dengan pertambahan ukuran sampel.

Dengan kata lain, jika diammbil banyak sampel dari populasi apa pun dan menghitung rata-rata dari setiap sampel, distribusi dari rata-rata tersebut akan cenderung membentuk kurva Normal. Hal ini memiliki implikasi penting dalam inferensi statistik, karena kita dapat menggunakan distribusi Normal untuk membuat perkiraan atau menguji hipotesis tentang rata-rata populasi atau parameter lainnya, terutama jika ukuran sampel besar. Dengan demikian, keterkaitan antara distribusi sampling dari rata-rata sampel dengan distribusi Normal adalah bahwa, ketika ukuran sampel cukup besar, distribusi dari rata-rata sampel akan mendekati distribusi Normal berdasarkan pada CLT. Ini memungkinkan penggunaan berbagai metode statistik yang bergantung pada distribusi normal untuk melakukan inferensi tentang populasi, meskipun distribusi populasi mungkin tidak terdistribusi secara Normal.

G. Soal Latihan

1. Untuk X berdistribusi Normal dengan $\mu = 50$ dan $\sigma = 10$. Untuk ukuran sampel 10, hitung:
   a. P($\bar{X}$ = 40)
   b. P($\bar{X} \leq 50$)
   c. P($\bar{X} \geq 30$)
2. Untuk X berdistribusi Binomial dengan n = 100 dan p = 0.5 , hitung:
   a. P(p < 0,4)
   b. P(0,4 ≤ p ≤ 0,5)
   c. P(p > 0,3)

3. Diketahui populasi $N = \{2, 3, 5, 7, 9, 11, 13, 17, 19, 23\}$. Hitunglah probabilitas bahwa sebuah sampel acak berukuran 9 yang diambil tanpa pengembalian akan menghasilkan mean sampel lebih besar dari 4,5.
4. Diketahui populasi $N = \{2, 3, 5, 7, 9, 11, 13, 17, 19, 23\}$.Hitunglah probabilitas bahwa sebuah sampel acak berukuran 5 yang diambil dengan pengembalian akan menghasilkan mean sampel lebih besar dari 4,

DAFTAR PUSTAKA

Arifin, M. H. (2014). Konsep-konsep Dasar statistika. *Jakarta: Universitas Terbuka*.

Bluman, A. G. (2018). *Elementary Statistics: A Step By Step Approach* (Tenth Edit). McGraw-Hill Education.

Janah, M., & Kartini, A. Y. (2022). Penerapan Metode Regresi Linier Berganda Pada Kasus Balita Gizi Buruk Di Kabupaten Bojonegoro. *Jurnal Statistika Dan Komputasi*, *1*(2), 74–82.

Mendenhall, W. (2013). *Introduction to Probability and Statistics, 3rd*. Nelson Educatio.

Mendenhall, W., Beaver, R. J., & Beaver, B. M. (2012). *Introduction to probability and statistics*. Cengage Learning.

Mendenhall, W., Beaver, R. J., & Beaver, B. M. (2020). *Introduction to probability and statistics* (Fifteenth). Cengage Learning, Inc.

Nugroho, S. (2007). *Dasar-dasar Metode Statistika*. Grasindo.

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

Santoso, S. (2016). *Panduan Lengkap SPSS Versi 23*. Gramedia Direct.

Steel, R. G. D., & Torrie, J. H. (1976). Introduction to statistics. *(No Title)*.

Sudjana. (2001). *Metode Statistika*. Tarsito.

Sudjana. (2015). *Metode Statistika*. PT Tarsito Bandung.

Walpole, R. E. (1995). *Pengantar statistika*.

Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2016). Probability & Statistics for Engineers & Scientists. In *Journal of the American Statistical Association* (Ninth Edit). Pearson Education Limited.

Yitnosumarto, S. (1990). Dasar-Dasar Statistika. *C. V Rajawali: Jakarta*.

BIOGRAFI PENULIS

Alif Yuanita Kartini, S.Si., M.Si, lahir di Tuban pada 21 April 1986 dan sekarang menetap di Bojonegoro. Penulis menyelesaikan pendidikan sarjana dan magister di Jurusan Statistika Institut Teknologi Sepuluh Nopember Surabaya. Saat ini penulis adalah dosen tetap di Universitas Nahdlatul Ulama Sunan Giri dengan homebase di Program Studi Statistika. Bidang keilmuan penulis adalah statistika pemodelan dan statistika non parametrik. Penulis dapat dihubungi melalui alifyuanita@unugiri.ac.id

Denny Nurdiansyah, S.Si., M.Si Lulus S1 di Jurusan Matematika Universitas Airlangga pada tahun 2010, lulus S2 di Jurusan Statistika Institut Teknologi Sepuluh Nopember Surabaya pada tahun 2014. Saat ini penulis adalah dosen tetap di Program Studi Statistika Universitas Nahdlatul Ulama Sunan Giri Bojonegoro. Bidang keahlian penulis adalah statistika komputasi dan peramalan. Penulis saat ini tinggal di Surabaya. Penulis dapat dihubungi melalui surat elektronik denny.nur@unugiri.ac.id

Nita Cahyani, S.Si., M.Stat menyelesaikan Pendidikan S1 di Jurusan Statistika Institut Teknologi Sepuluh Nopember Surabaya pada tahun 2013 dan menyelesaikan Pendidikan S2 di JuJurusan Statistika Institut Teknologi Sepuluh Nopember Surabaya pada tahun 2018. Penulis merupakan dosen tetap di Program Studi Statistika Universitas Nahdlatul Ulama Sunan Giri Bojonegoro. Penulis ahli di bidang statistika komputasi dan pemodelan statistika. Penulis dapat dihubungi melalui surat elekronik nitacahyani@unugiri.ac.id

buku Bu Alif

ORIGINALITY REPORT

14%	14%	2%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	pustaka.ut.ac.id Internet Source	4%
2	lms-paralel.esaunggul.ac.id Internet Source	4%
3	vdocuments.site Internet Source	2%
4	Submitted to Myanmar Imperial College Student Paper	1%
5	journal.unugiri.ac.id Internet Source	1%
6	mipa.ub.ac.id Internet Source	1%
7	pdfcoffee.com Internet Source	1%
8	www.scirp.org Internet Source	1%

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	On		