

# STATISTIKA DASAR

Prima Kurniati Hamzah, M. Ahkam Alwi, R. Lutfi Zakaria,  
Susana Labuem, Ahmad Ridfah, Vera Rosalina Bulu, Farida,  
Herry Agung Prabowo, Agung Wahyudi Biantoro, Yani Quarta  
Mondiana, Sawarni Hasibuan, Daniel Williams Fointuna



# **STATISTIKA DASAR**

**Prima Kurniati Hamzah**

**M. Ahkam Alwi**

**R Lutfi Zakaria**

**Susana Labuem**

**Ahmad Ridfah**

**Vera Rosalina Bulu**

**Farida**

**Herry Agung Prabowo**

**Agung Wahyudi Biantoro**

**Yani Quarta Mondiana**

**Sawarni Hasibuan**

**Daniel Williams Fointuna**



**CV HEI PUBLISHING INDONESIA**

# STATISTIKA DASAR

## **Penulis:**

Prima Kurniati Hamzah  
M. Ahkam Alwi  
R Lutfi Zakaria  
Susana Labuem  
Ahmad Ridfah  
Vera Rosalina Bulu  
Farida  
Herry Agung Prabowo  
Agung Wahyudi Biantoro  
Yani Quarta Mondiana  
Sawarni Hasibuan  
Daniel Williams Fointuna

**ISBN: 978-634-7214-32-4**

**Editor : Ariyanto, S.Pd, M.Pd**

**Penyunting : Atyka Trianisa, S.Pd**

**Desain Sampul dan Tata Letak : Ipah Kurnia Putri S.St**

**Penerbit : CV HEI PUBLISHING INDONESIA**

**Nomor IKAPI 043/SBA/2023**

## **Redaksi :**

Jl. Air Paku No.29 RSUD Rasidin, Kel. Sungai Sapih, Kec Kuranji  
Kota Padang Sumatera Barat  
Website : [www.HeiPublishing.id](http://www.HeiPublishing.id)  
Email : [heipublishing.id@gmail.com](mailto:heipublishing.id@gmail.com)

**Cetakan pertama, April 2025**

Hak cipta dilindungi undang-undang  
Dilarang memperbanyak karya tulis ini dalam bentuk  
dan dengan cara apapun tanpa izin tertulis dari penerbit.

## KATA PENGANTAR

Puji syukur kepada Allah subhanahu wa'taala atas rahmat dan karunia-Nya sehingga buku "Statistika Dasar", dapat terselesaikan dengan baik. Buku ini berisikan tentang Pengantar Statistika, Data dalam Statistika, Penyajian Data, Ukuran Pemusatan Data, Ukuran Penyebaran Data, Probabilitas Dasar, Distribusi Peluang, Sampling dan Populasi, Distribusi Sampling, Uji Hipotesis, Analisis Regresi dan Korelasi, Analisis Variansi (ANOVA).

Semoga buku ini dapat menjadi referensi yang bermanfaat bagi mahasiswa, dosen, dan para profesional di bidang Statistika Dasar, serta siapa saja yang tertarik mempelajari Statistika Dasar. Terima kasih kepada semua pihak yang telah berkontribusi dalam penyusunan buku ini, Harapan terbesar buku ini dapat memberikan manfaat dan kontribusi positif dalam perkembangan ilmu pengetahuan.

Selamat membaca dan semoga bermanfaat.

Padang, April 2025

Penulis

## DAFTAR ISI

KATA PENGANTAR .....	i
DAFTAR ISI .....	ii
DAFTAR TABEL.....	vi
DAFTAR GAMBAR .....	vii
BAB 1 PENGANTAR STATISTIK.....	1
1.1 Pengertian Statistik .....	1
1.2 Jenis-jenis Statistika.....	1
1.2.1 Jenis Statistika berdasarkan tujuan pengolahan data.....	1
1.2.2 Jenis Statistika berdasarkan variabel .....	5
1.2.3 Jenis Statistika berdasarkan parameter .....	5
1.3 Jenis-jenis data.....	6
1.3.1 Jenis Statistika berdasarkan bentuknya.....	6
1.3.2 Jenis Statistika berdasarkan skala.....	6
1.3.3 Jenis Statistika berdasarkan sumbernya .....	7
DAFTAR PUSTAKA.....	8
BAB 2 DATA DALAM STATISTIKA.....	9
2.1 Pendahuluan.....	9
2.2 Jenis-Jenis Data.....	9
2.2.1 Berdasarkan Sifatnya .....	9
2.2.2 Berdasarkan Sumbernya .....	10
2.2.3 Berdasarkan Skala Pengukurannya.....	11
2.3 Metode Pengumpulan Data .....	13
2.4 Penyajian Data .....	18
2.4.1 Tabel.....	18
2.4.2 Diagram dan Grafik.....	19
2.4.3 Diagram Lingkaran .....	20
2.4.4 Histogram.....	22
2.4.5 Diagram Garis .....	23
2.5 Kesimpulan .....	24
DAFTAR PUSTAKA.....	25
BAB 3 PENYAJIAN DATA.....	31
3.1 Pendahuluan.....	31
3.2 Jenis Penyajian Data.....	32
3.2.1 Penyajian Data dalam Bentuk Tabel.....	32

3.2.2 Penyajian Data dalam Bentuk Grafik.....	34
3.3 Penyajian Data dalam Kehidupan Sehari-hari.....	37
3.4 Kesimpulan.....	38
DAFTAR PUSTAKA.....	39
<b>BAB 4 UKURAN PEMUSATAN DATA .....</b>	<b>41</b>
4.1 Konsep Dasar Ukuran Pemusatan.....	41
4.2 Data Tunggal .....	41
4.3 Data Kelompok.....	50
DAFTAR PUSTAKA.....	56
<b>BAB 5 UKURAN PENYEBARAN DATA.....</b>	<b>57</b>
5.1 Pendahuluan.....	57
5.2 Jenis-jenis Ukuran Penyebaran .....	57
5.2.1 Rentang ( <i>Range</i> ) .....	58
5.2.2 Rentang Antar Kuartil ( <i>Interquartile Range - IQR</i> ) ..	58
5.2.3 Deviasi Rata-rata ( <i>Mean Deviation</i> ).....	59
5.2.4 Varians.....	59
5.2.5 Simpangan Baku (Standar Deviasi) .....	60
5.3 Keunggulan <i>Mean Deviation</i> dibandingkan Standar Deviasi .....	61
5.3.1 <i>Mean Deviation</i> sebagai Alternatif Standar Deviasi .	61
5.3.2 Mengapa Standar Deviasi Mendominasi?.....	62
5.3.3 Kritik terhadap Standar Deviasi.....	62
5.4 Kesalahan Umum dalam Interpretasi Ukuran Penyebaran.....	63
5.5 <i>Reasoning</i> tentang Ukuran Penyebaran Data.....	63
5.6 Jamovi untuk Ukuran Penyebaran Data.....	63
DAFTAR PUSTAKA.....	67
<b>BAB 6 PROBABILITAS DASAR .....</b>	<b>69</b>
6.1 Pengertian Probabilitas.....	69
6.2 Pendekatan probabilitas.....	70
6.3 Aturan dasar Probabilitas.....	71
6.4 Sifat – Sifat Probabilitas .....	72
6.5 Teorema Bayes.....	72
6.6 Faktorial, Permutasi dan Kombinasi .....	74
DAFTAR PUSTAKA.....	77
<b>BAB 7 DISTRIBUSI PROBABILITAS.....</b>	<b>79</b>
7.1 Pendahuluan .....	79

7.1.1 Bilangan Faktorial.....	79
7.1.2 Permutasi .....	79
7.1.3 Kesimpulan.....	80
7.2 Konsep Dasar Probabilitas.....	80
7.2.1 Definisi Probabilitas .....	80
7.2.2 Ruang Sampel ( <i>Sample Space</i> ).....	80
7.2.3 Kejadian ( <i>Event</i> ).....	81
7.2.5 Kesimpulan.....	81
7.3 Distribusi Probabilitas .....	81
7.3.1 Variabel Acak.....	82
7.3.2 Variabel Acak Diskrit.....	82
7.3.2 Variabel Acak Kontinu.....	85
DAFTAR PUSTAKA.....	89
<b>BAB 8 SAMPLING DAN POPULASI.....</b>	<b>91</b>
8.1 Pendahuluan.....	91
8.2 Populasi dalam Penelitian.....	92
8.3 Teknik Sampling .....	93
8.4 Kesimpulan .....	98
DAFTAR PUSTAKA.....	99
<b>BAB 9 DISTRIBUSI SAMPLING.....</b>	<b>101</b>
9.1 Populasi Dan Sampel .....	101
9.1.1 Populasi .....	101
9.1.2 Sampel.....	102
9.2 Sampling .....	103
9.3 Distribusi Sampling .....	104
DAFTAR PUSTAKA.....	112
<b>BAB 10 PENGUJIAN HIPOTESIS.....</b>	<b>113</b>
10.1 Pendahuluan.....	113
10.2 Konsep Teoritis.....	113
10.3 Kriteria pengujian Hipotesis Melalui Statistik Uji .....	116
10.4 Uji Hipotesis Terhadap Rata-rata Populasi .....	116
10.5 Uji Hipotesis terhadap ragam populasi.....	119
10.6 Uji Hipotesis terhadap selisih rata-rata populasi .....	120
DAFTAR PUSTAKA.....	124
<b>BAB 11 ANALISIS REGRESI DAN KORELASI .....</b>	<b>125</b>
11.1 Pendahuluan.....	125
11.2 Teori Dasar Analisis Regresi.....	125

11.2.1 Metode Estimasi Parameter .....	126
11.2.2 Asumsi Klasik.....	128
11.2.3 Uji Hipotesis .....	132
11.2.4 Uji Parsial (Uji t) .....	132
11.2.5 Uji Simultan (Uji F) .....	133
11.2.5 Analisis Determinasi (R-square) .....	133
11.3 Teori Dasar Analisis Korelasi.....	133
11.4 Studi Kasus.....	137
11.5 Penutup.....	150
DAFTAR PUSTAKA.....	152
<b>BAB 12 ANALISIS VARIANS (ANOVA) .....</b>	<b>153</b>
12.1 Pendahuluan .....	153
12.2 Analisis Varians Satu Jalur (ANOVA Satu Jalur).....	154
12.3 Analisis Varians Dua Jalur (ANOVA Dua Jalur) .....	160
DAFTAR PUSTAKA.....	167
<b>BIODATA PENULIS</b>	



## DAFTAR TABEL

<b>Tabel 2.1.</b> Kepuasan Pelanggan .....	18
<b>Tabel 3.1.</b> Contoh Tabel Satu Arah .....	32
<b>Tabel 3.2.</b> Contoh Tabel Dua Arah .....	33
<b>Tabel 3.3.</b> Contoh Table Distribusi Frekuensi .....	33
<b>Tabel 4.1.</b> Menemukan Rata-Rata dari Suatu Distribusi Frekuensi .....	46
<b>Tabel 9.1.</b> Rumus distribusi sampling rata-rata .....	105
<b>Tabel 9.2.</b> Rumus terkait dengan distribusi sampling proporsi1 .....	08
<b>Tabel 9.3.</b> Rumus yang berkaitan dengan kekeliruan baku.....	110
<b>Tabel 9.4.</b> Rumus Distribusi sampling selisih rata rata.....	111
<b>Tabel 9.5.</b> Rumus Distribusi sampling selisih proporsi.....	111
<b>Tabel 10.1.</b> Kriteria pengujian hipotesis terhadap satu rata-rata populasi .....	118
<b>Tabel 10.2.</b> Statistik Uji dan Wilayah Kritis Uji Ragam Populasi .....	119
<b>Tabel 10.3.</b> Statistik uji dan Daerah Kritis pada Uji hipotesis terhadap Selisih rata-rata 2 populasi. .	121
<b>Tabel 11.1 .</b> Uji Statistik Durbin-Watson.....	131
<b>Tabel 11.2.</b> Hasil uji normalitas.....	146
<b>Tabel 12.1.</b> Hasil Percobaan Tikus dalam Labirin.....	158
<b>Tabel 12.2.</b> Ringkasan ANAVA Satu Jalur pada Percobaan Tikus dalam Labirin. ....	160
<b>Tabel 12.3.</b> Nilai Kuis Statistika Berdasarkan Tingkat Fobia dan Gender. ....	164
<b>Tabel 12.4.</b> Penentuan Nilai $F_{tabel}$ .....	166
<b>Tabel 12.5.</b> Ringkasan ANAVA Dua Jalur. ....	166

## DAFTAR GAMBAR

<b>Gambar 2.1.</b> Contoh diagram batam Distribusi Nilai Ujian Siswa.....	20
<b>Gambar 2.2.</b> Contoh diagram Lingkaran Proporsi Penjualan Produk.....	21
<b>Gambar 2.3.</b> Contoh Gambar Diagram Garis dalam berbagai bidang .....	23
<b>Gambar 3.1.</b> Contoh Grafik Batang .....	34
<b>Gambar 3.2.</b> Contoh Grafik Garis.....	35
<b>Gambar 3.3.</b> Contoh Grafik Lingkaran .....	35
<b>Gambar 3.4.</b> Contoh Histogram.....	36
<b>Gambar 3. 5.</b> Contoh Polygon Frekuensi.....	37
<b>Gambar 4.1.</b> Skewness.....	50
<b>Gambar 5.1.</b> Input Data.....	64
<b>Gambar 5.2.</b> Memindahkan Variabel .....	65
<b>Gambar 5.3.</b> Dispersi.....	65
<b>Gambar 8.1.</b> Populasi dan Sampel.....	92
<b>Gambar 8.2.</b> Pembagian Teknik Sampling.....	95
<b>Gambar 9.1.</b> Luas daerah menggunakan kurva distribusi z.....	106
<b>Gambar 9.2.</b> Luas daerah menggunakan kurva distribusi z .....	107
<b>Gambar 9.3.</b> Luas daerah menggunakan kurva distribusi z .....	109
<b>Gambar 11.1.</b> Model korelasi monotonik dan nonmonotonik.....	136
<b>Gambar 11.2.</b> Scatter diagram usia vs tekanan darah sistolik .....	145
<b>Gambar 11.3.</b> Box plot sebaran usia dan tekanan darah sistolik .....	146
<b>Gambar 12.1.</b> Pembagian Variasi pada Desain Faktorial Dua Jalur.....	161



# **BAB 1**

## **PENGANTAR STATISTIK**

**Oleh Prima Kurniati Hamzah**

### **1.1 Pengertian Statistik**

Statistika, yang berasal dari kata Latin 'status' atau 'statista' yang berarti 'negara', awalnya digunakan untuk mencatat urusan kenegaraan seperti jumlah penduduk, pajak, dan gaji guru. Pada abad ke-17 dan ke-18, istilah ini bersaing dengan 'political arithmetic' dan 'publisistika'.

Pada pertengahan abad ke-18, istilah 'statistika' muncul sebagai yang paling bertahan dari beberapa istilah serupa. Statistika adalah ilmu yang mempelajari pengumpulan, deskripsi, analisis data, dan penarikan kesimpulan untuk pengambilan keputusan. Seiring waktu, penerapannya meluas dari urusan negara ke berbagai bidang seperti pertanian, ekonomi, kedokteran, biologi, psikologi, dan pendidikan. Contohnya, statistika dalam ekonomi dikenal sebagai ekonometrika, dalam psikologi sebagai psikometrika, dan dalam biologi sebagai biometrika.

### **1.2 Jenis-jenis Statistika**

Pengelompokan jenis-jenis statistika dapat dilakukan berdasarkan tujuan pengolahan data, variabel yang terlibat, dan parameter yang dianalisis.

#### **1.2.1 Jenis Statistika berdasarkan tujuan pengolahan data**

Berikut adalah rincian jenis-jenis statistika menurut klasifikasi tersebut.

##### **1. Statistika Deskriptif**

Statistika deskriptif merupakan cabang ilmu statistika yang fokus pada bagaimana cara mengumpulkan dan menyajikan data sedemikian rupa sehingga mudah dipahami dan memberikan informasi yang berguna. Tujuan utama dari statistika deskriptif

adalah untuk menjelaskan dan menggambarkan kondisi atau permasalahan yang ada tanpa menarik kesimpulan yang lebih luas.

Jika data yang diolah berasal dari sampel, statistika deskriptif akan menghasilkan ukuran-ukuran yang disebut statistik sampel. Sebaliknya, jika data yang diolah mencakup seluruh populasi, maka hasil yang diperoleh adalah ukuran-ukuran populasi yang disebut parameter.

Ruang lingkup kajian statistika deskriptif meliputi berbagai teknik penyajian data, seperti:

Tentu, sebagai ahli statistik, saya akan menjelaskan secara rinci konsep-konsep tersebut:

a. Distribusi Frekuensi:

Distribusi frekuensi adalah pengelompokan data ke dalam beberapa kategori atau kelas yang saling terpisah, dan menunjukkan berapa kali setiap kategori muncul dalam kumpulan data.

- 1) Tujuannya adalah untuk memberikan gambaran yang jelas dan ringkas tentang bagaimana data tersebar.
- 2) Komponennya adalah kelas atau kategori: Rentang nilai atau kelompok data. Frekuensi: Jumlah kemunculan data dalam setiap kelas.
- 3) Manfaat
  - a) Memudahkan pemahaman tentang pola data.
  - b) Mengidentifikasi nilai-nilai yang sering muncul atau jarang muncul.
  - c) Menyajikan data dalam bentuk yang lebih mudah dibaca.

b. Tabel:

Tabel adalah cara untuk menyajikan data dalam format baris dan kolom. Setiap baris dan kolom mewakili kategori atau variabel tertentu.

1) Jenis:

Tabel frekuensi berfungsi menampilkan distribusi frekuensi data. Tabel kontingensi berfungsi menampilkan hubungan antara dua atau lebih variabel kategorikal.

- 2) Manfaat:
  - a) Menyajikan data secara terstruktur dan sistematis.
  - b) Memudahkan perbandingan antar kategori.
  - c) Memungkinkan analisis data yang lebih mendalam.
- c. Grafik adalah memvisualisasikan Data dalam Bentuk Diagram
  - 1) Pengertian:

Grafik adalah representasi visual dari data. Tujuannya adalah untuk membuat data lebih mudah dipahami dan diinterpretasikan.
  - 2) Jenis:
    - a) Histogram: Menampilkan distribusi frekuensi data numerik.
    - b) Diagram batang: Membandingkan nilai antar kategori.
    - c) Diagram lingkaran: Menampilkan proporsi data dalam bentuk persentase.
    - d) Diagram garis: Menampilkan perubahan data dari waktu ke waktu.
  - 3) Manfaat:
    - a) Menyajikan data secara visual yang menarik.
    - b) Memudahkan identifikasi pola dan tren data.
    - c) Meningkatkan pemahaman tentang hubungan antar variabel.
- d. Ukuran Pusat: Menentukan Nilai Rata-rata, Median, dan Modus
  - 1) Pengertian merupakan ukuran pusat adalah nilai yang mewakili nilai tipikal atau rata-rata dari suatu kumpulan data.
  - 2) Jenis:
    - a) Rata-rata (mean): Jumlah semua nilai dibagi dengan jumlah data.
    - b) Median: Nilai tengah dalam kumpulan data yang diurutkan.
    - c) Modus: Nilai yang paling sering muncul dalam kumpulan data.

- 3) Manfaat:
  - a) Memberikan gambaran tentang nilai tipikal dari suatu data.
  - b) Memudahkan perbandingan antar kumpulan data.
  - c) Berguna dalam menganalisa data.
- e. Ukuran Letak: Menentukan Posisi Data dalam Suatu Kumpulan Data
  - 1) Pengertian: Ukuran letak adalah nilai yang menunjukkan posisi relatif suatu data dalam kumpulan data yang diurutkan.
  - 2) Jenis:
    - a) Kuartil: Membagi data menjadi empat bagian yang sama.
    - b) Desil: Membagi data menjadi sepuluh bagian yang sama.
    - c) Persentil: Membagi data menjadi seratus bagian yang sama.
  - 3) Manfaat:
    - a) Menentukan posisi relatif suatu data dalam distribusi.
    - b) Mengidentifikasi nilai-nilai ekstrem atau pencilan.
    - c) Berguna dalam menganalisa distribusi data.
- f. Ukuran Simpangan: Mengukur Seberapa Jauh Data Menyebarkan dari Nilai Pusat
  - 1) Pengertian:
 

Ukuran simpangan adalah ukuran yang menunjukkan seberapa jauh data menyebarkan dari nilai pusat (biasanya rata-rata).
  - 2) Jenis:
    - a) Rentang (*range*): Selisih antara nilai maksimum dan minimum.
    - b) Variansi: Rata-rata kuadrat simpangan dari rata-rata.
    - c) Simpangan baku (*standard deviation*): Akar kuadrat dari variansi.

3) Manfaat:

- a) Memberikan gambaran tentang variabilitas data.
- b) Menilai seberapa representatif nilai pusat terhadap data secara keseluruhan.
- c) Berguna dalam menganalisa variabilitas data.

## 2. Statistik Inferensial

Statistika inferensial adalah cabang ilmu statistika yang fokus pada penggunaan data sampel untuk membuat kesimpulan, perkiraan, dan prediksi tentang populasi yang lebih besar. Proses ini melibatkan analisis data dari sampel yang dipilih secara acak untuk menggeneralisasi temuan ke seluruh populasi. Kegiatan utama dalam statistika inferensial mencakup pengujian hipotesis, estimasi parameter populasi, dan pengambilan keputusan berdasarkan hasil analisis. Beberapa teknik yang umum digunakan dalam statistika inferensial antara lain analisis korelasi, pengujian rata-rata, analisis regresi linier sederhana, analisis varians, dan analisis kovarians.

### 1.2.2 Jenis Statistika berdasarkan variabel

#### 1. Statistika Univariat

Metode analisa statistika dimana dalam penganalisaanya hanya melibatkan satu variabel yang menjadi hasil dari berapapun banyaknya variabel yang menjadi penyebab.

#### 2. Statistika Bivariat

Metode analisa statistika yang berfokus pada hubungan antara dua variabel terikat, tanpa mempedulikan jumlah variabel bebas yang mungkin mempengaruhinya.

#### 3. Statistika Multivariat

Metode statistik yang digunakan untuk meneliti hubungan antara tiga atau lebih variabel, terlepas dari jumlah variabel bebas yang ada."

### 1.2.3 Jenis Statistika berdasarkan parameter

#### 1. Statistik Parametrik

Statistik parametrik adalah metode analisis yang mengasumsikan data terdistribusi normal dan memiliki variansi yang sama.



Namun, dalam praktik, asumsi-asumsi ini sering kali tidak begitu ketat. Uji t tetap valid kecuali jika distribusi data sangat tidak normal, dan homogenitas variansi tidak masalah jika ukuran kelompok sama.

## **2. Statistik Non Parametrik**

Alat analisis yang digunakan saat parameter populasi atau asumsi distribusi populasi data tidak mengikuti pola distribusi tertentu, atau tidak terikat pada distribusi tertentu. Selain itu, variansi data tidak harus sama. Inilah sebabnya mengapa statistika nonparametrik sering disebut sebagai statistika yang tidak bergantung pada distribusi.

### **1.3 Jenis-jenis data**

Salah satu cara menentukan uji statistika yang digunakan ialah melihat jenis data. Berikut ini jenis-jenis data berdasarkan bentuknya, skalanya dan sumbernya.

#### **1.3.1 Jenis Statistika berdasarkan bentuknya**

1. Data kuantitatif: data yang dinyatakan dalam bentuk angka. Contoh: lama bekerja, jumlah gaji, usia, dan hasil ulangan.
2. Data kualitatif: Data yang dinyatakan dalam bentuk bukan angka. Contoh: jenis pekerjaan, status marital, dan tingkat kepuasan kerja.

#### **1.3.2 Jenis Statistika berdasarkan skala**

1. Data nominal adalah data yang diperoleh dengan cara kategorisasi atau klasifikasi. ciri: posisi data setara, tidak bisa dilakukan operasi matematika (+, -, x, :). Contoh: jenis kelamin, jenis pekerjaan.
2. Data ordinal adalah data yang diperoleh dengan cara kategorisasi atau klasifikasi, tetapi di antara data tersebut terdapat hubungan. Ciri: posisi data tidak setara, tidak bisa dilakukan operasi matematika (+, -, x, :). Contoh: kepuasan kerja, motivasi.
3. Data rasio adalah data yang diperoleh dengan cara pengukuran, di mana jarak antara dua titik skala sudah diketahui dan mempunyai titik 0 absolut. Ciri: tidak ada kategorisasi bisa

- dilakukan operasi matematika. Contoh: gaji, skor ujian, dan jumlah buku.
4. Data interval adalah data yang diperoleh dengan cara pengukuran, di mana jarak antara dua titik skala sudah diketahui. Ciri: Tidak ada kategorisasi, bisa dilakukan operasi matematika. Contoh: temperatur yang diukur berdasarkan OC dan OF, system kalender dan lain-lain.

### **1.3.3 Jenis Statistika berdasarkan sumbernya**

1. Data PRIMER adalah data yang berasal langsung dari sumbernya, diperoleh dengan cara wawancara, observasi dan lain sebagainya.
2. Data SEKUNDER adalah data yang tidak langsung diperoleh dari sumbernya, tetapi melalui departemen, lembaga dan lain sebagainya seperti BPS, Sekolah dan Bank.

## DAFTAR PUSTAKA

- Achi Rinaldi, Novalia, Muhamad Syazali. 2020. Statistik Inferensial untuk Pendidikan dan Sosial. Edisi Pertama. Jilid 1. Bogor: IPB Press
- Sugiono. 2012. Statistik untuk Penelitian. Cetakan ke 21. Bandung: Alfabeta.
- Sugiono. Metode Penelitian Kuantitatif. 2022. Bandung: Alfabeta.
- Sutanto Priyo Hastono, Luknis Sabri. 2011. Cetakan ke 6. Jakarta: Kharisma Putra Utama Offiset.

# BAB 2

## DATA DALAM STATISTIKA

Oleh M. Ahkam Alwi

### 2.1 Pendahuluan

Data adalah sekumpulan fakta, angka, atau informasi yang dikumpulkan, dianalisis, dan digunakan untuk keperluan tertentu. Dalam statistika, data berperan sebagai dasar dalam pengambilan keputusan dan analisis fenomena. Data memiliki peran penting dalam pengambilan keputusan dan analisis di berbagai bidang, berfungsi sebagai dasar untuk memahami fenomena melalui metode empiris.

### 2.2 Jenis-Jenis Data

#### 2.2.1 Berdasarkan Sifatnya

1. **Data Kualitatif:** Data yang tidak berbentuk angka, melainkan kategori atau deskripsi. Contoh: warna mata, jenis kelamin, status pernikahan. Data kualitatif merupakan informasi non-numerik yang digunakan untuk mengategorikan atau menggambarkan karakteristik tertentu, seperti warna mata, jenis kelamin, dan status pernikahan. Jenis data ini memiliki peran penting dalam berbagai bidang penelitian karena dapat memberikan wawasan mendalam terkait perilaku, keyakinan, serta faktor kontekstual yang memengaruhi individu atau kelompok.

Data kualitatif terdiri dari dua jenis utama. **Data deskriptif** mencakup atribut kategoris non-numerik, seperti warna mata dan jenis kelamin, untuk mengelompokkan individu berdasarkan karakteristik tertentu. **Data kontekstual** menggambarkan keadaan atau situasi seseorang, seperti status pernikahan, yang dapat memengaruhi perilaku dan persepsi dalam konteks sosial tertentu). (Chandrasekar et al., 2024).

2. **Data Kuantitatif:** Data Kuantitatif adalah data berbentuk angka yang dapat diukur, seperti tinggi badan, berat badan, dan jumlah siswa. Data ini memungkinkan analisis akurat terhadap berbagai

fenomena, mendukung penelitian dalam ilmu sosial dengan memfasilitasi pemeriksaan hubungan antar variabel. Dengan pendekatan berbasis angka, data kuantitatif memberikan analisis objektif dan sistematis untuk pengambilan keputusan berbasis bukti.

Dalam kesehatan, data kuantitatif digunakan untuk mengukur kesehatan gigi dan ketahanan hidup pasien kanker. Survei kuantitatif, seperti OHIP-14 dan GOHAI, membantu menilai kondisi kesehatan gigi berdasarkan laporan individu, memahami tren kesehatan, serta merancang strategi intervensi yang lebih efektif (Maida et al., 2022)

Di konservasi, data kuantitatif menilai keanekaragaman hayati dan menentukan prioritas perlindungan. Pendekatan berbasis geografi kuantitatif mengidentifikasi ancaman terhadap biodiversitas serta merancang langkah-langkah konservasi yang efektif, mendukung kebijakan berbasis bukti dan pemetaan area yang membutuhkan perlindungan mendesak (Minin et al., 2022).

Dalam ilmu olahraga, data kuantitatif digunakan untuk meneliti cedera, menganalisis pola, dan merancang strategi pencegahan. Pendekatan ini semakin kuat dengan integrasi penelitian kualitatif dan kuantitatif untuk memahami faktor risiko dan solusi pencegahan yang lebih efektif (Sutter et al., 2023).

## **2.2.2 Berdasarkan Sumbernya**

### **1. Data Primer**

Data primer adalah informasi yang dikumpulkan langsung oleh peneliti melalui wawancara, kuesioner, eksperimen, atau diskusi kelompok terfokus (FGD). Data ini krusial dalam penelitian karena meningkatkan kualitas dan keandalan hasil. Studi di Universitas Buea menunjukkan bahwa lebih dari 90% responden mengakui signifikansinya dalam penyusunan disertasi (Loveline, 2024). Integrasi data primer memastikan ketelitian ilmiah dan meningkatkan kredibilitas penelitian (Loveline, 2024).

### **2. Data Sekunder**

Data sekunder adalah informasi yang diperoleh dari sumber lain, seperti laporan, jurnal, atau arsip yang sudah tersedia. Data ini menjadi aset penting dalam penelitian karena

memungkinkan analisis tanpa perlu pengumpulan langsung, menghemat waktu dan sumber daya (Davis-Kean et al., 2015; Spurlock, 2020).

Salah satu keunggulan utama data sekunder adalah efisiensinya. Peneliti dapat melakukan analisis longitudinal atau mengkaji fenomena langka tanpa menunggu data baru (Davis-Kean et al., 2015). Dalam kesehatan, rekam medis elektronik (EHR) sering digunakan untuk mendukung kebijakan dan meningkatkan kualitas layanan (Scott et al., 2017; Speedie et al., 2016).

Namun, tantangan dalam analisis data sekunder mencakup kualitas dan relevansi data, termasuk inkonsistensi serta keterbatasan variabel yang tersedia (Riley et al., 2022). Dalam pendidikan, literasi data semakin ditekankan, di mana pemanfaatan data sekunder membantu pembelajaran lebih efektif. Pendidik diharapkan membekali siswa dengan keterampilan analisis data dari berbagai sumber, seperti laporan penelitian dan basis data pemerintah (Gebre, 2022).

## **2.2.3 Berdasarkan Skala Pengukurannya**

### **1. Skala Nominal:**

Data dalam bentuk kategori tanpa urutan tertentu. Contoh: jenis kelamin (laki-laki/perempuan). Skala nominal merupakan jenis data yang digunakan untuk mengelompokkan objek, individu, atau fenomena ke dalam kategori tertentu tanpa adanya hierarki atau peringkat di antara kategori tersebut. Salah satu contoh umum dari data nominal adalah jenis kelamin, yang dikelompokkan ke dalam dua kategori, yaitu laki-laki dan perempuan. Dalam bidang statistik, skala nominal memiliki peran yang signifikan dalam proses pengumpulan serta analisis data, terutama dalam penelitian sosial dan perilaku manusia, di mana pengelompokan data sering kali diperlukan untuk memahami berbagai variabel dalam masyarakat (Razak et al., 2024).

### **2. Skala Ordinal:**

Skala ordinal adalah bentuk data kategori yang memiliki tingkatan atau urutan, tetapi tanpa jarak seragam antar tingkatannya. Contohnya adalah tingkat kepuasan pelanggan,

seperti *sangat puas*, *puas*, *kurang puas*, dan *tidak puas*, yang menunjukkan urutan relatif tanpa perbedaan kuantitatif yang tetap (Wang et al., 2022).

Skala ini memungkinkan peneliti mengevaluasi pengaruh variabel independen terhadap kategori tertentu, memberikan pemahaman yang lebih dalam dibandingkan analisis biner (Youdom & Basco, 2021).

Keunggulan utama skala ordinal adalah kemampuannya menangkap perbedaan halus dalam data, menjadikannya relevan di berbagai bidang, seperti penelitian kesehatan dan pemasaran. Dalam studi kepuasan pelanggan, skala ini tidak hanya mengidentifikasi kepuasan, tetapi juga mengukur tingkatannya secara lebih spesifik, membantu dalam perancangan strategi peningkatan layanan yang lebih efektif (Keeble et al., 2016)

### **3. Skala Interval**

Skala interval adalah jenis pengukuran numerik dengan urutan yang jelas tetapi tanpa nol absolut. Contohnya suhu dalam Celsius, di mana nol tidak menunjukkan ketiadaan suhu secara mutlak (Huang et al., 2020). Berbeda dengan skala rasio, skala ini memungkinkan perbandingan antar nilai tetapi tidak mendukung perhitungan rasio yang bermakna.

Skala interval banyak digunakan dalam penelitian psikologi, pendidikan, dan pengukuran fisik. Misalnya, dalam studi suhu lingkungan, skala ini memungkinkan perbandingan suhu antar lokasi atau periode tertentu tanpa memperhitungkan arti nol dalam skala tersebut (Liu et al., 2019; Loureiro & Boldyrev, 2017). Selain itu, skor IQ dan hasil tes akademik juga menggunakan skala ini, di mana selisih skor bermakna meskipun nol tidak bersifat absolut (Torquato et al., 2019). Dengan memahami karakteristik dan keterbatasannya, skala interval dapat dioptimalkan untuk analisis data yang lebih akurat dalam berbagai konteks penelitian dan pengambilan keputusan.

### **4. Skala Rasio**

Skala rasio adalah data numerik dengan nol absolut dan perbandingan bermakna, seperti berat dan tinggi badan. Skala ini penting dalam nutrisi, kedokteran olahraga, dan studi pertumbuhan, di mana pengukuran akurat mendukung analisis

yang tepat. Nol absolut memungkinkan operasi matematis seperti perkalian dan pembagian, yang tidak dapat dilakukan pada skala ordinal atau interval (Waterworth et al., 2022)

Dalam kedokteran olahraga, skala rasio digunakan untuk menormalkan pengukuran, seperti rasio kekuatan hamstring terhadap quadriceps dalam evaluasi performa atlet (Atkinson & Batterham, 2012). Namun, interpretasi data perlu hati-hati, karena metode koreksi rasio mungkin tidak efektif dalam mengatasi variabel pengganggu. Pendekatan statistik alternatif, seperti ANCOVA, sering kali lebih sesuai dalam kondisi tertentu (Karp et al., 2012)

Meskipun skala rasio menyediakan kerangka pengukuran yang kuat, penggunaannya harus memperhitungkan asumsi statistik dan potensi bias. Oleh karena itu, ketelitian metodologi sangat penting untuk menghasilkan kesimpulan yang valid.

## **2.3 Metode Pengumpulan Data**

Beberapa metode umum dalam mengumpulkan data:

### **1. Observasi**

Observasi adalah teknik penelitian yang melibatkan pengamatan langsung terhadap fenomena dalam lingkungan alaminya. Metode ini memungkinkan peneliti memperoleh data relevan dengan mengamati perilaku, interaksi, dan konteks situasi subjek (Fadli, 2021). Observasi sering digunakan dalam penelitian kualitatif karena memberikan pemahaman mendalam mengenai fenomena sosial dan perilaku manusia (Rahayu et al., 2024).

Dalam penelitian, observasi dilakukan secara sistematis dan terdokumentasi untuk memastikan konsistensi data. Metode ini efektif dalam memahami dinamika sosial dibandingkan survei atau kuesioner (Rahayu et al., 2024).

Contohnya, dalam studi sosial, observasi digunakan untuk menganalisis perilaku anak saat pembelajaran, termasuk interaksi siswa dan respons terhadap materi (Dwirian & Dwiastuti, 2021).



Selain itu, observasi membantu mengidentifikasi pola atau perubahan yang tidak terdeteksi melalui metode lain (Zulaiha & Rohman, 2020). Misalnya, penelitian tentang pengaruh media sosial terhadap interaksi sosial dapat menggunakan observasi untuk mengamati perubahan pola komunikasi siswa (Rahayu et al., 2024).

Secara keseluruhan, observasi adalah metode penting dalam penelitian kualitatif karena menghasilkan data kontekstual dan realistis, memungkinkan analisis yang lebih mendalam serta kesimpulan yang lebih valid.

## **2. Wawancara**

Mengumpulkan data melalui pertanyaan langsung kepada responden. Wawancara merupakan teknik pengumpulan data yang melibatkan komunikasi langsung antara peneliti dan responden melalui pertanyaan terbuka maupun tertutup. Metode ini berperan penting dalam penelitian kualitatif karena memungkinkan eksplorasi mendalam terhadap pandangan, pengalaman, serta persepsi responden. Dalam wawancara, peneliti menyusun pertanyaan yang dirancang untuk mendorong diskusi guna memperoleh informasi yang kaya dan kontekstual (Kallio et al., 2016).

Wawancara dapat dilakukan secara langsung, melalui telepon, atau daring via platform seperti Zoom atau Skype, masing-masing dengan kelebihan dan kekurangannya. Wawancara tatap muka lebih efektif dalam menggali informasi mendalam karena interaksi langsung membangun hubungan lebih kuat antara peneliti dan responden (Khan & MacEachen, 2022). Namun, dalam situasi tertentu, seperti pandemi COVID-19, wawancara daring menjadi alternatif yang lebih aman dan praktis, meskipun tantangan teknis dan privasi tetap perlu diperhatikan (Nair et al., 2022).

Berdasarkan tingkat strukturnya, wawancara dapat dikategorikan menjadi tiga jenis utama:

- a. Wawancara Terstruktur: Menggunakan daftar pertanyaan yang telah ditentukan sebelumnya untuk memastikan konsistensi dalam pengumpulan data.

- b. Wawancara Semi-Terstruktur: Menggabungkan pertanyaan yang telah disiapkan dengan fleksibilitas untuk mengeksplorasi topik yang lebih luas sesuai dengan respons responden.
- c. Wawancara Tidak Terstruktur: Tidak memiliki format pertanyaan yang kaku, memungkinkan responden berbicara lebih bebas mengenai topik yang dianggap relevan (Hotaman, 2020).

Pemilihan jenis wawancara yang tepat bergantung pada tujuan penelitian dan tingkat kebebasan yang ingin diberikan kepada responden dalam berbagi pengalaman mereka. Meskipun wawancara efektif untuk mengumpulkan data mendalam, tantangannya termasuk potensi bias pewawancara yang dapat memengaruhi responden. Oleh karena itu, standar etika seperti transparansi tujuan penelitian dan menjaga kerahasiaan data perlu diterapkan. (Iacono et al., 2016). Selain itu, analisis data wawancara harus dilakukan secara cermat agar interpretasi hasil tetap objektif dan akurat.

Secara keseluruhan, wawancara merupakan metode yang sangat berharga dalam penelitian kualitatif. Dengan perencanaan yang matang dan pendekatan yang sesuai, wawancara dapat membantu peneliti memperoleh wawasan mendalam mengenai subjek penelitian, sehingga berkontribusi pada pengembangan ilmu pengetahuan di berbagai bidang.

### 3. Kuesioner

Menggunakan formulir pertanyaan untuk memperoleh informasi dari banyak responden. Kuesioner adalah instrumen yang banyak digunakan dalam penelitian untuk memperoleh data dari sejumlah besar responden melalui serangkaian pertanyaan yang terstruktur. Metode ini memiliki peran penting dalam penelitian empiris karena kemampuannya dalam mengumpulkan informasi secara efisien dan sistematis dari populasi yang luas (Aithal & Aithal, 2020). Dalam penyusunan kuesioner, aspek seperti format, jenis pertanyaan, serta teknik validasi menjadi

faktor utama yang harus diperhatikan agar data yang diperoleh akurat dan dapat diandalkan (Aithal & Aithal, 2020).

Salah satu aspek krusial dalam penggunaan kuesioner adalah pemilihan serta penyesuaian pertanyaan yang relevan dengan konteks penelitian. Ini mencakup proses penerjemahan dan adaptasi kuesioner yang telah divalidasi agar sesuai dengan bahasa atau budaya yang berbeda (Kalfoss, 2019). Sebagai contoh, dalam penelitian mengenai kebutuhan informasi dan kesehatan, validitas kuesioner menjadi faktor penting untuk memastikan bahwa hasil yang diperoleh benar-benar mencerminkan opini serta perilaku responden (Dam et al., 2022).

Kuesioner dapat berisi berbagai tipe pertanyaan, mulai dari pertanyaan terbuka yang memungkinkan responden memberikan jawaban secara bebas hingga pertanyaan tertutup yang mengharuskan mereka memilih dari sejumlah opsi yang telah ditentukan. Pemilihan jenis pertanyaan ini akan mempengaruhi kualitas data yang dikumpulkan (Aithal & Aithal, 2020). Dengan strategi pengumpulan data yang tepat, kuesioner dapat memberikan wawasan mendalam mengenai berbagai aspek, seperti kepuasan, sikap, dan perilaku responden dalam bidang kesehatan, pendidikan, serta pemasaran (Aithal & Aithal, 2020).

Selain sebagai alat pengumpulan data, kuesioner yang dirancang dengan baik juga dapat membangun kepercayaan antara peneliti dan responden. Dalam penelitian yang mencakup isu-isu sensitif seperti kesehatan mental atau pengalaman pribadi, menciptakan lingkungan yang nyaman bagi responden agar mereka merasa aman dalam memberikan informasi merupakan aspek yang sangat penting (Roodhof & Veen, 2022). Pendekatan yang empatik dan komunikasi yang baik dari peneliti dapat meningkatkan validitas serta keandalan data yang diperoleh (Roodhof & Veen, 2022).

Kuesioner adalah instrumen penting dalam penelitian sosial dan ilmu data, memungkinkan pengumpulan informasi secara sistematis dari populasi luas. Data yang diperoleh dapat dianalisis untuk mengidentifikasi pola, tren, dan wawasan yang mendukung pengambilan keputusan berbasis bukti.

#### 4. Eksperimen

Melakukan percobaan terkontrol untuk mengumpulkan data berdasarkan variabel tertentu. Eksperimen adalah metode penelitian yang melibatkan pelaksanaan percobaan terkontrol untuk mengumpulkan data berdasarkan variabel tertentu. Dalam konteks ini, eksperimen bertujuan untuk mengeksplorasi hubungan sebab akibat dengan menguji hipotesis yang telah ditentukan sebelumnya. Melalui pengaturan yang ketat dan pengendalian variabel, peneliti bisa menilai pengaruh satu atau lebih variabel bebas terhadap variabel tergantung (Keusch et al., 2019).

Keuntungan utama eksperimen adalah kemampuannya menghasilkan data yang andal dan dapat direplikasi. Dalam psikologi dan ilmu kesehatan, eksperimen memberikan wawasan mendalam tentang perilaku manusia serta respons fisiologis terhadap stimulasi tertentu. Misalnya, penelitian tentang pengalaman pasien dalam konteks layanan kesehatan sering menggunakan metode eksperimen untuk mengukur hasil dari berbagai intervensi perawatan (Greenhalgh et al., 2017). Dengan desain yang cermat, peneliti dapat mengisolasi variabel yang memengaruhi kepuasan dan efikasi layanan serta mengidentifikasi area perbaikan. Eksperimen modern semakin efisien dengan teknologi canggih, seperti perangkat lunak untuk desain dan pengumpulan data otomatis. Sebuah penelitian yang menggunakan sistem berbasis aplikasi smartphone untuk pengumpulan data menunjukkan bagaimana kemajuan teknologi dapat meningkatkan partisipasi dan kemudahan pengumpulan data (Ságvári et al., 2021).

Pendekatan ini memungkinkan peneliti mengumpulkan data secara real-time dengan cakupan lebih luas, meningkatkan keberagaman dan jumlah responden. Namun, eksperimen juga menghadapi tantangan, seperti potensi bias dalam pengumpulan data. Faktor-faktor seperti identitas responden dapat mempengaruhi keakuratan data yang dikumpulkan melalui survei atau eksperimen (Brenner & DeLamater, 2016).

Untuk mengatasi tantangan ini, peneliti harus merancang eksperimen dan metode pengumpulan data dengan cermat agar

hasilnya valid dan representatif. Kepatuhan terhadap **etika penelitian** juga krusial, termasuk memperoleh persetujuan peserta dan menjaga transparansi, terutama dalam bidang yang melibatkan informasi sensitif seperti kesehatan (Greenhalgh et al., 2017).

**Eksperimen** merupakan metode pengumpulan data yang kuat dalam penelitian ilmiah, memberikan wawasan signifikan jika dilakukan dengan pendekatan yang tepat dan etis.

## 2.4 Penyajian Data

Setelah data dikumpulkan, langkah selanjutnya adalah menyajikan data agar lebih mudah dipahami. Beberapa bentuk penyajian data antara lain:

### 2.4.1 Tabel

Tabel menyusun data dalam baris dan kolom untuk memberikan struktur yang jelas, memudahkan analisis, perbandingan, dan interpretasi. Alat ini banyak digunakan di berbagai bidang seperti sains, bisnis, pendidikan, dan penelitian untuk menyajikan informasi kompleks secara efisien(Lenzen et al., 2014).

**Tabel 2.1.** Kepuasan Pelanggan

Kategori Usia	Laki-laki %	Perempuan %
18-25	78	80
26-35	82	85
36-45	75	78
46-55	68	70
56+	60	65

Contoh Tabel Kepuasan pelanggan berdasarkan usia dan jenis kelamin

Salah satu fungsi utama tabel dalam pengumpulan data adalah untuk menunjukkan hubungan antar data serta memberikan

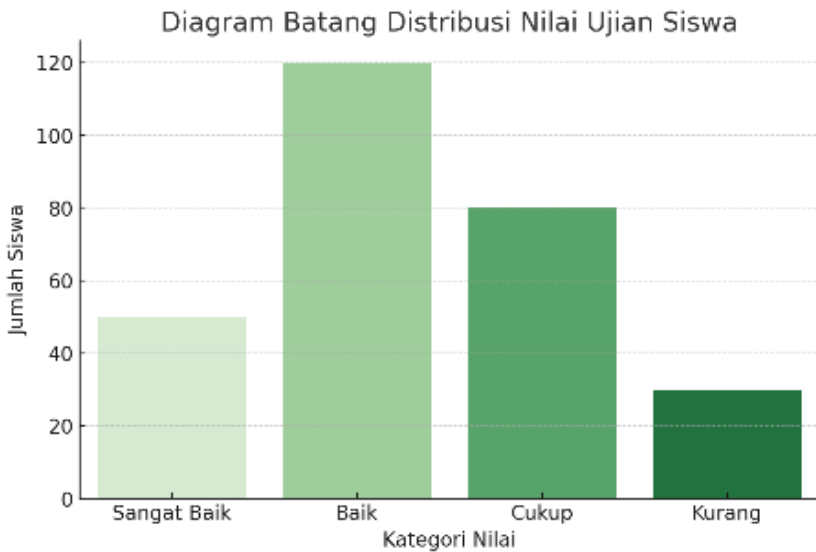
konteks bagi informasi yang disajikan. Dalam penelitian, misalnya, pengorganisasian data dalam bentuk tabel mempermudah peneliti dalam membandingkan kategori atau kelompok serta menginterpretasikan hasil dengan lebih cepat (Zheng et al., 2021). Selain itu, tabel sering digunakan untuk menampilkan hasil analisis statistik, di mana data dikelompokkan berdasarkan kriteria tertentu, seperti lokasi, periode waktu, atau variabel lainnya (Lenzen et al., 2017).

Tabel menyederhanakan informasi kompleks, memungkinkan penyajian data yang ringkas dan mudah dibandingkan. Peneliti menggunakannya untuk mempermudah analisis, seperti dalam studi kepuasan pelanggan, di mana tabel membantu menampilkan peringkat kepuasan berdasarkan faktor seperti usia atau jenis kelamin, memberikan gambaran yang lebih komprehensif (Novieto et al., 2023).

#### **2.4.2 Diagram dan Grafik**

Diagram Batang menyajikan data dalam bentuk batang vertikal atau horizontal untuk membandingkan kategori. Metode visualisasi ini efektif dalam menampilkan perbedaan nilai antar kategori, dengan panjang atau tinggi batang mewakili data yang diukur. Dalam pendidikan dan analisis data, diagram batang digunakan untuk menyederhanakan informasi kompleks, memungkinkan perbandingan visual yang lebih jelas dan ringkas (Sugiyono, 2023).

Diagram batang sering digunakan dalam berbagai penelitian untuk menyajikan data secara lebih mudah dipahami. Misalnya, dalam studi tentang kepuasan pelanggan, kategori seperti "sangat puas," "puas," "kurang puas," dan "tidak puas" dapat ditampilkan dalam bentuk batang, di mana tinggi batang menunjukkan jumlah responden dalam setiap kategori. Representasi ini mempermudah analisis dan memberikan pemahaman yang lebih mendalam terkait persepsi responden (Hisyam & Dini, 2023).



**Gambar 2.1.** Contoh diagram batang Distribusi Nilai Ujian Siswa

Selain itu, diagram batang juga kerap diterapkan dalam evaluasi hasil ujian, di mana distribusi nilai dapat dibandingkan dalam satu tampilan grafis. Dengan menggunakan diagram batang, peneliti dapat mengidentifikasi pola distribusi nilai secara lebih cepat, sehingga memudahkan interpretasi terhadap tren akademik suatu kelompok (AZKIYA et al., 2024).

Diagram batang adalah alat visualisasi yang efektif dalam berbagai bidang penelitian, membantu menyederhanakan informasi kompleks dan memperjelas perbandingan data. Dengan tampilan yang mudah dipahami, diagram ini mempermudah interpretasi serta meningkatkan pemahaman terhadap hasil analisis

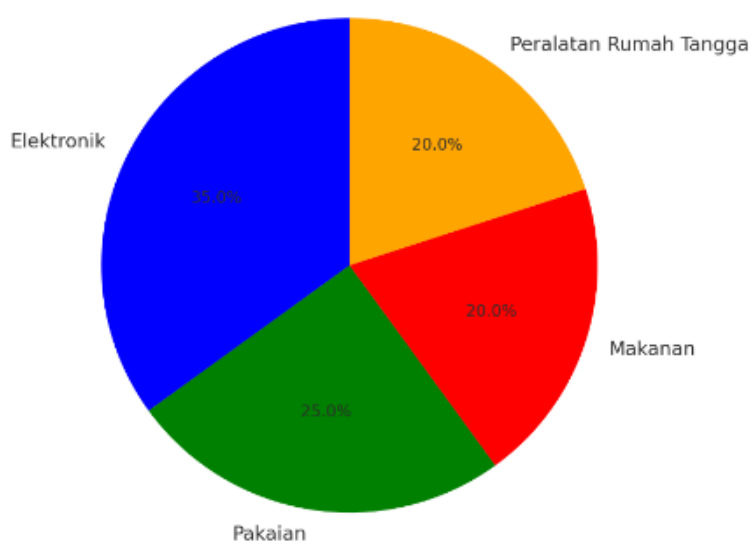
#### 2.4.3 Diagram Lingkaran

Diagram lingkaran atau *diagram pai* menampilkan data dalam bentuk sektor untuk menunjukkan proporsi tiap kategori dalam satu lingkaran. Setiap sektor mewakili bagian dari keseluruhan, memudahkan pengguna dalam memahami hubungan relatif antar kategori, dengan total sudut selalu mencapai 360 derajat (Ramadanti et al., 2021). Diagram ini banyak digunakan dalam pendidikan dan bisnis. Dalam pengajaran matematika, diagram lingkaran membantu

siswa memahami konsep proporsi dan representasi data, meskipun mereka sering menghadapi tantangan dalam menganalisis serta mengelompokkan data sebelum divisualisasikan (Ramadanti et al., 2021).

Dalam bisnis, diagram lingkaran digunakan untuk menyajikan pembagian pasar atau proporsi pendapatan berdasarkan kategori produk atau layanan. Misalnya, dalam analisis industri, diagram ini menggambarkan persentase penjualan berbagai produk dalam sebuah perusahaan, membantu mengidentifikasi kontribusi terbesar terhadap pendapatan keseluruhan (Kosasi et al., 2022). Setiap sektor biasanya dilengkapi dengan label dan persentase untuk mempermudah interpretasi.

Diagram Lingkaran Proporsi Penjualan Produk



Gambar 2.2. Contoh diagram Lingkaran Proporsi Penjualan Produk

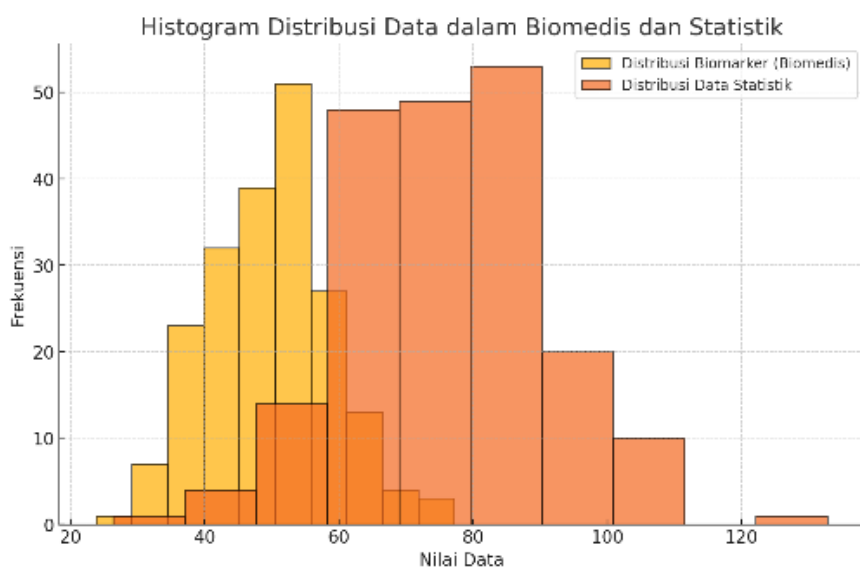
Secara keseluruhan, diagram lingkaran adalah alat visualisasi efektif untuk menyajikan data numerik secara jelas. Dengan penggunaan yang tepat, baik siswa maupun profesional dapat memanfaatkannya untuk menyederhanakan dan mengomunikasikan informasi kompleks dalam format yang intuitif dan mudah dipahami.



## 2.4.4 Histogram

Histogram menyajikan data dalam batang berkesinambungan untuk menggambarkan distribusi frekuensi, memudahkan identifikasi pola, tren, dan outlier. Alat ini banyak digunakan dalam penelitian ilmiah, teknik, dan industri, seperti analisis biomarker dalam biomedis serta evaluasi distribusi data dalam statistik. Tantangan utama terletak pada pemilihan ukuran bin, yang memengaruhi akurasi interpretasi dan identifikasi subpopulasi dalam data.

Histogram menampilkan data dalam batang berkesinambungan untuk distribusi frekuensi, membantu mengidentifikasi pola, tren, dan outlier. Alat ini digunakan dalam penelitian ilmiah, teknik, dan industri, seperti analisis biomarker dalam biomedis dan evaluasi distribusi data dalam statistik. Tantangan utama adalah pemilihan ukuran bin, yang memengaruhi akurasi interpretasi dan identifikasi subpopulasi dalam data.

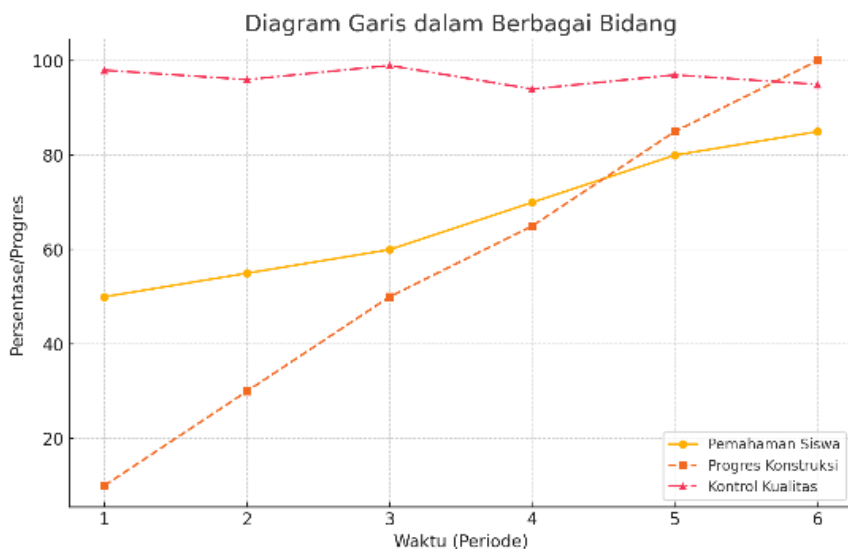


Berikut adalah histogram yang menunjukkan distribusi biomarker dalam biomedis dan data statistik. Diagram ini membantu memahami pola distribusi, frekuensi data dalam interval tertentu, serta mendeteksi outlier.

## 2.4.5 Diagram Garis

Diagram garis digunakan untuk menunjukkan tren atau perubahan data dalam rentang waktu tertentu. Representasi visual ini membantu mengidentifikasi pola, variasi, dan titik penting dalam data, memberikan pemahaman lebih jelas dibandingkan tabel statis.

Dalam pendidikan, diagram garis membantu siswa memahami konsep matematika dan menyelesaikan masalah aritmetika dengan lebih efektif (Nurjanah et al., 2021). Selain itu, penggunaannya dalam sistem koordinat Kartesius mendukung analisis kinerja berdasarkan variabel pada sumbu X dan Y, meskipun aplikasinya dalam analisis hasil masih perlu dikaji lebih lanjut (Farid et al., 2023).



**Gambar 2.3.** Contoh Gambar Diagram Garis dalam berbagai bidang

Berikut adalah diagram garis yang menggambarkan:

1. Pemahaman Siswa dalam Pendidikan. Menunjukkan tren peningkatan pemahaman siswa terhadap konsep matematika seiring waktu.
2. Progres Konstruksi. Mengilustrasikan perkembangan proyek konstruksi dari awal hingga penyelesaian.
3. Kontrol Kualitas Produk. Memantau stabilitas kualitas produk dalam suatu proses produksi.

Diagram ini berfungsi sebagai alat visual yang mempermudah interpretasi data berbasis waktu, membantu dalam analisis tren, serta mendukung pengambilan keputusan berdasarkan informasi yang tersedia.

## **2.5 Kesimpulan**

Data merupakan elemen penting dalam statistika karena menjadi dasar dalam analisis dan pengambilan keputusan. Jenis data yang berbeda membutuhkan metode pengumpulan dan penyajian yang sesuai agar dapat diinterpretasikan dengan baik.

## DAFTAR PUSTAKA

- Aithal, A., & Aithal, P. S. (2020). Development and Validation of Survey Questionnaire & Experimental Data – A Systematical Review-Based Statistical Approach. *International Journal of Management Technology and Social Sciences*, 233–251. <https://doi.org/10.47992/ijmts.2581.6012.0116>
- Atkinson, G., & Batterham, A. M. (2012). The use of ratios and percentage changes in sports medicine: time for a rethink? *International Journal of Sports Medicine*, 33(07), 505–506, 33(07), 55–506.
- AZKIYA, W. R., REFFIANE, F., SELUNAWATI, K., & MUSHAFANAH, Q. (2024). Analisis Asesmen Diagnostik Tentang Gaya Belajar Untuk Pembelajaran Matematika Berdiferensiasi Di Sekolah Dasar. *Elementary Jurnal Inovasi Pendidikan Dasar*, 4(3), 150–159. <https://doi.org/10.51878/elementary.v4i3.3082>
- Brenner, P. S., & DeLamater, J. (2016). Lies, Damned Lies, and Survey Self-Reports? Identity as a Cause of Measurement Bias. *Social Psychology Quarterly*, 79(4), 333–354. <https://doi.org/10.1177/0190272516628298>
- Chandrasekar, A., Clark, S. E., Martin, S., Vanderslott, S., Flores, E. C., Aceituno, D., Barnett, P., Vindrola-Padros, C., & Vera San Juan, N. (2024). Making the most of big qualitative datasets: a living systematic review of analysis methods. *Frontiers in Big Data*, 7. <https://doi.org/10.3389/fdata.2024.1455399>
- Dam, L. T., Heidler, P., & King, I. (2022). Access, Understanding, Promotion and Maintenance of Good Health: Evaluation of Knowledge Transfer of People With Intellectual Disabilities to Bridge the Health Information and Disease Prevention in Public Health. *Frontiers in Public Health*, 10. <https://doi.org/10.3389/fpubh.2022.915970>
- Davis-Kean, P., Jager, J., & Maslowsky, J. (2015). Answering Developmental Questions Using Secondary Data. *Child Development Perspectives*, 9(4), 256–261. <https://doi.org/10.1111/cdep.12151>

- Dwirian, R., & Dwiastuti, I. (2021). Ego Transcendence Pada Anggota Majelis Ta'lim as-Sakinah Yang Telah Lanjut Usia. *Flourishing Journal*, 1(4), 310–314. <https://doi.org/10.17977/um070vli42021p310-314>
- Fadli, M. R. (2021). Memahami Desain Metode Penelitian Kualitatif. *Humanika*, 21(1), 33–54. <https://doi.org/10.21831/hum.v21i1.38075>
- Farid, A., Zahroh, H., & Sugiarti, T. (2023). Martajasah Mangrove Ecotourism Development Strategy, Bangkalan Regency, Madura. *Economic and Social of Fisheries and Marine*, 010(02), 221–235. <https://doi.org/10.21776/ub.ecsofim.2023.010.02.08>
- Gebre, E. H. (2022). Conceptions and Perspectives of Data Literacy in Secondary Education. *British Journal of Educational Technology*, 53(5), 1080–1095. <https://doi.org/10.1111/bjet.13246>
- Greenhalgh, J., Dalkin, S., Gooding, K., Gibbons, E., Wright, J., Meads, D., Black, N., Valderas, J. M., & Pawson, R. (2017). Functionality and Feedback: A Realist Synthesis of the Collation, Interpretation and Utilisation of Patient-Reported Outcome Measures Data to Improve Patient Care. *Health Services and Delivery Research*, 5(2), 1–280. <https://doi.org/10.3310/hsdr05020>
- Hisyam, F. R., & Dini, S. K. (2023). Persepsi Aparatur Sipil Negara (ASN) Kabupaten Sleman Terhadap Keamanan Data Pribadi Dengan Metode Statistika Deskriptif. *Esds*, 1(3), 263–277. <https://doi.org/10.20885/esds.vol1.iss.3.art34>
- Hotaman, D. (2020). The Effect of Formative Assessment on the Academic Achievement Levels of Prospective Teachers. *Journal of Curriculum and Teaching*, 9(3), 33. <https://doi.org/10.5430/jct.v9n3p33>
- Huang, D. Z., Xu, K., Farhat, C., & Darve, E. (2020). Learning Constitutive Relations From Indirect Observations Using Deep Neural Networks. *Journal of Computational Physics*, 416, 109491. <https://doi.org/10.1016/j.jcp.2020.109491>
- Iacono, V. L., Symonds, P. S., & Brown, D. (2016). Skype as a Tool for Qualitative Research Interviews. *Sociological Research Online*, 21(2), 103–117. <https://doi.org/10.5153/sro.3952>
- Kallio, H., Pietilä, A., Johnson, M., & Kangasniemi, M. (2016). Systematic Methodological Review: Developing a Framework for a Qualitative Semi-structured Interview Guide. *Journal of*

- Advanced Nursing*, 72(12), 2954–2965.  
<https://doi.org/10.1111/jan.13031>
- Karp, N. A., Segonds-Pichon, A., Gerdin, A. K. B., Ramírez-Solis, R., & White, J. K. (2012). The fallacy of ratio correction to address confounding factors. *Laboratory Animals*, 46(3), 245–252.  
<https://doi.org/10.1258/la.2012.012003>
- Keeble, C., Baxter, P. D., Gislason-Lee, A. J., Treadgold, L. A., & Davies, A. G. (2016). Methods for the Analysis of Ordinal Response Data in Medical Image Quality Assessment. *British Journal of Radiology*, 89(1063), 20160094.  
<https://doi.org/10.1259/bjr.20160094>
- Keusch, F., Struminskaya, B., Antoun, C., Couper, M. P., & Kreuter, F. (2019). Willingness to Participate in Passive Mobile Data Collection. *Public Opinion Quarterly*, 83(S1), 210–235.  
<https://doi.org/10.1093/poq/nfz007>
- Khan, T. H., & MacEachen, E. (2022). An Alternative Method of Interviewing: Critical Reflections on Videoconference Interviews for Qualitative Data Collection. *International Journal of Qualitative Methods*, 21.  
<https://doi.org/10.1177/16094069221090063>
- Kosasi, S., Millah, S., & Santoso, N. P. L. (2022). Manajemen Dalam Konsep Dan Prinsip Pengelolaan Pendidikan Menggunakan Komputasi Awan. *Jurnal Mentari Manajemen Pendidikan Dan Teknologi Informasi*, 1(1), 38–45.  
<https://doi.org/10.34306/mentari.v1i1.137>
- Lenzen, M., Geschke, A., Rahman, M. D. A., Xiao, Y., Fry, J., Reyes, R. C., Dietzenbacher, E., Inomata, S., Kanemoto, K., Los, B., Moran, D., Bäumen, H. S. in den, Tukker, A., Walmsley, T., Wiedmann, T., Wood, R., & Yamano, N. (2017). The Global MRIO Lab – Charting the World Economy. *Economic Systems Research*, 29(2), 158–186. <https://doi.org/10.1080/09535314.2017.1301887>
- Lenzen, M., Geschke, A., Wiedmann, T., Lane, J., Anderson, N. G., Baynes, T., Boland, J., Daniels, P., Dey, C., Fry, J., Hadjikakou, M., Kenway, S., Malik, A., Moran, D., Murray, J., Nettleton, S. J., Poruschi, L., Reynolds, C., Rowley, H. V., ... West, J. (2014). Compiling and Using Input-output Frameworks Through Collaborative Virtual Laboratories. *The Science of the Total*

- Environment*, 485–486, 241–251.  
<https://doi.org/10.1016/j.scitotenv.2014.03.062>
- Liu, L., Cao, W., Shi, B., & Tang, M. (2019). Large-Scale Green Supplier Selection Approach Under a Q-Rung Interval-Valued Orthopair Fuzzy Environment. *Processes*, 7(9), 573.  
<https://doi.org/10.3390/pr7090573>
- Loureiro, N., & Boldyrev, S. (2017). Role of Magnetic Reconnection in Magnetohydrodynamic Turbulence. *Physical Review Letters*, 118(24). <https://doi.org/10.1103/physrevlett.118.245101>
- Maida, C. A., Xiong, D., Marcus, M., Zhou, L., Huang, Y., Lyu, Y., Shen, J., Garcia, A. O., & Liu, H. (2022). Quantitative data collection approaches in subject - reported oral health research: a scoping review. *BMC Oral Health*, 1–14.  
<https://doi.org/10.1186/s12903-022-02399-5>
- Minin, E. Di, Correia, R. A., & Toivonen, T. (2022). Quantitative conservation geography. *Trends in Ecology & Evolution*, 37(1), 42–52. <https://doi.org/10.1016/j.tree.2021.08.009>
- Nair, S., Hijam, M., Alee, N. T., Haobijam, N., Kaur, H., & Rao, M. V. V. (2022). Continuing Qualitative Data Collection During the COVID 19 Pandemic: Lessons Learnt From a Study Among Tribes in Manipur, India. *National Journal of Community Medicine*, 13(08), 578–580. <https://doi.org/10.55489/njcm.130820222062>
- Novieto, D. T., Kulor, F., Apprey, M. W., & Ayeke, E. (2023). Appraisal of Students' Perceptions on Green Building Concepts In a technical University. *Frontiers in Engineering and Built Environment*, 3(2), 122–136. <https://doi.org/10.1108/febe-08-2022-0034>
- Nurjanah, M. T., Sa'dijah, C., & Susiswo, S. (2021). Representasi Skematis Siswa Dalam Menyelesaikan Masalah Trend in International Mathematics and Science Study (TIMSS) Ditinjau Dari Self Efficacy. *Jurnal Pendidikan Teori Penelitian Dan Pengembangan*, 4(4), 622.  
<https://doi.org/10.17977/jptpp.v6i4.14725>
- Rahayu, A., Pebriani, E., & Julinda, J. (2024). Dampak Media Sosial Terhadap Pola Interaksi Sosial Budaya Siswa Di SD N Talang Duku. *Jis*, 2(2), 159–170. <https://doi.org/10.59024/jis.v2i2.764>

- Ramadanti, F., Mutaqin, A., & Hendrayana, A. (2021). Pengembangan E-Modul Matematika Berbasis PBL (Problem Based Learning) Pada Materi Penyajian Data Untuk Siswa SMP. *Jurnal Cendekia Jurnal Pendidikan Matematika*, 5(3), 2733–2745. <https://doi.org/10.31004/cendekia.v5i3.759>
- Razak, M. I., Taan, H., & Niode, I. Y. (2024). Pengaruh Digital Marketing Terhadap Keputusan Pembelian Melalui Brand Awareness Pada McDonald's Gorontalo. *Cendekia*, 1(6), 291–304. <https://doi.org/10.62335/9kzpkp07>
- Roodhof, A. M., & Veen, E. J. (2022). Finding Quality in Quantitative Methods: The Case of Food Forestry in the Netherlands. *Urban Agriculture & Regional Food Systems*, 1(1). <https://doi.org/10.1002/uar2.20024>
- Ságvári, B., Gulyás, A., & Koltai, J. (2021). Attitudes Towards Participation in a Passive Data Collection Experiment. *Sensors*, 21(18), 6085. <https://doi.org/10.3390/s21186085>
- Scott, P. J., Rigby, M., Ammenwerth, E., McNair, J., Georgiou, A., Hyppönen, H., Keizer, N. d., Magrabi, F., Nykänen, P., Gude, W. T., & Hackl, W. (2017). Evaluation Considerations for Secondary Uses of Clinical Data: Principles for an Evidence-Based Approach to Policy and Implementation of Secondary Analysis. *Yearbook of Medical Informatics*, 26(01), 59–67. <https://doi.org/10.1055/s-0037-1606530>
- Speedie, S. M., Simon, G., Kumar, V., Westra, B. L., & Johnson, S. (2016). Application of an Ontology for Characterizing Data Quality for a Secondary Use of EHR Data. *Applied Clinical Informatics*, 07(01), 69–88. <https://doi.org/10.4338/aci-2015-08-ra-0107>
- Spurlock, D. (2020). Scholarship During a Pandemic: Secondary Data Analysis. *Journal of Nursing Education*, 59(5), 245–247. <https://doi.org/10.3928/01484834-20200422-02>
- Sugiyono, S. (2023). Kemampuan Representasi Visual Mahasiswa PGSD Pada Mata Kuliah Statistik Pendidikan. *Jurnal Penelitian Pendidikan*, 15(1), 1–9. <https://doi.org/10.21137/jpp.2023.15.1.1>
- Sutter, S. E. H., Gennäs, K. B., Schubring, A., Grau, S., Jungmalm, J., & Ruchti, N. B. (2023). Interdisciplinary sport injury research and the integration of qualitative and quantitative data. *BMC Medical Research Methodology*, 23(110), 1–9.



<https://doi.org/10.1186/s12874-023-01929-1>

- Torquato, S., Zhang, G., & Courcy-Ireland, M. d. (2019). Hidden Multiscale Order in the Primes. *Journal of Physics a Mathematical and Theoretical*, 52(13), 135002. <https://doi.org/10.1088/1751-8121/ab0588>
- Wang, F., Pollock, K., & Hauseman, C. (2022). Time Demands and Emotionally Draining Situations Amid Work Intensification of School Principals. *Educational Administration Quarterly*, 59(1), 112–142. <https://doi.org/10.1177/0013161x221132837>
- Waterworth, S. P., Kerr, C. J., McManus, C. J., Costello, R., & Sandercock, G. R. H. (2022). Obese individuals do not underreport dietary intake to a greater extent than nonobese individuals when data are allometrically-scaled. *American Journal of Human Biology*, 34(7), 1–11. <https://doi.org/10.1002/ajhb.23743>
- Youdom, S. W., & Basco, L. K. (2021). Methodological Approaches for Analysing Data From Therapeutic Efficacy Studies. *Malaria Journal*, 20(1). <https://doi.org/10.1186/s12936-021-03768-1>
- Zheng, H., Többen, J., Dietzenbacher, E., Moran, D., Meng, J., Wang, D., & Guan, D. (2021). Entropy-Based Chinese City-Level MRIO Table Framework. *Economic Systems Research*, 34(4), 519–544. <https://doi.org/10.1080/09535314.2021.1932764>
- Zulaiha, D., & Rohman, A. (2020). Strategi Guru Dan Keterlibatan Orangtua Dalam Pemahaman Konsep Sains Anak Selama Covid-19. *Jurnal Obsesi Jurnal Pendidikan Anak Usia Dini*, 5(2). <https://doi.org/10.31004/obsesi.v5i2.816>

# BAB 3

## PENYAJIAN DATA

Oleh Raden Lutfi Zakaria

### 3.1 Pendahuluan

Data merupakan elemen utama dalam analisis statistika. Setelah data dikumpulkan, langkah berikutnya adalah menyajikan data dalam bentuk yang lebih mudah dipahami. Penyajian data bertujuan untuk menyederhanakan, mengorganisir, serta memvisualisasikan data sehingga lebih informatif bagi pengambil keputusan (*Wickham & Grolemund, 2017*). Penyajian data dapat dilakukan melalui tabel, grafik, dan diagram.

Dalam era digital, data tersedia dalam jumlah yang sangat besar (*Big Data*), sehingga menuntut metode penyajian yang lebih efisien dan interaktif. Teknologi seperti *data visualization tools* (misalnya *Tableau* dan *Power BI*) semakin banyak digunakan untuk membantu analisis data dalam berbagai bidang, termasuk bisnis, kesehatan, dan rekayasa (*Chen et al., 2020*). Penyajian data yang baik tidak hanya membantu dalam memahami informasi, tetapi juga mendukung pengambilan keputusan yang lebih akurat.

Isu terbaru dalam statistika dasar mencakup pentingnya data yang bersih dan bebas dari bias. Banyak penelitian menunjukkan bahwa penyajian data yang kurang baik dapat menyebabkan misinterpretasi informasi, yang pada akhirnya mempengaruhi keputusan strategis. Misalnya, dalam dunia kesehatan, kesalahan dalam menampilkan data epidemiologi dapat mengarah pada kebijakan yang tidak efektif dalam menangani wabah penyakit (*World Health Organization, 2021*).

Seiring berkembangnya teknologi, metode penyajian data juga semakin beragam. Misalnya, teknik *storytelling* dalam data *visualization* kini banyak digunakan untuk menyampaikan informasi yang kompleks dengan cara yang lebih menarik dan mudah dipahami. Dengan memahami berbagai teknik penyajian data, seseorang dapat

lebih efektif dalam mengomunikasikan informasi statistik kepada audiens yang lebih luas.

### 3.2 Jenis Penyajian Data

Penyajian data secara umum dapat dibedakan menjadi dua bentuk utama, yaitu:

#### 3.2.1 Penyajian Data dalam Bentuk Tabel

Tabel adalah cara penyajian data yang menggunakan kolom dan baris untuk mengorganisir informasi. Tabel dapat dibedakan menjadi beberapa jenis:

##### Tabel Satu Arah

Tabel satu arah adalah tabel yang hanya menampilkan satu variabel atau kategori dalam satu dimensi. Tabel ini sering digunakan untuk menyajikan jumlah objek dalam berbagai kelompok. Tabel 3.1 adalah contoh jumlah mahasiswa di suatu universitas berdasarkan fakultas.

**Tabel 3.1.** Contoh Tabel Satu Arah

Fakultas	Jumlah Mahasiswa
Teknik	350
Ekonomi	420
Hukum	300
Kedokteran	200

Tabel satu arah memberikan gambaran sederhana tentang distribusi data, tetapi kurang efektif untuk membandingkan hubungan antar variabel.

##### Tabel Dua Arah

Tabel dua arah digunakan untuk menyajikan data yang melibatkan dua variabel atau kategori. Tabel ini sering digunakan dalam survei atau penelitian untuk menunjukkan hubungan antara dua faktor. Cohntohnuya adalah:

**Tabel 3.2.** Contoh Tabel Dua Arah

Fakultas	Laki-laki	Perempuan	Total
Teknik	200	150	350
Ekonomi	180	240	420
Hukum	130	170	300
Kedokteran	90	110	200

Tabel dua arah lebih informatif dibandingkan tabel satu arah karena memungkinkan analisis lebih dalam terhadap hubungan antar variabel.

### **Tabel Distribusi Frekuensi**

Tabel distribusi frekuensi digunakan untuk menyajikan data dalam kelompok-kelompok atau interval tertentu. Tabel ini berguna untuk data kuantitatif yang memiliki banyak variasi. Contohnya adalah:

**Tabel 3.3.** Contoh Table Distribusi Frekuensi

Rentang Nilai	Frekuensi
50 - 60	5
61 - 70	8
71 - 80	12
81 - 90	10
91 - 100	5

Tabel distribusi frekuensi mempermudah analisis tren data dan pola distribusi. Dalam era modern, metode otomatisasi seperti penggunaan *software* statistik (*SPSS, R, atau Python*) semakin banyak digunakan untuk membuat tabel ini secara efisien.

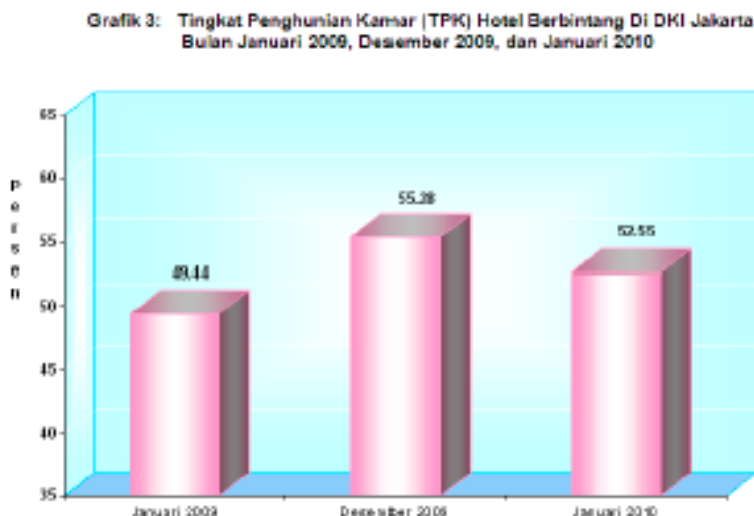
Isu terbaru dalam penyajian tabel adalah meningkatnya kebutuhan akan data *real-time* dan interaktif. Banyak organisasi kini menggunakan *dashboard* digital yang memungkinkan tabel dinamis dengan filter dan analisis langsung, yang sangat berguna dalam bisnis dan penelitian ilmiah.

### 3.2.2 Penyajian Data dalam Bentuk Grafik

Grafik adalah representasi visual dari data yang memudahkan interpretasi. Jenis-jenis grafik yang sering digunakan dalam statistika meliputi:

#### Grafik Batang

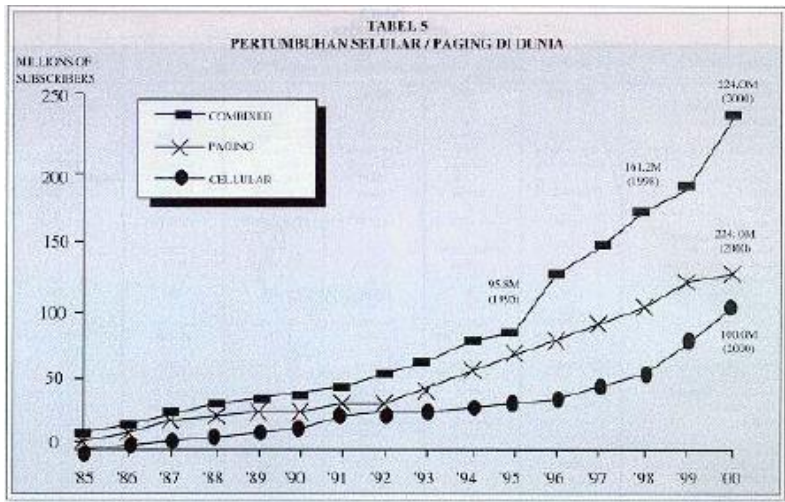
Grafik batang digunakan untuk membandingkan kategori data. Grafik ini menampilkan data dalam bentuk batang vertikal atau horizontal yang panjangnya sebanding dengan frekuensi atau nilai kategori yang diwakili. Grafik batang sering digunakan dalam survei kepuasan pelanggan atau perbandingan penjualan antar produk. Contohnya adalah:



Gambar 3.1. Contoh Grafik Batang

#### Grafik Garis

Grafik garis digunakan untuk menunjukkan perubahan atau tren data dari waktu ke waktu. Grafik ini biasanya digunakan dalam bidang ekonomi, seperti memantau pertumbuhan inflasi atau harga saham. Isu terbaru dalam grafik garis adalah penggunaannya dalam analisis big data untuk memprediksi tren masa depan. Contohnya adalah:

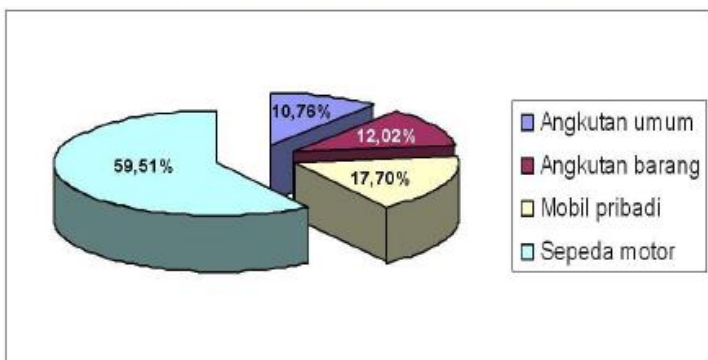


**Gambar 3.2. Contoh Grafik Garis**

### Grafik Lingkaran (*Pie Chart*)

Grafik lingkaran menampilkan data dalam bentuk lingkaran yang terbagi menjadi beberapa sektor, dengan setiap sektor mewakili persentase dari keseluruhan data. Grafik ini sangat populer dalam analisis demografi atau distribusi anggaran. Tantangan utama dalam penggunaan grafik lingkaran adalah ketika terdapat terlalu banyak kategori, yang dapat membuatnya sulit dibaca. Contohnya adalah:

#### Kendaraan Terlibat Kecelakaan (Januari-Juni 2010)



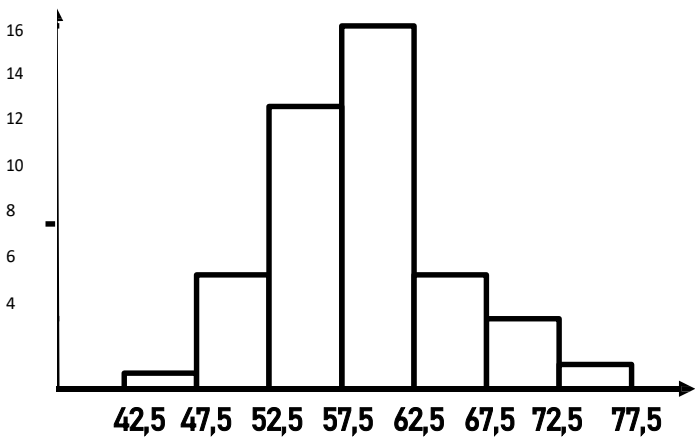
Sumber: Ditlantas Polda Metro Jaya

**Gambar 3.3. Contoh Grafik Lingkaran**

**Histogram**

Histogram menyerupai grafik batang, tetapi digunakan untuk data numerik yang dikelompokkan dalam kelas-kelas interval. Histogram sangat berguna dalam analisis distribusi data, seperti pola nilai ujian mahasiswa atau variasi ukuran produk industri. Contohnya:

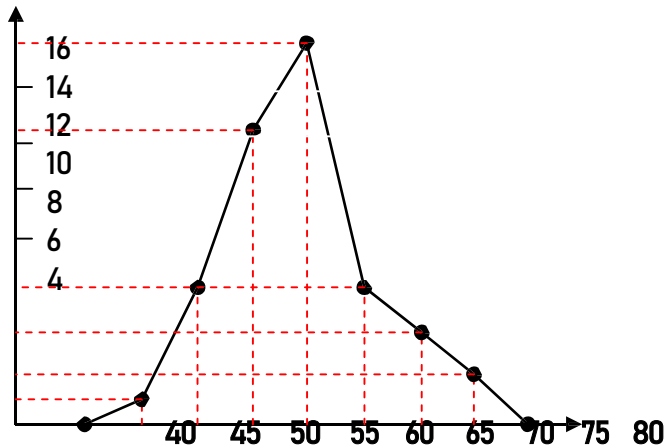
Hasil Pengukuran	Frekuensi
43 – 47	1
48 – 52	6
53 – 62	13
58 – 62	16
63 – 67	6
68 – 72	4
73 – 77	2



Gambar 3.4. Contoh Histogram

**Polygon Frekuensi**

Polygon frekuensi adalah versi garis dari histogram yang menghubungkan titik tengah setiap kelas interval. Grafik ini memudahkan analisis tren distribusi data dan sering digunakan dalam penelitian akademik dan pengendalian kualitas industri. Contohnya:



Gambar 3.5. Contoh Polygon Frekuensi

### 3.3 Penyajian Data dalam Kehidupan Sehari-hari Bidang Kesehatan

Dalam bidang kesehatan, penyajian data sangat penting untuk memantau tren penyakit, efektivitas vaksin, dan penyebaran wabah. Misalnya, grafik epidemiologi digunakan untuk melacak jumlah kasus COVID-19 harian di berbagai wilayah. Penyajian data yang akurat dapat membantu otoritas kesehatan dalam mengambil keputusan cepat dan efektif.

#### Bidang Ekonomi

Di bidang ekonomi, penyajian data digunakan dalam analisis inflasi, pertumbuhan ekonomi, dan pengangguran. Statistik yang disajikan dalam bentuk grafik garis atau tabel distribusi dapat membantu pemerintah dan pelaku bisnis dalam merumuskan kebijakan ekonomi yang lebih baik. Isu terbaru dalam penyajian data ekonomi adalah penggunaan *Big Data* dan *Machine Learning* untuk membuat prediksi keuangan yang lebih akurat.



## **Bidang Pendidikan**

Dalam dunia pendidikan, penyajian data digunakan untuk menganalisis kinerja siswa, efektivitas metode pembelajaran, dan distribusi anggaran pendidikan. Contohnya, tabel distribusi frekuensi nilai ujian dapat membantu dosen dalam mengidentifikasi pola keberhasilan akademik mahasiswa. Teknologi analisis data kini semakin banyak diterapkan dalam sistem pendidikan untuk memberikan wawasan yang lebih mendalam.

## **3.4 Kesimpulan**

Penyajian data merupakan langkah penting dalam analisis statistika yang bertujuan untuk menyederhanakan dan menginterpretasikan informasi dengan lebih baik. Melalui berbagai metode seperti tabel, grafik, dan diagram, data dapat disajikan secara lebih jelas dan mudah dipahami oleh pengambil keputusan.

Dalam dunia modern, tantangan utama dalam penyajian data adalah memastikan keakuratan dan keterbacaan informasi. Dengan berkembangnya teknologi, semakin banyak alat bantu yang dapat digunakan untuk membuat visualisasi data yang lebih interaktif dan informatif. Big data, kecerdasan buatan, serta machine learning kini mulai diterapkan dalam penyajian data guna meningkatkan keandalan dan efektivitas analisis.

Selain itu, penting untuk memperhatikan aspek etika dalam penyajian data. Manipulasi visualisasi data yang tidak sesuai dengan standar ilmiah dapat menyebabkan misinformasi dan pengambilan keputusan yang keliru. Oleh karena itu, pemahaman yang mendalam tentang metode penyajian data yang benar menjadi sangat penting bagi akademisi, praktisi bisnis, serta pemerintah dalam menyusun kebijakan yang berbasis data.

Dengan semakin berkembangnya dunia digital dan akses terhadap data, keterampilan dalam menyajikan data secara efektif akan menjadi kompetensi yang sangat berharga. Oleh karena itu, memahami berbagai teknik dan isu terkini dalam penyajian data merupakan langkah yang esensial dalam penerapan statistika di berbagai bidang kehidupan.

## DAFTAR PUSTAKA

- Montgomery, D. C., & Runger, G. C. (2020). *Applied Statistics and Probability for Engineers* (7th ed.). Wiley.
- Levine, D. M., Stephan, D. F., Krehbiel, T. C., & Berenson, M. L. (2020). *Statistics for Managers Using Microsoft Excel* (9th ed.). Pearson.
- McClave, J. T., Benson, P. G., & Sincich, T. (2018). *Statistics for Business and Economics* (13th ed.). Pearson.
- Moore, D. S., McCabe, G. P., & Craig, B. A. (2021). *Introduction to the Practice of Statistics* (10th ed.). W. H. Freeman.
- McKinney, W. (2017). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython* (2nd ed.). O'Reilly Media.
- Triola, M. F. (2021). *Elementary Statistics* (14th ed.). Pearson.
- Badan Pusat Statistik. (2022). *Laporan statistik tahunan*. Jakarta: BPS.
- Wickham, H., & Golemund, G. (2017). *R for data science: Import, tidy, transform, visualize, and model data*. O'Reilly Media.
- World Health Organization. (2021). *COVID-19 weekly epidemiological update*. Retrieved from <https://www.who.int>



# BAB 4

## UKURAN PEMUSATAN DATA

Oleh Susana Labuem

### 4.1 Konsep Dasar Ukuran Pemusatan

Definisi:

Ukuran pemusatan adalah nilai pada pusat atau tengah dari kumpulan data.

### 4.2 Data Tunggal

#### 1. Mean

Definisi:

Rata-rata aritmatika atau mean sekumpulan data adalah ukuran pemusatan yang ditemukan dengan menjumlahkan nilai-nilai data dan membagi totalnya dengan banyaknya nilai data.

Definisi ini dapat diekspresikan seperti pada formula 3-1 dimana huruf Yunani  $\Sigma$  mengindikasikan bahwa nilai-nilai data harus dijumlahkan. Yaitu,  $\Sigma x$  merepresentasikan jumlah dari semua nilai data. Simbol  $n$  menotasikan **ukuran sampel**, yaitu banyaknya nilai data.

#### Formula 3-1

$$\text{mean} = \frac{\Sigma x}{n} \quad \begin{array}{l} \leftarrow \text{jumlah semua nilai data} \\ \text{banyaknya nilai data} \end{array}$$

Jika data adalah sampel dari sebuah populasi, mean dinotasikan dengan  $\bar{x}$  (dibaca "x bar"); jika data adalah keseluruhan populasi, mean dinotasikan dengan  $\mu$  (huruf kecil Yunani "mu").

### Notasi

- $\Sigma$  : melambangkan jumlah dari sekumpulan nilai data  
 $x$  : adalah variable yang biasanya digunakan untuk merepresentasikan nilai data tunggal  
 $n$  : merepresentasikan banyaknya nilai data dalam sampel  
 $N$  : merepresentasikan banyaknya nilai data dalam populasi

#### Contoh 1:

Temukan mean dari ukuran tinggi lima ekor ayam jantan berikut (ukuran dalam satuan cm) : 27,531; 15,684; 5, 638; 27,997; dan 25,433.

Solusi:

Mean dihitung dengan menggunakan Formula 3-1.

Pertama jumlahkan nilai data, kemudian bagi dengan banyaknya nilai data.

$$\begin{aligned}\bar{x} &= \frac{\sum x}{n} = \frac{27,531 + 15,684 + 5,638 + 27,997 + 25,433}{5} \\ &= \frac{102,283}{5} \\ &= 20,456,6\end{aligned}$$

## 2. Median

Tidak seperti mean, median merupakan ukuran pemusatan yang resistan, karena tidak berubah dalam nilai yang besar hanya dikarenakan oleh kehadiran beberapa nilai ekstrem. Median dapat diartikan secara lebih luas sebagai “nilai tengah” dalam arti bahwa sekitar setengah dari nilai dalam kumpulan data berada di bawah median dan setengahnya lagi berada di atasnya.

Definisi:

Median dari sekumpulan data adalah ukuran pemusatan yang merupakan nilai tengah ketika nilai-nilai data asli besarnya diurutkan naik (atau turun). Median seringkali dinotasikan dengan  $\tilde{x}$  (dibaca “x-tilde”).

Untuk menemukan median, pertama urutkan nilai-nilai data, kemudian ikuti satu dari dua prosedur berikut:

- Jika banyaknya nilai data ganjil, maka median adalah bilangan yang terletak tepat di tengah urutan.

- b. Jika banyaknya nilai data genap, maka median ditemukan dengan menghitung mean dari dua bilangan yang berada ditengah urutan.

Contoh 2:

Temukan nilai median dari sampel nilai data yang ada pada contoh 1:

27,531; 15,684; 5, 638; 27,997; dan 25,433.

Solusi:

Pertama-tama nilai datanya diurutkan:

5, 638   15,684   25,433   27,531   27,997

Karena banyaknya nilai data adalah bilangan ganjil (5), maka median adalah bilangan yang terletak tepat di tengah-tengah daftar terurut , yaitu 25,433.

Contoh 3:

Ulangi contoh 2 setelah memasukan nilai data tambahan 8,077.

Temukan nilai median dari data tersebut.

Solusi:

5, 638   8,077   15,684   25,433   27,531   27,997

Karena banyaknya nilai data adalah bilangan genap (6), maka median dicari dengan menghitung rata-rata kedua bilangan yang di tengah, yaitu 15,684 dan 25,433.

$$\text{Median} = \frac{15,684 + 25,433}{2} = \frac{41,117}{2} = 20,558.5$$

### 3. Modus

Modus juga merupakan salah satu ukuran pemusatan.

Definisi

Modus dari suatu data adalah nilai yang muncul dengan frekuensi terbesar.

Sekumpulan data dapat memiliki satu modus, lebih dari satu modus, atau tidak memiliki modus.

- Ketika ada dua nilai yang muncul dengan frekuensi terbesar yang sama, masing-masing diantaranya merupakan modus dan data tersebut disebut dengan **bimodal**.
- Ketika lebih dari dua nilai yang muncul dengan frekuensi terbesar yang sama, masing-masing diantaranya merupakan modus dan data tersebut disebut dengan **multimodal**.
- Ketika tidak ada nilai data yang berulang, maka data tersebut dikatakan **tidak memiliki modus**.

Contoh 4:

Tentukan modus dari data berikut.

18360      18360      27531      15684      5638      27997      25433

Solusi:

Modusnya adalah 18360 karena merupakan nilai dengan frekuensi terbesar.

Dalam Contoh 4, data hanya memiliki satu modus. Berikut adalah dua kemungkinan keadaan lain:

Dua modus: 0, 0, 0, 1, 1, 2, 3, 5, 5, 5 memiliki dua modus: 0 dan 5.

Tidak ada modus: 0, 1, 2, 3, 5 tidak memiliki mode karena tidak ada nilai yang muncul lebih dari sekali.

Pada kenyataannya, modus ini tidak banyak digunakan dalam data numerik. Namun, modus adalah satu-satunya ukuran pemusatan yang dapat digunakan pada data nominal. (Ingat, data nominal terdiri dari: nama, label, atau kategori saja.).

#### 4. Midrange

Ukuran pemusatan lainnya adalah *midrange*. Karena *midrange* hanya menggunakan nilai maksimum dan minimum, terlalu sensitif terhadap ekstrim, sehingga *midrange* jarang digunakan. Namun, *midrange* memiliki tiga kelebihan:

- Sangat mudah untuk dihitung;
- Membantu memperkuat poin penting bahwa ada beberapa cara berbeda untuk mendefinisikan pusat data;
- Kadang-kadang salah digunakan untuk median, sehingga kebingungan dapat dikurangi dengan mendefinisikan dengan jelas *midrange* bersama dengan median.

Definisi:

Midrange (nilai tengah) adalah ukuran pemusatan yang merupakan nilai tengah antara nilai maksimum dan minimum dalam data. Ditemukan dengan menambahkan nilai data maksimum ke nilai data minimum dan kemudian dibagi 2, seperti dalam rumus berikut:

$$\text{midrange} = \frac{\text{nilai data maksimum} + \text{nilai data minimum}}{2}$$

Contoh 5:

Tentukan midrange dari data berikut:

27531; 15684; 5638; 27997; 25433

Solusi:

$$\begin{aligned}\text{midrange} &= \frac{\text{nilai data maksimum} + \text{nilai data minimum}}{2} \\ &= \frac{27997 + 5638}{2} \\ &= 16871,5\end{aligned}$$

Istilah rata-rata sering digunakan untuk mean, tetapi kadang-kadang digunakan untuk ukuran pemusatan lainnya. Untuk menghindari kebingungan atau ambiguitas, kita menggunakan istilah yang benar dan spesifik, seperti mean atau median. Istilah rata-rata tidak digunakan oleh ahli statistik dan tidak akan digunakan pada pembahasan lanjutan di buku ini. Ketika menghitung ukuran pemusatan, kita sering perlu membulatkan hasilnya. Kita menggunakan aturan berikut.

#### **Aturan Pembulatan untuk Mean, Median, dan Midrange**

Membulatkan pada satu tempat desimal lebih dari yang digunakan untuk nilai aslinya.

(Karena nilai modulus sama dengan beberapa nilai data asli, nilai tersebut dapat dibiarkan apa adanya tanpa pembulatan apa pun.

Saat menerapkan aturan ini, bulatkan hanya jawaban akhir, bukan nilai tengah yang muncul selama perhitungan. Misalnya, rata-rata dari 2, 3, 5, adalah 3,333333..., yang dibulatkan menjadi 3,3, yang memiliki satu tempat desimal lebih dari nilai asli 2, 3, 5. Sebagai contoh lain, rata-rata dari 80,4 dan 80,6 adalah 80,50 (satu tempat desimal lebih dari yang digunakan untuk nilai aslinya). Karena modulus adalah satu atau lebih dari nilai data asli,



kami tidak membulatkan nilai modus; kami hanya menggunakan nilai asli yang sama.

### 5. Rata-Rata Dari Suatu Distribusi Frekuensi

Saat bekerja dengan data yang dirangkum dalam distribusi frekuensi, kita tidak mengetahui secara pasti nilai yang termuat dalam suatu kelas tertentu. Untuk memungkinkan perhitungan, kita asumsikan bahwa semua nilai sampel di setiap kelas sama dengan titik tengah kelas. Sebagai contoh, pertimbangkan interval kelas 0–9,999 dengan frekuensi 46 seperti pada tabel 3.1 berikut ini.

**Tabel 4.1.** Menemukan Rata-Rata dari Suatu Distribusi Frekuensi

Word Counts from Men	Frequency <i>f</i>	Class Midpoint <i>x</i>	<i>f</i> · <i>x</i>
0–9,999	46	4,999.5	229,977.0
10,000–19,999	90	14,999.5	1,349,955.0
20,000–29,999	40	24,999.5	999,980.0
30,000–39,999	7	34,999.5	244,996.5
40,000–49,999	3	44,999.5	134,998.5
Totals:	Σ <i>f</i> = 186		Σ( <i>f</i> · <i>x</i> ) = 2,959,907
			$\bar{x} = \frac{\Sigma (f \cdot x)}{\Sigma f} = \frac{2,959,907}{186} = 15,913.5$

Nilai pada kelas 0 – 9,999 (kelas pertama) memiliki frekuensi 46, sehingga dengan menggunakan pendekatan titik tengah kelas kita mendapatkan nilai 4.999,5 sebanyak 46 pada kelas pertama. Oleh sebab itu, pada kelas pertama kita memiliki total nilai 4.999,5 × 46 = 229.977. Kemudian kita dapat menemukan total semua nilai sampel, yang kemudian untuk menemukan rata-ratanya, dibagi dengan jumlah frekuensi (Σ*f*). Formula 4-2 digunakan untuk menghitung rata-rata ketika data sampel diringkaskan dalam distribusi frekuensi. Formula 4-2 ini bukan konsep baru, namun variasi dari Formula 4-1.

### Formula 3-2

$$\text{rata - rata distribusi frekuensi } \bar{x} = \frac{\sum(f.x)}{\sum f}$$

Contoh 6:

Menghitung mean dari distribusi frekuensi.

Dua kolom pertama dalam Tabel 4.1 merupakan distribusi frekuensi yang meringkas jumlah kata dari 186 orang dalam 8 kumpulan data dari lampiran B. Gunakan distribusi frekuensi untuk mencari mean.

Solusi:

Tabel 4.1 mengilustrasikan penggunaan formula 4-2 saat menghitung mean dari data yang diringkas dalam distribusi frekuensi. Nilai titik tengah kelas disajikan di kolom ketiga, dan hasil kali frekuensi  $f.x$  disajikan di kolom terakhir. Perhitungan menggunakan formula 4-2 ditunjukkan di bagian bawah tabel 4.1. Hasilnya adalah  $\bar{x} = 15,913.5$ . Jika kita menggunakan cara yang manual, kita akan mendapatkan hasil  $\bar{x} = 15,668.5$ . Distribusi frekuensi menghasilkan perkiraan  $\bar{x}$  karena tidak didasarkan pada daftar nilai sampel asli yang tepat.

### 6. Weighted Mean (Rata-Rata Tertimbang)

Ketika nilai data diberi bobot yang berbeda, kita dapat menghitung **rata-rata tertimbang**. Formula 3-3 dapat digunakan untuk menghitung rata-rata tertimbang,  $w$ .

### Formula 3-3

$$\text{weighted mean} = \bar{x} = \frac{\sum(w \cdot x)}{\sum w}$$

Formula 4-3 memberitahu kita terlebih dahulu untuk mengalikan setiap bobot  $w$  dengan nilai  $x$  yang sesuai, kemudian menjumlahkan produknya, dan akhirnya membagi total itu dengan jumlah bobot  $\sum w$ .

Contoh 7:

Menghitung nilai rata-rata.

Pada semester pertama kuliahnya, seorang mahasiswa penulis mengambil lima mata kuliah. Nilai akhir beserta jumlah sks untuk setiap mata kuliah adalah: A (3 sks); A (4 sks); B (3 sks), C (3 sks), dan F (1 sks). Sistem penilaian memberikan indeks kualitas ke nilai huruf sebagai berikut: A=4; B=3; C=2; D=1; F=0. Hitunglah rata-rata indeks prestasinya.

Solusi:

Gunakan banyak sks sebagai bobot  $w = 3, 4, 3, 3, 1$ . Ganti nilai huruf dari A, A, B, C, dan F dengan poin kualitas yang sesuai:  $x = 4, 4, 3, 2, 0$ . Sekarang kita menggunakan Rumus 4-3 seperti yang ditunjukkan di bawah ini. Hasilnya adalah nilai rata-rata semester pertama 3,07. (Menggunakan aturan pembulatan sebelumnya, hasilnya harus dibulatkan menjadi 3.1, tetapi adalah umum untuk membulatkan nilai rata-rata dengan dua tempat desimal).

$$\begin{aligned}\bar{x} &= \frac{\sum(w \cdot x)}{\sum w} \\ &= \frac{(3 \times 4) + (4 \times 4) + (3 \times 3) + (3 \times 2) + (1 \times 0)}{3 + 4 + 3 + 3 + 1} \\ &= \frac{43}{14} \\ &= 3,07\end{aligned}$$

## 7. SKEWED

Perbandingan rata-rata, median, dan modus dapat mengungkapkan informasi tentang karakteristik *skewness*, yang didefinisikan di bawah ini dan diilustrasikan pada Gambar 6.1.

Definisi:

Suatu distribusi data adalah *skewed* jika tidak simetris dan meluas lebih ke satu sisi daripada yang lain. (Suatu distribusi data simetris jika bagian kiri histogramnya kira-kira merupakan bayangan cermin dari bagian kanannya).

Data *skewed* ke kiri (juga disebut *skewed* negatif) memiliki ekor kiri yang lebih panjang, dan rata-rata dan median berada di sebelah kiri mode. Data yang condong ke kanan (juga disebut *skewed* positif) memiliki ekor kanan yang lebih panjang, dan rata-rata dan median berada di sebelah kanan modus.

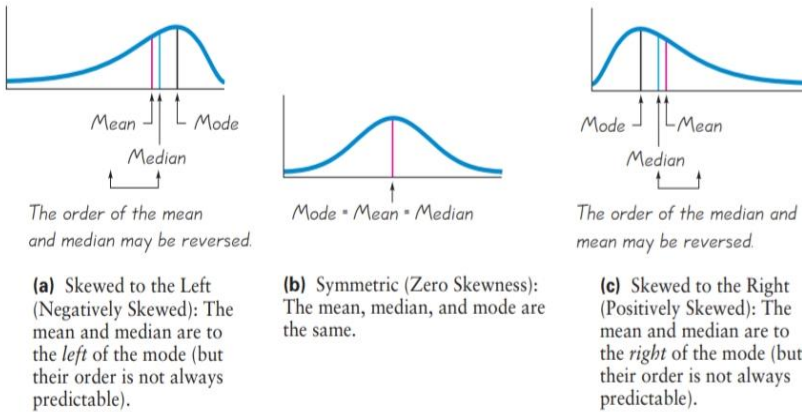
Data *skewed* biasanya (tetapi tidak selalu) memiliki rata-rata yang terletak lebih jauh di ekor yang lebih panjang daripada median. Gambar 6.1(a) menunjukkan rata-rata di sebelah kiri median untuk data yang condong ke kiri, dan Gambar 6.1(c) menunjukkan rata-rata di sebelah kanan median untuk data yang condong ke kanan, tetapi relatif posisi rata-rata dan median tidak selalu seperti yang ditunjukkan pada gambar.

Sebagai contoh dimungkinkan untuk memiliki data *skewed* ke kiri dengan median kurang dari rata-rata, bertentangan dengan urutan yang ditunjukkan pada Gambar 6.2(a).

Untuk nilai  $-100, 1.0, 1.5, 1.7, 1.8, 2.0, 3.0, 4.0, 5.0, 50.0, 50.0, 60.0$ , suatu histogram menunjukkan bahwa data *skewed* ke kiri, tetapi rata-rata 6,7 lebih besar dari median 2,5, bertentangan dengan urutan rata-rata dan median yang ditunjukkan pada Gambar 4.1(a).

Dalam praktiknya, banyak distribusi data yang mendekati simetris dan tanpa *skewness*. Distribusi yang *skewed* ke kanan lebih umum daripada yang *skewed* ke kiri karena seringkali lebih mudah untuk mendapatkan nilai yang sangat besar daripada nilai yang sangat kecil. Dengan pendapatan tahunan, misalnya, tidak mungkin mendapatkan nilai di bawah nol, tetapi ada beberapa orang yang menghasilkan jutaan atau miliaran

dolar dalam setahun. Oleh karena itu, pendapatan tahunan cenderung condong ke kanan, seperti pada Gambar 4.1(c).



**Gambar 4.1. Skewness**

## 4.3 Data Kelompok

### 1. Rata-Rata atau Ukuran Tendensi Tengah

Rata-rata adalah nilai yang khas, atau mewakili, dari sekumpulan data. Karena nilai tipikal seperti itu cenderung terletak secara terpusat di dalam sekumpulan data yang disusun menurut besarnya, rata-rata juga disebut ukuran tendensi sentral. Beberapa jenis rata-rata dapat didefinisikan, yang paling umum adalah rata-rata aritmatika, median, modus, rata-rata. Masing-masing memiliki kelebihan dan kekurangan, tergantung pada data dan tujuan yang dimaksudkan.

#### a. Rata-Rata Aritmatika

Rata-rata aritmatika, atau singkatnya rata-rata, dari himpunan  $N$  bilangan  $X_1, X_2, X_3, \dots, X_N$  dilambangkan dengan  $\bar{X}$  (dibaca "X bar") dan didefinisikan sebagai

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \dots + X_N}{N} = \frac{\sum_{j=1}^N X_j}{N} = \frac{\sum X}{N}$$

Contoh 8:

Rata-rata aritmatika bilangan 8, 3, 5, 12, dan 10 adalah

$$\overline{X} = \frac{8+3+5+12+10}{5} = \frac{38}{5} = 7,6$$

Jika bilangan  $X_1, X_2, \dots, X_K$  muncul  $f_1, f_2, \dots, f_K$  kali, masing-masing (yaitu, terjadi dengan frekuensi  $f_1, f_2, \dots, f_K$ ), mean aritmatikanya adalah:

$$\overline{X} = \frac{f_1 X_1 + f_2 X_2 + \dots + f_k X_k}{f_1 + f_2 + \dots + f_k} = \frac{\sum_{j=1}^k f_j X_j}{\sum_{j=1}^k f_j} = \frac{\sum fX}{\sum f} = \frac{\sum fX}{\sum N}$$

Dimana  $N = \sum f$  adalah frekuensi total (yaitu, jumlah total kasus)

Contoh 9:

Jika 5, 8, 6, dan 2 berturut-turut muncul dengan frekuensi 3, 2, 4, dan 1, rata-rata aritmatikanya adalah

$$\overline{X} = \frac{(3)(5) + (2)(8) + (4)(6) + (1)(2)}{3 + 2 + 4 + 1} = \frac{15 + 16 + 24 + 2}{10} = 5,7$$

b. Sifat-sifat Aritmatika

Sifat-sifat aritmatika, yaitu:

- 1) Jumlah aljabar simpangan himpunan bilangan dari rata-rata aritmatikanya adalah nol.

Contoh 10:

Penyimpangan angka 8, 3, 5, 12, dan 10 dari mean aritmatika, perhatikan (contoh 1) adalah 8-7.6, 3-7.6, 5-7.6, 12-7.6, dan 10-7.6, atau 0.4, -4.6, -2.6, 4.4, dan 2.4, dengan jumlah aljabar 0.4-4.6-2.6+4.4+2.4=0

- 2) Jika bilangan  $f_1$  memiliki mean  $m_1$ , bilangan  $f_2$  memiliki mean  $m_2, \dots$ , bilangan  $f_K$  memiliki mean  $m_K$ , maka mean dari semua bilangan tersebut adalah

$$\bar{X} = \frac{f_1 m_1 + f_2 m_2 + \dots + f_k m_k}{f_1 + f_2 + \dots + f_k}$$

- 3) Jika  $A$  adalah sembarang rata-rata aritmatika yang diduga atau diasumsikan (yang dapat berupa bilangan apa saja) dan jika  $d_j = X_j - A$  adalah deviasi  $X_j$  dari  $A$ , maka persamaan (1) dan (2) masing-masing menjadi:

$$\bar{X} = A + \frac{\sum_{j=1}^N d_j}{N} = A + \frac{\sum d}{N}$$

$$\bar{X} = A + \frac{\sum_{j=1}^K f_j d_j}{\sum_{j=1}^K f_j} = A + \frac{\sum f d}{N}$$

c. Rata-Rata Aritmatika Yang Dihitung dari Data Yang Dikelompokkan

Ketika data disajikan dalam distribusi frekuensi, semua nilai yang berada dalam interval kelas tertentu dianggap bertepatan dengan tanda kelas, atau titik tengah, dari interval tersebut. Rumus (2) dan (5) valid untuk kelompok data tersebut jika kita menginterpretasikan  $X_j$  sebagai tanda kelas,  $f_j$  sebagai frekuensi kelas yang bersesuaian,  $A$  sebagai tanda kelas yang diduga atau diasumsikan, dan  $d_j = X_j - A$  sebagai simpangan dari  $X_j$  dari  $A$ .

Contoh 11:

Gunakan distribusi frekuensi tinggi badan pada Tabel 2.1 untuk mencari rata-rata tinggi badan dari 100 mahasiswa laki-laki di universitas XYZ.

**Table 2.1 Heights of 100 male students at XYZ University**

Height (in)	Number of Students
60–62	5
63–65	18
66–68	42
69–71	27
72–74	8
Total	100

Solusi:

Perhatikan bahwa semua siswa yang memiliki tinggi 60 hingga 62 inci (inci), 63 hingga 65 inci, dll., dianggap memiliki tinggi badan 61 inci, 64 inci, dst. Masalahnya kemudian dikurangi menjadi mencari tinggi rata-rata 100 siswa jika 5 siswa memiliki tinggi 61 inci, 18 memiliki tinggi 64 inci, dll. Perhitungan yang terlibat bisa menjadi membosankan, terutama untuk kasus di mana jumlahnya besar dan banyak kelas hadir.

**Table 3.1**

Height (in)	Class Mark ( $X$ )	Frequency ( $f$ )	$fX$
60–62	61	5	305
63–65	64	18	1152
66–68	67	42	2814
69–71	70	27	1890
72–74	73	8	584
$N = \sum f = 100$			$\sum fX = 6745$

$$\bar{X} = \frac{\sum fX}{\sum f} = \frac{\sum fX}{N} = \frac{6745}{100} = 67.45 \text{ in}$$

## 2. Median

Median dari sekumpulan angka yang diatur dalam urutan besarnya (yaitu, dalam array) adalah nilai tengah atau rata-rata aritmatika dari dua nilai tengah. Untuk data yang dikelompokkan, median, yang diperoleh dengan interpolasi, diberikan oleh:



$$Median = L_i + \left( \frac{\frac{N}{2} - (\sum f)_{\text{I}}}{f_{\text{median}}} \right) c$$

Dimana:

- $L_i$  : batas kelas bawah dari kelas median  
(yaitu, kelas yang berisi median)
- $N$  : jumlah item dalam data (yaitu, frekuensi total)
- $(\sum f)_{\text{I}}$  : jumlah frekuensi semua kelas lebih rendah dari kelas median
- $f_{\text{median}}$  : frekuensi kelas median
- $c$  : ukuran interval kelas median

### 3. Modus

Modus suatu himpunan bilangan adalah nilai yang muncul dengan frekuensi terbesar; yaitu, itu adalah nilai yang paling umum. Mode mungkin tidak ada, dan bahkan jika memang ada, mode itu mungkin tidak unik.

Contoh 12:

Himpunan 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12, dan 18 memiliki modus 9.

Contoh 13:

Himpunan 3, 5, c, 10, 12, 15, dan 16 tidak memiliki modus.

Contoh 14:

Himpunan 2, 3, 4, 4, 4, 5, 5, 7, 7, 7, dan 9 memiliki dua modus, 4 dan 7, dan disebut bimodal.

Distribusi yang hanya memiliki satu modus disebut unimodal. Dalam kasus data yang dikelompokkan di mana kurva frekuensi telah dibangun agar sesuai dengan data, modus akan menjadi nilai (atau nilai) dari  $X$  yang sesuai dengan titik (atau titik). Nilai  $X$  ini kadang-kadang dilambangkan dengan  $\hat{X}$ . Dari distribusi frekuensi atau histogram modus dapat diperoleh dari rumus:

$$Modus = L_i + \left( \frac{\Delta_1}{\Delta_1 + \Delta_2} \right) c$$

dimana:

$L_i$  : batas kelas bawah dari kelas modus (yaitu, kelas yang berisi modus)

$\Delta_1$  : selisih frekuensi kelas modus dengan kelas sebelumnya

$\Delta_1$  : selisih frekuensi kelas modus dengan kelas sesudahnya

$c$  : ukuran interval kelas modus

## DAFTAR PUSTAKA

- Triola, M. F. 2010. *Elementary Statistic*. Edisi kesebelas. USA: Pearson Education. Inc.
- Spiegel, M.R dan Stephen, L. J. 2008. *Theory and Problem of Statistics*. Edisi keempat. McGRAW-HILL

# **BAB 5**

## **UKURAN PENYEBARAN DATA**

**Oleh Ahmad Ridfah**

### **5.1 Pendahuluan**

Ukuran penyebaran data dikenal dengan istilah dispersi, selain itu akrab juga dengan istilah variabilitas. Variabilitas adalah salah satu konsep utama dalam statistika yang menunjukkan sejauh mana data tersebar atau berbeda satu sama lain. Variabilitas sangat penting karena membantu kita memahami seberapa konsisten atau beragam suatu kumpulan data. Tanpa memahami variabilitas, kita hanya akan melihat nilai tengah atau pemusatan data tanpa mengetahui sejauh mana nilai-nilai lain dalam kumpulan data berbeda dari nilai tengah tersebut. Variabilitas hadir dalam hampir semua aspek kehidupan.

Dalam konteks statistika, pemahaman terhadap variabilitas sangat penting karena membantu dalam pengambilan keputusan berdasarkan data (Garfield & Ben-Zvi, 2004). Selain itu, Kelley (2021) menyoroti bahwa ukuran penyebaran data memiliki peran krusial dalam memahami distribusi dan keandalan data yang dikumpulkan. Menurut Vetter (2017), statistik deskriptif berfungsi untuk menyajikan data dengan cara yang logis dan efisien. Ukuran penyebaran seperti rentang, standar deviasi, dan rentang antar kuartil adalah alat penting dalam pelaporan data karena memberikan pemahaman tentang distribusi data yang lebih luas. Dalam penelitian ilmiah, statistik deskriptif sering kali digunakan sebagai langkah awal sebelum melakukan analisis inferensial lebih lanjut.

### **5.2 Jenis-jenis Ukuran Penyebaran**

Ukuran penyebaran adalah cara untuk mengukur seberapa besar variasi dalam suatu kumpulan data. Berikut adalah beberapa ukuran penyebaran yang umum digunakan dalam statistika:

### 5.2.1 Rentang (*Range*)

Rentang adalah ukuran penyebaran yang paling sederhana. Rentang dihitung dengan cara mengurangi nilai tertinggi dalam data dengan nilai terendah. Rumusnya adalah:

$$\text{Rentang} = \text{Nilai Maksimum} - \text{Nilai Minimum}$$

Rentang memberikan gambaran kasar tentang sebaran data, tetapi tidak menunjukkan bagaimana data tersebar di antara nilai maksimum dan minimum tersebut (Garfield & Ben-Zvi, 2004). Kelley (2012) mencatat bahwa rentang memiliki keterbatasan dalam penggunaannya karena sangat dipengaruhi oleh ukuran sampel dan nilai ekstrem. Bhardwaj & Sharma (2013) menyoroti bahwa rentang sangat berguna dalam pengendalian kualitas karena dapat memberikan informasi cepat tentang penyebaran data. Namun, rentang kurang dapat diandalkan karena hanya mempertimbangkan dua nilai ekstrem dan mengabaikan distribusi nilai di antaranya.

Keunggulan utama rentang adalah kemudahannya dalam perhitungan, tetapi memiliki keterbatasan karena sangat dipengaruhi oleh nilai pencilon dan tidak mencerminkan variasi di antara nilai data lainnya. Oleh karena itu, dalam analisis statistik yang lebih kompleks, rentang sering tidak cukup untuk menggambarkan variabilitas data secara menyeluruh (Manikandan, 2011).

### 5.2.2 Rentang Antar Kuartil (*Interquartile Range - IQR*)

IQR digunakan untuk mengukur penyebaran data di sekitar median, dengan mengabaikan nilai ekstrem. Rentang antar kuartil dihitung dengan rumus:

$$\text{IQR} = \text{Kuartil Ketiga (Q3)} - \text{Kuartil Pertama (Q1)}$$

IQR lebih stabil dibanding rentang karena tidak dipengaruhi oleh nilai ekstrem dalam kumpulan data (Garfield & Ben-Zvi, 2004). Kelley (2012) menyarankan bahwa IQR adalah ukuran yang lebih andal dibandingkan dengan rentang karena menghilangkan pengaruh nilai pencilon. Menurut Vetter (2017), IQR sering digunakan dalam penelitian medis dan epidemiologi karena memberikan pemahaman lebih baik

tentang distribusi data, terutama ketika data tidak berdistribusi normal.

Whaley (2005) menunjukkan bahwa IQR memiliki karakteristik berbeda tergantung pada jenis distribusi:

1. Distribusi Normal: IQR dapat dengan mudah dihitung dengan menggunakan tabel distribusi normal standar.
2. Distribusi Eksponensial: IQR lebih tinggi dibandingkan distribusi normal karena sifat distribusinya yang skewed.
3. Distribusi Gamma dan Weibull: Dalam distribusi ini, IQR dapat bervariasi secara signifikan bergantung pada parameter distribusinya.

Whaley (2005) menunjukkan bahwa IQR memiliki beberapa keunggulan dibandingkan ukuran penyebaran lainnya, seperti standar deviasi atau rentang:

1. Lebih tahan terhadap pencilan: Karena hanya mempertimbangkan nilai antara kuartil pertama dan ketiga, IQR tidak dipengaruhi oleh nilai pencilan ekstrem seperti standar deviasi.
2. Merepresentasikan penyebaran data lebih akurat pada distribusi skewed: IQR memberikan informasi yang lebih tepat dibandingkan standar deviasi ketika data tidak terdistribusi normal.
3. Sederhana dan mudah dihitung: IQR hanya membutuhkan nilai kuartil dan dapat dihitung secara manual dengan mudah.

### **5.2.3 Deviasi Rata-rata (*Mean Deviation*)**

Deviasi rata-rata adalah ukuran penyebaran yang dihitung dengan mengambil rata-rata dari nilai absolut deviasi tiap data terhadap nilai pusatnya (mean, median, atau modus). Menurut Bhardwaj & Sharma (2013), deviasi rata-rata lebih baik daripada rentang karena mempertimbangkan semua nilai dalam distribusi, bukan hanya dua nilai ekstrem. Namun, kelemahan utama deviasi rata-rata adalah tidak dapat digunakan untuk analisis matematis lebih lanjut seperti standar deviasi.

### **5.2.4 Varians**

Varians mengukur seberapa jauh data menyebar dari rata-rata dengan menghitung selisih kuadrat antara setiap data dengan

rata-rata. Varians membantu kita memahami sejauh mana nilai dalam suatu dataset berbeda dari rata-rata (Garfield & Ben-Zvi, 2004). Menurut Kelley (2012), varians memiliki keunggulan dibandingkan ukuran penyebaran lainnya karena lebih stabil terhadap perubahan ukuran sampel.

Standar deviasi adalah ukuran variabilitas yang paling umum digunakan karena lebih mudah diinterpretasikan dibandingkan varians. Varians dihitung dengan cara mengambil rata-rata dari selisih kuadrat antara setiap nilai dalam kumpulan data dengan mean. Namun, karena hasil varians memiliki satuan kuadrat dari satuan asli data (misalnya, dalam satuan meter menjadi meter kuadrat), sering kali lebih praktis menggunakan standar deviasi yang merupakan akar kuadrat dari varians.

Joseph dkk. (2005) menekankan bahwa standar deviasi memberikan gambaran lebih intuitif tentang seberapa jauh nilai-nilai dalam kumpulan data tersebar dari mean. Sebagai contoh, dalam penelitian psikologi, standar deviasi skor kesejahteraan seseorang setelah terapi dapat menunjukkan apakah efek terapi relatif seragam di antara semua peserta atau apakah ada variasi yang signifikan.

### **5.2.5 Simpangan Baku (Standar Deviasi)**

Simpangan baku adalah akar kuadrat dari varians. Menurut Bhardwaj & Sharma (2013), standar deviasi lebih unggul dibandingkan rentang dan deviasi rata-rata karena dapat digunakan dalam analisis statistik yang lebih kompleks. Standar deviasi juga lebih tahan terhadap fluktuasi sampel dibandingkan ukuran penyebaran lainnya.

Menurut Manikandan (2011), standar deviasi memiliki beberapa keunggulan, terutama dalam distribusi normal:

1. 68% dari data berada dalam  $\pm 1$  SD dari mean.
2. 95% dari data berada dalam  $\pm 2$  SD dari mean.
3. 99,7% dari data berada dalam  $\pm 3$  SD dari mean.

Distribusi normal memungkinkan penggunaan standar deviasi untuk menentukan probabilitas suatu nilai jatuh dalam rentang tertentu. Dalam analisis statistik inferensial, informasi ini sangat berguna untuk pengambilan keputusan berbasis data.

Pemilihan ukuran penyebaran yang tepat bergantung pada jenis data:

1. Jika data berdistribusi normal, standar deviasi adalah pilihan terbaik.
2. Jika data memiliki pencilan atau distribusi *skewed*, rentang interkuartil lebih disarankan.
3. Jika hanya ingin mendapatkan gambaran kasar sebaran data, rentang dapat digunakan.

Standar deviasi tidak cocok digunakan jika data sangat *skewed* karena bisa memberikan gambaran yang menyesatkan mengenai variabilitas data (Manikandan, 2011).

### 5.3 Keunggulan *Mean Deviation* dibandingkan Standar Deviasi

Gorard (2005) dalam artikelnya *Revisiting a 90-Year-Old Debate: The Advantages of the Mean Deviation* mengemukakan bahwa *mean deviation* memiliki banyak keunggulan dibandingkan standar deviasi. Meskipun standar deviasi telah menjadi ukuran penyebaran yang dominan dalam statistik, *mean deviation* menawarkan sejumlah manfaat, termasuk kemudahan interpretasi, efisiensi dalam menghadapi data dengan kesalahan kecil, serta stabilitas dalam distribusi non-normal.

#### 5.3.1 *Mean Deviation* sebagai Alternatif Standar Deviasi

*Mean deviation* adalah rata-rata absolut dari selisih setiap nilai terhadap rata-rata dataset. Keunggulan utama mean deviation menurut Gorard (2005) adalah:

1. Lebih mudah dipahami dan dihitung – Tidak perlu melakukan operasi kuadrat dan akar kuadrat seperti dalam standar deviasi.
2. Lebih tahan terhadap kesalahan kecil dalam pengukuran – Dalam data kehidupan nyata yang sering mengandung error kecil, mean deviation lebih efisien daripada standar deviasi.
3. Lebih stabil untuk distribusi non-normal – Ketika data memiliki pencilan atau distribusi *skewed*, mean deviation memberikan gambaran penyebaran yang lebih representatif dibandingkan standar deviasi.



### 5.3.2 Mengapa Standar Deviasi Mendominasi?

Meskipun mean deviation memiliki banyak keunggulan, standar deviasi tetap menjadi ukuran utama dalam statistik. Gorard (2005) menguraikan beberapa alasan mengapa standar deviasi lebih banyak digunakan:

1. Alasan historis – Standar deviasi telah menjadi standar dalam statistik sejak abad ke-20 berkat teori-teori yang dikembangkan oleh Fisher (1920).
2. Kemudahan dalam manipulasi matematis – Standar deviasi lebih mudah digunakan dalam perhitungan lanjutan, seperti analisis varians (*ANOVA*), regresi linier, dan distribusi probabilitas normal.
3. Adopsi luas dalam metode statistik – Banyak uji statistik bergantung pada standar deviasi, termasuk perhitungan efektivitas ukuran seperti Cohen's *d* dan analisis variabilitas dalam eksperimen.

Namun, Gorard (2005) berpendapat bahwa dominasi standar deviasi lebih karena alasan kebiasaan daripada karena keunggulan mutlak. Dalam situasi tertentu, terutama dalam penelitian sosial dan pendidikan, mean deviation bisa menjadi pilihan yang lebih baik.

### 5.3.3 Kritik terhadap Standar Deviasi

Beberapa kritik utama terhadap standar deviasi yang diajukan oleh Gorard (2005) adalah:

1. Distorsi akibat efek *squaring* – Karena standar deviasi menggunakan kuadrat selisih dari rata-rata, nilai pencilan memiliki dampak yang sangat besar terhadap hasil akhir.
2. Kurang intuitif dibandingkan *mean deviation* – Standar deviasi sulit untuk dijelaskan kepada peneliti pemula karena melibatkan operasi kuadrat dan akar kuadrat.
3. Kurang efisien dalam data dengan kesalahan kecil – Gorard menunjukkan bahwa dalam data yang mengandung kesalahan kecil, *mean deviation* memberikan estimasi yang lebih baik terhadap variabilitas sebenarnya.

## 5.4 Kesalahan Umum dalam Interpretasi Ukuran Penyebaran

Beberapa kesalahan umum dalam interpretasi ukuran penyebaran yang diidentifikasi oleh Vetter (2017) meliputi:

1. Menggunakan *standar error* sebagai ukuran variabilitas – SEM bukanlah ukuran variabilitas data, tetapi lebih kepada ketepatan estimasi rata-rata sampel.
2. Menyimpulkan distribusi data hanya berdasarkan nilai rata-rata – Nilai rata-rata saja tidak cukup untuk memahami distribusi data secara keseluruhan tanpa mempertimbangkan variabilitas.
3. Mengabaikan pengaruh pencilan – Data ekstrem dapat mempengaruhi nilai ukuran penyebaran, sehingga perlu dipertimbangkan secara hati-hati dalam analisis.

## 5.5 Reasoning tentang Ukuran Penyebaran Data

Ben-Zvi dan Garfield (2004) menyatakan bahwa pemahaman variabilitas merupakan inti dari pemikiran statistik. Variabilitas tidak hanya sekadar ukuran penyebaran dalam data, tetapi juga aspek penting dalam membuat kesimpulan statistik.

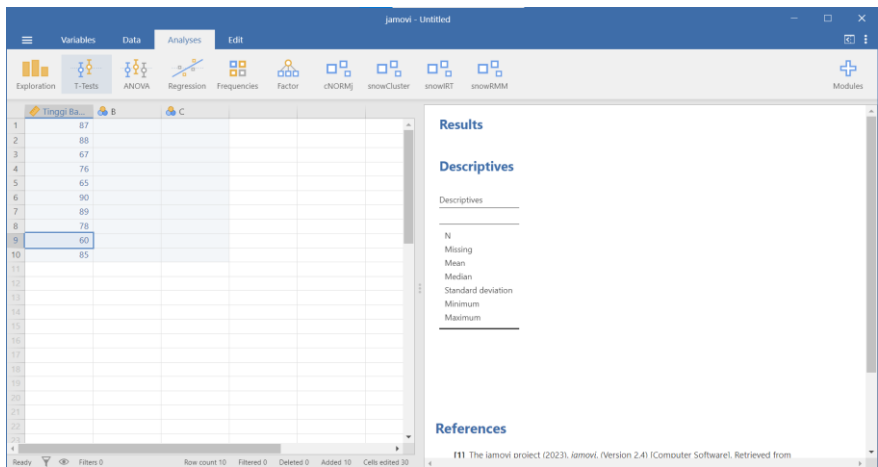
Ben-Zvi dan Garfield (2004) mengidentifikasi berbagai tantangan yang dihadapi siswa dalam memahami konsep variabilitas:

1. Mengabaikan Variabilitas – Banyak siswa cenderung hanya melihat nilai rata-rata tanpa mempertimbangkan variasi data.
2. Kesulitan dalam Membandingkan Variabilitas Antar Kelompok – Ketika membandingkan dua kumpulan data, siswa sering hanya melihat perbedaan rata-rata tanpa menilai seberapa besar variabilitas dalam setiap kelompok.
3. Kesalahan dalam Menginterpretasikan Standar Deviasi – Beberapa siswa menganggap standar deviasi sebagai ukuran kesalahan dalam data, bukan sebagai ukuran penyebaran data.

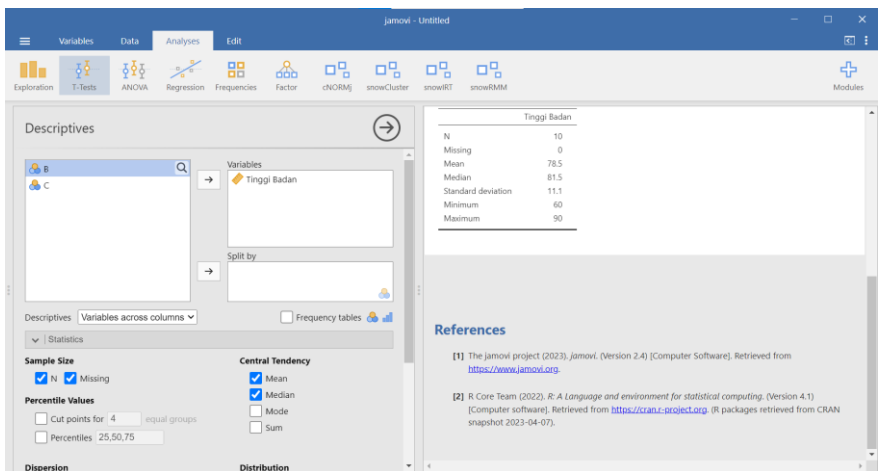
## 5.6 Jamovi untuk Ukuran Penyebaran Data

Jamovi merupakan salah satu perangkat lunak gratis yang dapat digunakan untuk mencari nilai-nilai ukuran penyebaran data. Pada menu Jamovi ukuran penyebaran data menggunakan istilah dispersi. Berikut langkah-langkahnya: Input data → Exploration →

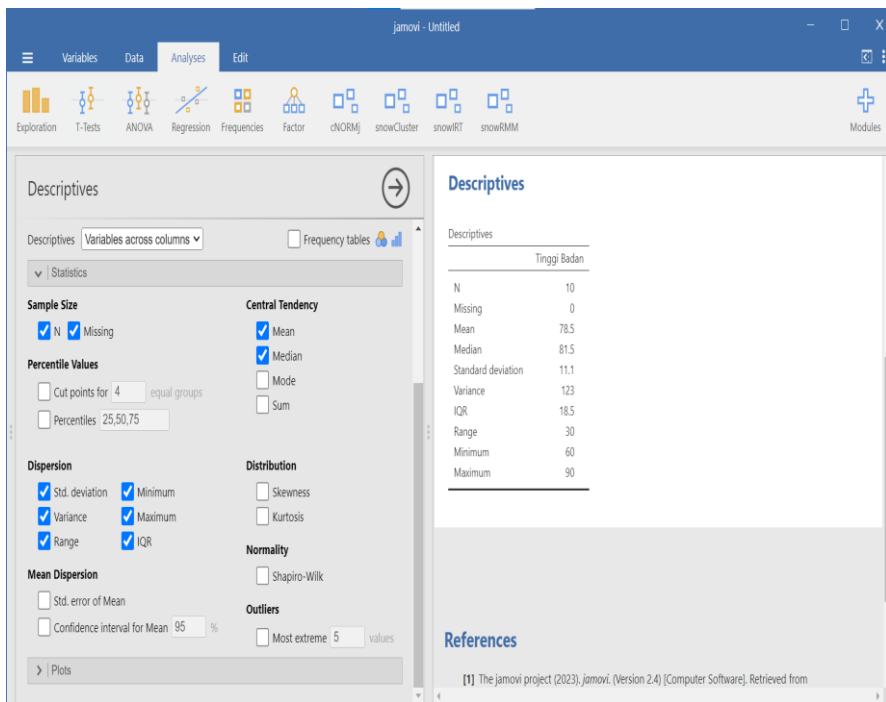
Descriptive → Pindahkan variabel yang ingin diolah → Statistics → Dispersion → Pilih menu dispersion yang diinginkan. Keseluruhan menu Sample Size dan Mean serta Median pada Central Tendency, akan otomatis tercentang oleh Jamovi.



Gambar 5.1. Input Data



Gambar 5.2. Memindahkan Variabel



**Gambar 5.3. Dispersi**

Contoh interpretasi berdasarkan luaran Jamovi di atas: Seorang peneliti ingin memahami distribusi tinggi badan dalam sampel kecil yang terdiri dari 10 peserta. Setelah mengumpulkan data dan menganalisisnya menggunakan metode statistik deskriptif, ia menemukan beberapa informasi menarik mengenai penyebaran tinggi badan dalam kelompok ini. Berdasarkan hasil analisis (lihat Tabel 1), rata-rata ( $M$ ) tinggi badan peserta adalah 78.5 cm ( $SD = 11.1$ ), dengan nilai median sebesar 81.5 cm. Ini menunjukkan bahwa separuh dari peserta memiliki tinggi badan di atas atau di bawah 81.5 cm, yang berarti distribusi data sedikit condong ke arah yang lebih rendah dibandingkan rata-rata. Lebih lanjut, data menunjukkan bahwa nilai minimum tinggi badan adalah 60 cm, sedangkan nilai maksimum mencapai 90 cm, menghasilkan rentang (range) sebesar 30 cm. Selain itu, *interquartile range* (IQR) sebesar 18.5 cm menunjukkan sebaran data di antara kuartil pertama dan ketiga, yang menggambarkan bagaimana mayoritas data tersebar di sekitar nilai tengah. Variabilitas dalam data juga ditunjukkan oleh variance sebesar 123, yang

mengindikasikan bahwa ada cukup banyak perbedaan tinggi badan antar peserta. Standar deviasi ( $SD = 11.1$ ) menunjukkan bahwa tinggi badan individu-individu dalam sampel ini bervariasi sekitar 11.1 cm dari rata-rata. Dari hasil ini, peneliti menyimpulkan bahwa terdapat variabilitas yang cukup tinggi dalam tinggi badan peserta dalam sampel kecil ini. Perbedaan antara nilai rata-rata dan median menunjukkan sedikit kecenderungan distribusi yang condong ke kiri (*negatively skewed*), yang mungkin dipengaruhi oleh adanya beberapa individu dengan tinggi badan lebih rendah dalam kelompok ini.

## DAFTAR PUSTAKA

- Ben-Zvi, D., & Garfield, J. (2004). Research on reasoning about variability. *Statistics Education Research Journal*, 3(2), 4-6.
- Bhardwaj, A., & Sharma, K. (2013). Comparative Study of Various Measures of Dispersion. *Journal of Advances in Mathematics*, 1(1), 6-9.
- Garfield, J., & Ben-Zvi, D. (2004). Research on statistical literacy, reasoning, and thinking. In D. Ben-Zvi & J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking* (pp. 397-409). Dordrecht, the Netherlands: Kluwer Academic Publishers.
- Gorard, S. (2005). Revisiting a 90-Year-Old Debate: The Advantages of the Mean Deviation. *British Journal of Educational Studies*, 53(4), 417-430.
- Joseph, S., Dyer, C., & Coolican, H. (2005). Means and Standard Deviations Explained. *Counselling and Psychotherapy Research*, 5(3), 256-257.  
<https://doi.org/10.1080/14733140500352280>
- Kelley, T. L. (2012). A New Measure of Dispersion. *Quarterly Publications of the American Statistical Association*, 117(134), 743-749. <https://doi.org/10.1080/15225445.1921.10503833>
- Manikandan, S. (2011). Measures of Dispersion. *Journal of Pharmacology and Pharmacotherapeutics*, 2(4), 315-316.  
<https://doi.org/10.4103/0976-500X.85931>
- Whaley, D. L. (2005). *The Interquartile Range: Theory and Estimation*. Master's Thesis, East Tennessee State University.
- Vetter, T. R. (2017). Descriptive Statistics: Reporting the Answers to the 5 Basic Questions of Who, What, Why, When, Where, and a Sixth, So What? *Anesthesia & Analgesia*, 125(5), 1797-1802.  
<https://doi.org/10.1213/ANE.00000000000002471>



# BAB 6

## PROBABILITAS DASAR

Oleh Vera Rosalina Bulu

### 6.1 Pengertian Probabilitas

Beberapa kejadian dalam kehidupan sehari – hari sulit diprediksi dengan pasti, apalagi kejadian tersebut adalah kejadian yang terjadi di masa mendatang. Walaupun sulit diketahui dengan pasti namun fakta – fakta yang ada dapat menuju derajat kepastian/derajat keyakinan. Derajat kepastian/ derajat keyakinan yang dikenal dengan istilah probabilitas.

Probabilitas berasal dari kata *probability* yang memiliki arti kemungkinan atau peluang sebuah kejadian akan terjadi. Probabilitas kejadian ditandai dengan interval  $0 < n < 1$ . Angka 0 menunjukkan peluang suatu kejadian tidak terjadi dan angka 1 menunjukkan peluang suatu kejadian pasti terjadi. Bila suatu kejadian X terjadi dalam N cara yang mungkin akan terjadi dimana masing – masing N cara mempunyai kesempatan atau kemungkinan yang sama untuk muncul, maka probabilitas kejadian E dapat ditulis dalam bentuk umum dari probabilitas yaitu :

$$P(E) = \frac{X}{N}$$

- P : probabilitas
- E : kejadian yang diharapkan
- X : banyaknya cara kejadian akan terjadi
- N : total kejadian yang mungkin terjadi

Konsep probabilitas memiliki manfaat untuk mengetahui kuantitas dari ketidakpastian, untuk membantu pengambilan keputusan yang tepat dari ketidakpastian yang ada.

Selain istilah probabilitas, terdapat juga istilah *percobaan* yaitu pengamatan terhadap beberapa aktivitas atau proses yang memungkinkan timbulnya paling sedikit dua peristiwa tanpa memperhatikan peristiwa mana yang akan terjadi. *Hasil (outcome)*



adalah hasil dari percobaan. Selanjutnya *peristiwa (event)* adalah kumpulan dari satu atau lebih hasil yang terjadi pada sebuah percobaan atau kegiatan.

Contoh:

Peluang untuk mendapatkan angka genap dari lemparan sebuah dadu. Jumlah kejadian A yaitu munculnya angka ganjil dalam 1 kali lemparan : 1, 3, 5 =  $n(A) = 3$ , dan jumlah seluruh kejadian yang mungkin terjadi dari 1 kali lemparan sebuah dadu: 1,2,3,4,5,6 =  $n(N) = 6$ , sehingga  $P(A) = \frac{n(A)}{n(N)} = \frac{3}{6} = \frac{1}{2}$ .

Berpeluang sama berarti kedua keadaan tersebut memiliki jumlah kemunculan kejadian yang sama.

## 6.2 Pendekatan probablilitas

Terdapat tiga pendekatan dalam probablilitas yaitu pendekatan klasik, relatif dan subjektif.

### 1. Pendekatan klasik

Pendekatan klasik menyatakan bahwa setiap peristiwa mempunyai kesempatan yang sama terjadi. Dengan kata lain jika dalam suatu eksperimen memiliki  $n$  kemungkinan hasil, maka pada tiap hasilnya mempunyai kemungkinan  $1/n$ .

Rumus yang digunakan yaitu

$$\text{Probabilitas suatu peristiwa} = \frac{\text{Jumlah kemungkinan hasil}}{\text{Jumlah total kemungkinan hasil}}$$

Contoh:

Pada suatu eksperimen pelemparan satu dadu, terdapat ruang sampel  $s = \{1,2,3,4,5,6\}$ . Maka setiap titik sampel memiliki probabilitas  $1/6$  untuk muncul.

### 2. Pendekatan relatif

Pendekatan relatif menyatakan bahwa probabilitas suatu kejadian tidak dianggap sama, tergantung dari berapa banyak suatu kejadian terjadi.

Rumus yang digunakan yaitu

$$\text{Probabilitas suatu peristiwa} = \frac{\text{Jumlah kemungkinan hasil}}{\text{Jumlah total percobaan}}$$

### 3. Pendekatan Subjektif

Pendekatan subjektif menyatakan bahwa probabilitas suatu kejadian didasarkan pada penilaian pribadi yang dinyatakan dalam suatu derajat kepercayaan.

## 6.3 Aturan dasar Probabilitas

### 1. Aturan penjumlahan (*Addition Rule*)

Aturan penjumlahan digunakan untuk menghitung probabilitas terjadinya satu atau beberapa peristiwa. Terdapat dua bentuk aturan penjumlahan:

- Aturan penjumlahan untuk peristiwa yang tidak dapat terjadi bersamaan (*mutually exclusive*). Jika peristiwa A dan B tidak dapat terjadi bersamaan, maka bentuk umum yang digunakan yaitu:

$$P(A \cup B) = P(A) + P(B) \quad P(A \cup B) = P(A) + P(B)$$

- Aturan penjumlahan untuk peristiwa yang dapat terjadi bersamaan (tidak *mutually exclusive*). Jika peristiwa A dan B dapat terjadi bersamaan, maka bentuk umum yang digunakan yaitu:

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ P(A \cup B) &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

Di mana  $P(A \cap B)$  adalah probabilitas peristiwa A dan B terjadi bersamaan.

### 2. Aturan perkalian (*Multiplication Rule*)

Aturan perkalian digunakan ketika kita ingin menghitung probabilitas terjadinya dua atau lebih peristiwa secara bersamaan.

- Aturan Perkalian untuk Peristiwa Independen (peristiwa yang tidak saling mempengaruhi):

$$P(A \cap B) = P(A) \times P(B) \quad P(A \cap B) = P(A) \times P(B)$$

- Aturan Perkalian untuk Peristiwa Bergantung (peristiwa yang saling mempengaruhi):

$$P(A \cap B) = P(A) \times P(B|A) \quad P(A \cap B) = P(A) \times P(B|A)$$

Di mana  $P(B|A)$  adalah probabilitas terjadinya B jika A sudah terjadi.

## 6.4 Sifat – Sifat Probabilitas

1. Bernilai antara 0 dan 1,  $0 \leq P(O_i) \leq 1$ , untuk setiap  $i$
2. Jumlahan probabilitas dari outcomes (semua hasil kemungkinan hasil eksperimen sama dengan 1)  
 $P(O_1) + P(O_2) + \dots + P(O_k) = 1$   
 $P(O_i)$  merupakan probabilitas outcome ke –  $i$

## 6.5 Teorema Bayes

Teorema Bayes digunakan untuk menghitung probabilitas terjadinya suatu peristiwa, berdasarkan pengaruh yang didapat dari hasil observasi peristiwa sebelumnya.

Teorema Bayes menyempurnakan teorema probabilitas bersyarat yang hanya dibatasi oleh 2 buah kejadian sehingga dapat diperluas untuk  $n$  buah kejadian.

Aplikasi teorema Bayes banyak ditemukan pada bidang komputer cerdas sebagai salah satu dasar dari metode *machine learning* dan *data mining*.

Jika kejadian  $B_1, B_2, B_3, \dots, B_k$  membentuk suatu partisi dari ruang sampel  $S$  sehingga  $p(B_i) \neq 0$  untuk  $i \in 1, 2, 3, \dots, k$ , maka untuk setiap kejadian  $A$  di  $S$  berlaku.

Teorema peluang total

$$P(A) = \sum_{i=1}^k p(B_i \cap A) = \sum_{i=1}^k p(B_i) \cdot P(A|B_i) \dots (1)$$

Probabilitas bersyarat suatu peristiwa  $A$ , dengan syarat peristiwa  $B$  didefinisikan sebagai:

Peluang bersyarat

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \dots (2)$$

Atau

$$P(B|A) = \frac{P(B \cap A)}{P(A)} \dots (3)$$

Sehingga diperoleh teori himpunan

$$P(A \cap B) = P(B \cap A) \dots (4)$$

Selanjutnya dari persamaan (3) dan (4) diperoleh

$$P(A \cap B) = p(A|B)p(B) = P(B|A)P(A) \dots (5)$$

$$P(B|A) = \frac{p(A|B)p(B)}{p(A)} \dots (6)$$

Dengan menggunakan teorema peluang total (1), aturan Bayes kemudian dapat dinyatakan dengan berlandaskan pada definisi peluang bersyarat (6), yaitu kasus ketika kita ingin mencari nilai dari  $p(B_i | A)$  maka diperoleh formula teorema Bayes

$$P(B_r|A) = \frac{p(B_r|A)}{p(A)} = \frac{p(B_r)p(A|B_r)}{\sum_i^k p(B_i)p(A|B_i)}$$

Atau

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

Aturan bayes menyatakan bahwa hasil dari hipotesis atau peristiwa (B) dapat diperkirakan berdasarkan pada beberapa evidence (A) yang diamati.

Hal penting dalam Bayes yaitu

1. Sebuah probabilitas awal/priori H atau  $P(B)$ , adalah probabilitas dari suatu hipotesis sebelum bukti diamati.
2. Sebuah probabilitas posterior H atau  $P(B|A)$ , adalah probabilitas dari suatu hipotesis setelah buktibukti yang diamati ada.

### ***Contoh konsep formula bayes***

Dalam suatu peramalan cuaca untuk memperkirakan terjadinya hujan, misal ada faktor yang mempengaruhi terjadinya hujan yaitu mendung.

Jika diterapkan dalam Naïve Bayes maka probabilitas terjadinya hujan jika bukti mendung sesudah diamati:

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

1.  $P(\text{hujan}|\text{mendung})$  adalah nilai probabilitas hipotesis hujan terjadi jika bukti mendung sudah diamati

2.  $P(\text{mendung}|\text{hujan})$  adalah probabilitas bahwa mendung yang diamati akan mempengaruhi terjadinya hujan.
3.  $P(\text{Hujan})$  adalah probabilitas awal hujan tanpa memandang bukti apapun.
4.  $P(\text{Mendung})$  adalah probabilitas terjadinya mendung.

### Contoh soal

Sebuah pabrik mempunyai 3 mesin A, B dan C yang memproduksi berturut turut 40%, 20% dan 10% dari total banyak unit yang diproduksi pabrik. Persentase kerusakan produk yang dihasilkan dari masing-masing mesin tersebut berturut turut adalah 1%, 4% dan 6%. Suatu unit dipilih secara random dan diketahui rusak. Hitung probabilitas bahwa unit tersebut berasal dari mesin C.

### Penyelesaian

Misal kejadian R adalah unit yang rusak, maka akan dihitung  $P(C|R)$  yaitu probabilitas bahwa suatu unit diproduksi oleh mesin C dengan diketahui unit tersebut rusak.

Dengan teorema Bayes, kejadian  $P(A)$ ,  $P(B)$  dan  $P(C)$  adalah peluang (persentase produksi) dari masing-masing mesin;  $P(R|A)$ ,  $P(R|B)$  dan  $P(R|C)$  adalah peluang (persentase kerusakan) dari masing-masing mesin.

$$P(C|R) = \frac{P(C)P(R|C)}{P(A)P(R|A) + P(B)P(R|B) + P(C)P(R|C)}$$

$$= \frac{(0,1)(0,04)}{(0,6)(0,02) + (0,3)(0,03) + (0,1)(0,04)} = \frac{4}{25}$$

## 6.6 Faktorial, Permutasi dan Kombinasi

### 1. Faktorial

*Faktorial* adalah perkalian antara bilangan bulat positif (bilangan asli) yang kurang dari atau sama dengan  $n$ . Faktorial ditulis sebagai  $n!$  dan disebut  $n$  faktorial. Secara umum dapat dituliskan sebagai:

$$n! = n(n-1) \cdot (n-2) \cdot (n-3) \cdot \dots \cdot 3 \cdot 2 \cdot 1$$

Contoh :

$$3! = 3 \cdot (3-1) \cdot (3-2) = 3 \cdot 2 \cdot 1 = 6$$

## 2. Permutasi

Permutasi adalah penyusunan atau pengaturan beberapa objek dalam urutan yang berbeda dari urutan yang semula.

Banyaknya permutasi dari  $n$  unsur diambil  $r$  unsur dinotasikan:

$${}_nP_r = n(n-1)(n-2)(n-3) \dots (n-r+1) \text{ atau}$$

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

Selanjutnya, banyaknya permutasi  $n$  unsur yang memuat  $k$ ,  $l$  dan  $m$  unsur yang sama dapat ditentukan dengan rumus

$$P = \frac{n!}{k! l! m!}$$

Jika permutasi yang cara menyusunnya melingkar, sehingga banyaknya menyusun  $n$  unsur yang berlainan dalam lingkaran disebut permutasi siklis dan ditulis dalam formulasi

$$\frac{n!}{n} = \frac{n(n-1)(n-2) \dots 3.2.1}{n} = (n-1)(n-2) \dots 3.2.1 \\ = (n-1)!$$

$$\text{Atau } P_{(\text{siklus})} = (n-1)!$$

## 3. Kombinasi

Kombinasi adalah pilihan yang dibuat dengan mengambil sebagian atau seluruh objek terlepas dari aturannya. Cara pemilihan objek dalam koleksi tersebut tidak mempermasalahkan urutannya.

Lambang notasi kombinasi adalah  $C$ . Jumlah anggota himpunan disebut dengan  $(n)$  dan jumlah objek yang harus dipilih disebut  $(r)$ . Dengan begitu, rumus kombinasi adalah

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

### Contoh soal

Terdapat himpunan huruf A,B,C,D. Akan dihitung susunan dengan dua huruf tanpa berurutan. Ada berapa banyak susunan dua huruf tersebut?

Dari contoh di atas, kamu bisa menggabungkan dengan manual

menjadi AB, AC, AD, BC, BD, CD. Namun jika menggunakan rumus, artinya dengan  $n = 4$  dan  $r=2$  maka :

$${}_4C_2 = \frac{4!}{(4-2)!2!} = \frac{4 \times 3 \times 2 \times 1}{2 \times 1 \times 2 \times 1} = 6$$

## DAFTAR PUSTAKA

- Adiwijaya. 2016. Matematika Diskrit dan Aplikasinya. Bandung: Telkom University.
- Budiyono. 2009. Statistika untuk Penelitian Edisi ke-2. Surakarta: Sebelas Maret University Press
- Munir R. 2005. Matematika Diskrit. Bandung: Informatika Bandung
- Walpole, R.E "Pengantar Statistik edisi 3". Jakarta, PT. Gramedia. 1995





# BAB 7

## DISTRIBUSI PROBABILITAS

Oleh Farida

### 7.1 Pendahuluan

Bilangan faktorial dan permutasi merupakan konsep fundamental yang sering digunakan dalam probabilitas. Faktorial digunakan untuk menghitung jumlah cara mengatur elemen dalam suatu himpunan, sedangkan permutasi membantu dalam menentukan cara menyusun objek dengan memperhatikan urutan.

#### 7.1.1 Bilangan Faktorial

Bilangan faktorial adalah hasil perkalian dari semua bilangan bulat positif hingga suatu bilangan tertentu  $n$ . Faktorial dinotasikan dengan simbol  $n!$  dan didefinisikan sebagai :

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1$$

Untuk  $n = 0$ , secara konvensional didefinisikan bahwa  $0! = 1$ .

Contoh :

- $5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$
- $3! = 3 \times 2 \times 1 = 6$
- $0! = 1$

#### 7.1.2 Permutasi

Permutasi adalah cara menyusun  $r$  objek dari  $n$  objek yang tersedia dengan memperhatikan urutan. Permutasi dinotasikan sebagai :

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

dengan:

- $n$  adalah jumlah total objek,
- $r$  adalah jumlah objek yang dipilih,
- $!$  menunjukkan faktorial.

Contoh :

Misalkan terdapat 5 objek dan kita ingin menyusun 3 di antaranya

$P(5,3) =$

$$\frac{5!}{(5-3)!} = \frac{5!}{2!} = \frac{5 \times 4 \times 3 \times 2!}{2!} = 5 \times 4 \times 3 = 60$$

### 7.1.3 Kesimpulan

Bilangan faktorial digunakan sebagai dasar perhitungan dalam permutasi dan berbagai konsep lainnya dalam matematika. Permutasi membantu dalam menentukan jumlah cara menyusun objek dengan mempertimbangkan urutan.

## 7.2 Konsep Dasar Probabilitas

### 7.2.1 Definisi Probabilitas

Probabilitas suatu kejadian didefinisikan sebagai ukuran kemungkinan terjadinya kejadian tersebut dalam suatu percobaan acak. Secara matematis, probabilitas suatu kejadian  $A$  dinyatakan sebagai :

$$P(A) = \frac{n(A)}{n(S)}$$

dengan:

- $P(A)$  adalah probabilitas kejadian  $A$ ,
- $n(A)$  adalah jumlah kejadian  $A$  yang mungkin terjadi,
- $n(S)$  adalah jumlah total kejadian dalam ruang sampel  $S$ .

Nilai probabilitas selalu berada dalam rentang  $0 \leq P(A) \leq 1$ , di mana 0 menunjukkan kejadian tidak mungkin terjadi, sedangkan 1 menunjukkan kejadian pasti terjadi.

### 7.2.2 Ruang Sampel (*Sample Space*)

Ruang sampel adalah himpunan semua kemungkinan hasil dari suatu percobaan. Ruang sampel biasanya dilambangkan dengan  $SS$ . Contoh:

- Melempar satu koin:  $S = \{H, T\}$
- Melempar dua dadu:  $S = \{(1,1), (1,2), \dots, (6,6)\}$

### 7.2.3 Kejadian (*Event*)

Kejadian adalah himpunan bagian dari ruang sampel yang berisi hasil-hasil tertentu yang menarik perhatian. Misalnya, dalam pelemparan dua dadu, kejadian mendapatkan jumlah mata dadu 7 adalah:  $A = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$

#### 1. Aturan Penjumlahan Probabilitas

Jika A dan B adalah dua kejadian yang saling lepas, maka probabilitas gabungan A atau B adalah :  $P(A \cup B) = P(A) + P(B)$

Jika A dan B tidak saling lepas, maka :  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

#### 2. Aturan Perkalian Probabilitas

Jika A dan B adalah dua kejadian independen, maka probabilitas keduanya terjadi bersama-sama adalah :  $P(A \cap B) = P(A) \times P(B)$

Jika A dan B tidak independen:  $P(A \cap B) = P(A) \times P(B|A)$  dengan  $P(B|A)$  adalah probabilitas B terjadi setelah A terjadi.

### 7.2.5 Kesimpulan

Konsep probabilitas membantu dalam memahami dan menganalisis kejadian-kejadian acak. Dengan memahami ruang sampel, aturan probabilitas, dan distribusi probabilitas, kita dapat membuat keputusan yang lebih baik dalam kondisi ketidakpastian.

## 7.3 Distribusi Probabilitas

Distribusi probabilitas adalah model matematis yang menggambarkan bagaimana probabilitas suatu variabel acak tersebar. Distribusi ini membantu dalam memahami bagaimana kemungkinan suatu kejadian tertentu terjadi dan digunakan secara luas dalam berbagai bidang seperti statistik, ekonomi, teknik, serta sains data.

Distribusi probabilitas dapat dikategorikan menjadi dua jenis utama:

1. Distribusi Diskrit: Digunakan untuk variabel acak diskrit, yang memiliki nilai tertentu dan dapat dihitung, seperti jumlah pelanggan yang datang ke toko dalam sehari.
2. Distribusi Kontinu: Digunakan untuk variabel acak kontinu, yang dapat mengambil nilai dalam suatu rentang, seperti tinggi badan atau waktu antara dua kejadian.

Memahami distribusi probabilitas sangat penting karena dapat digunakan untuk membuat prediksi, mengambil keputusan berdasarkan data, dan mengembangkan model dalam berbagai bidang ilmu. Dalam bahasan ini, kita akan membahas berbagai jenis distribusi probabilitas beserta penerapannya.

### **7.3.1 Variabel Acak**

Variabel acak adalah konsep fundamental dalam teori probabilitas dan statistik yang digunakan untuk merepresentasikan hasil dari suatu eksperimen acak. Pemahaman tentang variabel acak sangat penting dalam analisis data, model probabilistik, dan pengambilan keputusan berbasis probabilitas.

Secara umum, variabel acak dibagi menjadi dua jenis:

1. Variabel Acak Diskrit: Variabel yang hanya dapat mengambil nilai tertentu dalam jumlah terbatas atau dapat dihitung.
2. Variabel Acak Kontinu: Variabel yang dapat mengambil nilai dalam rentang yang tidak terbatas atau kontinu.

### **7.3.2 Variabel Acak Diskrit**

Variabel acak diskrit memiliki himpunan nilai yang dapat dihitung, seperti jumlah pelanggan dalam sehari atau jumlah kepala dalam pelemparan koin. Tiga distribusi probabilitas diskrit yang umum digunakan dalam berbagai bidang adalah Distribusi Binomial, Distribusi Poisson, dan Distribusi Hipergeometrik.

#### **1. Distribusi Binomial**

Distribusi Binomial adalah distribusi probabilitas diskrit yang menggambarkan jumlah keberhasilan dalam sejumlah percobaan independen yang masing-masing memiliki dua kemungkinan hasil (sukses atau gagal).

Syarat Distribusi Binomial

- a. Setiap percobaan hanya memiliki dua kemungkinan hasil (sukses atau gagal).
- b. Probabilitas sukses ( $p$ ) tetap untuk setiap percobaan.
- c. Setiap percobaan bersifat independen.
- d. Jumlah percobaan ( $n$ ) tetap.

Jika X adalah jumlah keberhasilan dalam nn percobaan dengan probabilitas sukses pp, maka distribusi binomial dinyatakan sebagai:

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

Dengan:

- n = jumlah percobaan
- k = jumlah sukses yang diinginkan
- p = probabilitas sukses dalam satu percobaan
- 1 - p = probabilitas kegagalan dalam satu percobaan
- $\binom{n}{k}$  = koefisien binomial, yaitu  $\frac{n!}{k!(n-k)!}$

Nilai Harapan dan Variansi

- Nilai harapan :  $E(X) = np$
- Variansi:  $\text{Var}(X) = np(1-p)$

Contoh Soal :

Sebuah dadu enam sisi dilempar 10 kali. Berapa probabilitas munculnya angka 6 sebanyak 3 kali ?

Diketahui :

- n=10
- k=3
- p=1/6

Gunakan rumus binomial untuk menyelesaikan.

## 2. Distribusi Poisson

Distribusi Poisson digunakan untuk menghitung probabilitas jumlah kejadian dalam suatu interval waktu atau ruang tertentu jika kejadian tersebut terjadi secara independen dengan rata-rata tertentu.

Syarat Distribusi Poisson

- Kejadian terjadi secara independen.
- Probabilitas terjadinya kejadian dalam selang waktu pendek sebanding dengan panjang selang waktu tersebut.
- Tidak ada dua kejadian yang terjadi dalam waktu yang sangat singkat secara bersamaan.

Jika X adalah jumlah kejadian dalam interval waktu tertentu dengan rata-rata kejadian  $\lambda$ , maka probabilitas nya sebesar :

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

Dengan:

- $\lambda$  = rata-rata kejadian dalam suatu interval
- k = jumlah kejadian yang diinginkan
- e = 2.71828 (bilangan Euler)

Nilai Harapan dan Variansi

- Nilai harapan:  $E(X)=\lambda$
- Variansi:  $\text{Var}(X)=\lambda$

Contoh Soal

Rata-rata jumlah kendaraan yang melewati sebuah jalan tol adalah 5 kendaraan per menit. Berapa probabilitas bahwa dalam satu menit ada tepat 7 kendaraan yang lewat?

Diketahui:

- $\lambda=5$
- k=7

Gunakan rumus distribusi Poisson untuk menyelesaikan.

### 3. Distribusi Hypergeometrik

Distribusi Hipergeometrik digunakan dalam kasus di mana pemilihan dilakukan tanpa pengembalian (*dependent trials*), misalnya dalam pengambilan sampel tanpa penggantian dari suatu populasi terbatas.

Syarat Distribusi Hipergeometrik

- Populasi terdiri dari dua jenis elemen: sukses dan gagal.
- Sampel diambil tanpa pengembalian.
- Probabilitas keberhasilan berubah setelah setiap pemilihan.

Jika dari populasi sebesar yang mengandung elemen sukses, kita mengambil sampel berukuran, maka probabilitas mendapatkan tepat elemen sukses diberikan oleh :

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

- N = total populasi

- $K$  = jumlah elemen sukses dalam populasi
- $n$  = ukuran sampel
- $k$  = jumlah sukses yang diinginkan dalam sampel

Nilai Harapan dan Variansi

- Nilai harapan:  $E(X) = n \frac{K}{N}$
- Variansi:  $Var(X) = n \frac{K}{N} \frac{N-K}{N} \frac{N-n}{N-1}$

Contoh Soal

Dalam sebuah kotak terdapat 10 bola merah dan 20 bola biru. Jika kita mengambil 5 bola secara acak tanpa pengembalian, berapa probabilitas bahwa 3 di antaranya adalah bola merah?

Diketahui :

- $N=30$
- $K=10$
- $n=5$
- $k=3$

Gunakan rumus distribusi hipergeometrik untuk menyelesaikan.

### 7.3.2 Variabel Acak Kontinu

Variabel acak kontinu adalah variabel yang dapat mengambil nilai dalam suatu interval tertentu, di mana kemungkinan untuk memperoleh nilai tertentu dalam interval itu tidak nol tetapi probabilitasnya dihitung dalam rentang nilai. distribusi probabilitas penting untuk variabel acak kontinu meliputi distribusi normal dan distribusi eksponensial.

#### 1. Distribusi Normal.

Distribusi normal merupakan distribusi probabilitas kontinu yang paling umum dan banyak digunakan dalam berbagai bidang seperti statistik, keuangan, fisika, dan ilmu sosial.

Distribusi ini sering disebut sebagai distribusi Gauss karena pertama kali dipelajari oleh Carl Friedrich Gauss.

Sebuah variabel acak kontinu dikatakan berdistribusi normal dengan parameter rata-rata  $\mu$  dan variansi  $\sigma^2$  jika fungsi kepadatan probabilitasnya (PDF) diberikan oleh :



$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad -\infty < x < \infty$$

Keterangan:

- $\mu$  = rata-rata (mean) distribusi
- $\sigma^2$  = variansi
- $\sigma$  = simpangan baku (standar deviasi)

#### Sifat-Sifat Distribusi Normal

- a. Bentuk Simetris: Kurva distribusi normal berbentuk lonceng dan simetris terhadap rata-rata  $\mu$ .
- b. Mean, Median, dan Modus Sama : Ketiga ukuran pemusatan ini berada pada titik yang sama, yaitu  $\mu$ .
- c. Asimtotik terhadap Sumbu X : Kurva tidak pernah benar-benar menyentuh sumbu X, meskipun semakin mendekati nilai nol di kedua ekornya.
- d. Total Luas di Bawah Kurva: Integral dari fungsi PDF pada seluruh domainnya selalu sama dengan 1.

#### Distribusi normal standar :

Merupakan distribusi normal dengan rata-rata  $\mu = 0$  dan simpangan baku  $\sigma = 1$ .

Setiap distribusi normal X dengan parameter  $\mu$  dan  $\sigma$  dapat diubah menjadi distribusi normal standar Z dengan transformasi :

$$Z = \frac{X - \mu}{\sigma}$$

#### Contoh Soal dan Penyelesaian

##### Menghitung Probabilitas dalam Distribusi Normal Standar

Seorang siswa mengikuti ujian dengan skor rata-rata dan simpangan baku. Tentukan probabilitas bahwa seorang siswa mendapatkan skor di bawah 85.

Penyelesaian :

- a. Standarisasi nilai ke dalam :

$$Z = \frac{85 - 70}{10} = 1.5$$

b. Cari nilai dari tabel distribusi normal:

$$P(Z \leq 1.5) = 0.9332$$

c. Jadi, probabilitas bahwa seorang siswa mendapatkan skor di bawah 85 adalah 0.9332 atau 93.32%.

## 2. Distribusi Eksponensial.

Distribusi eksponensial adalah salah satu distribusi probabilitas kontinu yang sering digunakan untuk memodelkan waktu antara kejadian dalam suatu proses Poisson. Distribusi ini banyak diterapkan dalam bidang teknik, ekonomi, dan ilmu komputer, seperti dalam analisis waktu kegagalan (*failure time analysis*), waktu tunggu layanan, dan model keandalan sistem.

Model untuk waktu antara kejadian dalam suatu proses.

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

Keterangan:

- $\lambda$  = rata-rata tingkat kejadian per unit waktu (rate parameter)
- $x$  = waktu antara dua kejadian
- $e$  = basis logaritma natural (sekitar 2.718)

Sifat utama dari distribusi eksponensial adalah bahwa nilai probabilitas terbesar terjadi pada  $x = 0$ , dan semakin besar nilai  $x$ , semakin kecil probabilitasnya.

Rataan dan variansi dari distribusi eksponensial diberikan oleh:

- Nilai harapan (*mean*) :

$$E(X) = \frac{1}{\lambda}$$

- Variansi :

$$Var(X) = \frac{1}{\lambda^2}$$

## Contoh Soal dan Penyelesaian

### Contoh 1: Menghitung Probabilitas

Sebuah lampu memiliki waktu hidup yang mengikuti distribusi eksponensial dengan  $\lambda = 0,2$  kegagalan per jam. Tentukan probabilitas bahwa lampu akan bertahan lebih dari 5 jam.

Penyelesaian:

Diketahui  $\lambda = 0,2$  dan kita ingin mencari  $P(X > 5)$ .

$$P(X > 5) = 1 - P(X \leq 5) = e^{-\lambda x} = e^{-0.2(5)}$$

$$P(X > 5) = e^{-1} \approx 0.3679$$

Jadi, probabilitas bahwa lampu bertahan lebih dari 5 jam adalah 0.3679 atau 36.79%.

Contoh 2: Menghitung Rataan Waktu Tunggu

Dalam suatu layanan pelanggan, waktu antara panggilan masuk mengikuti distribusi eksponensial dengan  $\lambda = 4$  panggilan per jam. Berapa rata-rata waktu antara panggilan masuk?

Penyelesaian:

Rataan waktu tunggu diberikan oleh:

$$E(X) = \frac{1}{\lambda} = \frac{1}{4} = 0.25 \text{ jam} = 15 \text{ menit}$$

Jadi, rata-rata waktu antara panggilan masuk adalah 15 menit.

Aplikasi Distribusi Eksponensial

1. Distribusi eksponensial banyak digunakan dalam berbagai bidang, antara lain:
2. Keandalan dan Pemeliharaan: Memodelkan waktu kegagalan komponen elektronik atau mekanis.
3. Sistem Antrian: Menentukan waktu tunggu dalam sistem layanan pelanggan, seperti call center dan bandara.
4. Telekomunikasi: Memodelkan waktu antara kedatangan paket data dalam jaringan komputer.
5. Analisis Risiko: Memodelkan waktu hingga terjadi peristiwa seperti klaim asuransi atau kecelakaan

## **DAFTAR PUSTAKA**

Montgomery, D. C., & Runger, G. C. (2018). Applied Statistics and Probability for Engineers (7th ed.). Wiley.



# BAB 8

## SAMPLING DAN POPULASI

Oleh Herry Agung Prabowo

### 8.1 Pendahuluan

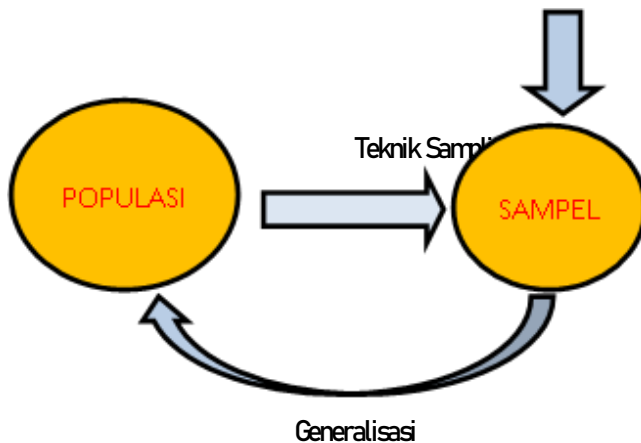
Dalam setiap penelitian, pemahaman tentang populasi dan teknik sampling sangat penting untuk memastikan validitas dan reliabilitas hasil yang diperoleh. Populasi mengacu pada keseluruhan objek atau individu yang menjadi subjek penelitian, sedangkan sampling merupakan metode pengambilan sebagian dari populasi untuk dianalisis. Penggunaan teknik sampling yang tepat dapat menghemat waktu, biaya, serta meningkatkan efisiensi penelitian tanpa mengurangi kualitas hasil yang diperoleh. (Creswell, 2021). Beberapa pengertian populasi menurut para ahli:

1. Menurut Ismiyanto, populasi mencakup seluruh subjek dalam penelitian, baik itu individu, benda, maupun aspek lain yang dapat memberikan atau menjadi sumber informasi bagi penelitian.
2. Sementara itu, Arikunto mendefinisikan populasi sebagai keseluruhan objek penelitian. Jika seorang peneliti mempelajari semua elemen dalam lingkup penelitian, maka penelitian tersebut dikategorikan sebagai penelitian populasi.
3. Adapun menurut Sugiyono, populasi merupakan suatu wilayah generalisasi yang terdiri dari objek atau subjek dengan karakteristik dan jumlah tertentu yang telah ditentukan oleh peneliti untuk dianalisis guna memperoleh kesimpulan.

Sampel atau contoh adalah bagian dari populasi yang karakteristiknya hendak diteliti. (himpunan bagian dari populasi). Sampel yang baik, adalah sampel yang dapat merepresentasikan atau menggambarkan karakteristik populasi secara benar dan akurat. Sampel sesungguhnya adalah sebagian kecil saja dari populasi yang memiliki sifat/karakteristik tertentu dan dipilih berdasarkan tujuan peneliti dengan prosedur tertentu agar dapat mewakili populasi. Dalam penelitian dengan populasi yang besar, sering kali tidak

memungkinkan untuk meneliti seluruh elemen yang ada karena minimannya sumber daya, seperti waktu, tenaga dan juga biaya. Oleh karena itu, peneliti dapat menggunakan sampel sebagai perwakilan populasi. Namun, sampel yang dipilih harus benar-benar representatif agar hasil penelitian dapat menggambarkan kondisi populasi secara akurat.

Lihat gambar 8.1 untuk ilustrasinya.



**Gambar 8.1.** Populasi dan Sampel

## 8.2 Populasi dalam Penelitian

Sel Populasi dalam penelitian adalah kelompok individu, objek, atau entitas lain dengan sifat/karakteristik tertentu yang menjadi fokus penelitian, dan dapat dikategorikan menjadi:

1. Populasi Target, yaitu seluruh kelompok yang memiliki karakteristik tertentu yang ingin diteliti.
2. Populasi Terjangkau, yaitu bagian dari populasi target yang secara praktis dapat dijangkau oleh peneliti untuk dianalisis. (Saunders, 2019)

Pemahaman yang baik mengenai karakteristik populasi sangat penting dalam menentukan teknik sampling yang akan digunakan, sehingga dapat menghindari bias dalam penelitian.

## 8.3 Teknik Sampling

Untuk melakukan suatu aktivitas riset langkah pertama adalah memilih metode yang paling tepat untuk *data collecting*-nya. Aktivitas *data collecting* yang dilakukan dalam ilmu statistik disebut dengan metode/teknik sampling. Teknik Sampling yaitu merupakan teknik pengambilan sampel, yaitu proses pemilihan sejumlah individu atau objek dari populasi untuk mewakili keseluruhan populasi.

Beberapa manfaat/alasan dilakukannya sampling:

1. Terkait biaya penelitian> akan lebih hemat.
2. Terkait waktu penelitian> akan lebih cepat/singkat
3. Terkait akurasi data> akan lebih akurat.
4. Terkait ruang lingkup penelitian> bisa lebih luas.

Teknik sampling terbagi menjadi dua kategori utama yaitu *probability sampling* dan *non-probability sampling*.

### 1. Sampling Probabilitas (*Probability Sampling*)

Sampling probabilitas digunakan ketika peluang yang sama untuk dipilih ada pada setiap anggota populasi. Teknik ini cukup banyak digunakan dalam penelitian kuantitatif karena dapat digunakan untuk generalisasi hasil ke seluruh populasi. Beberapa teknik sampling probabilitas meliputi:

- a. **Simple Random Sampling:** Setiap elemen dalam populasi memiliki peluang yang sama untuk dipilih secara acak. (Taherdoost, 2016).
- b. **Stratified Random Sampling:** Populasi dibagi menjadi beberapa strata berdasarkan karakteristik tertentu, lalu sampel diambil secara acak dari masing-masing strata secara proporsional jumlahnya.
- c. **Disproportionate Statified Sampling:** Teknik ini menggunakan metode sampel bertingkat, di mana populasi dikelompokkan ke dalam beberapa strata yang bersifat homogen. Dari setiap strata, dipilih sejumlah sampel, namun jumlahnya tidak selalu proporsional. Sebagai ilustrasi, seorang manajer SDM ingin menganalisis rata-rata pendapatan bulanan karyawan di perusahaannya. Untuk memperoleh data tersebut,



karyawan dikelompokkan berdasarkan tingkat pendidikan sebagai berikut:

Latar Pendidikan	Karyawan (Orang)	Sampel
Sekolah Dasar-Menengah	20	10
Menengah Atas	250	50
Diploma/PolTek	220	30
Universitas	70	10

- d. **Cluster Sampling:** Populasi dibagi menjadi kelompok-kelompok (cluster), lalu beberapa cluster dipilih secara acak untuk dianalisis dan pada salah satu cluster diambil sebagai sampel. Lazimnya metode ini menggunakan dasar/berbasis geografis. (Etikan & Bala, 2017). Ciri utama dari kelompok sampel ini adalah memiliki sifat heterogen dalam satu kelompok, namun variasinya relatif kecil antar kelompok (varian kecil). Sebagai ilustrasi, ada lembaga survei ingin meneliti jumlah Hp per keluarga (rata-rata) di suatu kota/kabupaten. Dalam penelitian ini, kota/kabupaten tersebut dibagi menjadi lima kecamatan, tetapi hanya satu kecamatan yang dipilih sebagai sampel.

## 2. Sampling Non-Probabilitas

Sampling non-probabilitas digunakan ketika tidak semua anggota populasi memiliki peluang yang sama untuk dipilih. Teknik ini lebih umum dalam penelitian eksploratif atau kualitatif. Beberapa jenisnya meliputi:

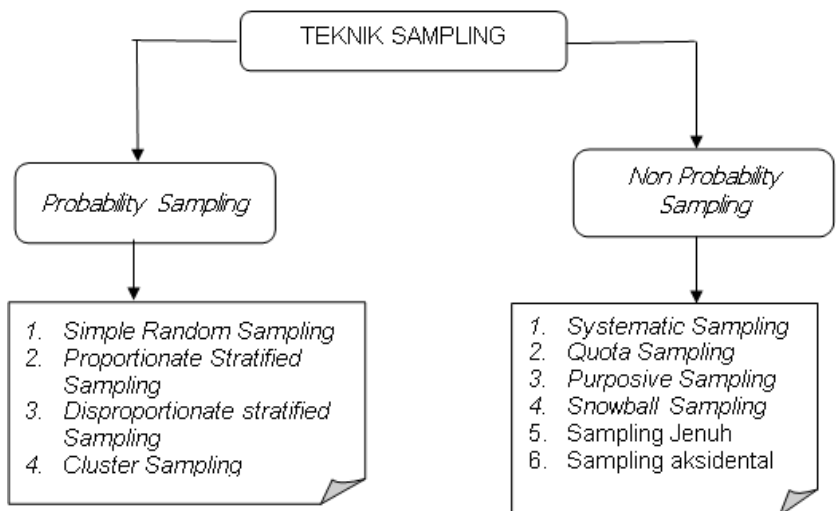
- a. ***Convenience Sampling.*** Sampel dipilih berdasarkan kemudahan akses.

- b. ***Systematic Sampling.***

Pemilihan anggota sampel dalam metode ini dilakukan dengan interval yang seragam, baik berdasarkan waktu maupun urutan tertentu. Contohnya, sampel dapat diambil pada interval waktu tertentu dalam sebuah penelitian, seperti pencatatan data setiap satu jam sekali. Contoh lainnya adalah pemilihan sampel berdasarkan nomor urut, misalnya dalam sebuah studi akademik, sampel dipilih dari nilai ujian mahasiswa yang memiliki nomor absen dengan akhiran angka ganjil. Metode ini memastikan bahwa sampel

tersebar secara merata dalam populasi tanpa harus memilih secara acak sepenuhnya.

- c. **Purposive Sampling.** Sampel dipilih berdasarkan kriteria tertentu yang sesuai dengan tujuan penelitian. Maksudnya penentuan sample sudah dipilih terlebih dahulu berdasarkan tujuan penelitian, jadi tidak acak/random. Contoh: Jika ingin membandingkan dari dua Klink yang murni dikelola swasta atau pemerintah, maka bila ada Klinik yang pengelolanya adalah swasta dan pemerintah secara bersama-sama tidak bisa dijadikan sampel
- d. **Snowball Sampling.** Sampel diperoleh melalui rekomendasi dari responden sebelumnya. (Palinkas et al., 2015).
- e. **Quota Sampling.** Sampel diambil hingga mencapai jumlah tertentu dalam kategori tertentu.
- f. **Sampling Jenuh:** Penambahan jumlah sampel secara terus-menerus tidak akan ada pengaruhnya terhadap hasil/nilai informasi yang didapat sebelumnya.



**Gambar 8.2.** Pembagian Teknik Sampling

## Ukuran Sampel dan Kesalahan Sampling

Ukuran sampel sangat menentukan validitas hasil penelitian. Beberapa metode penentuan ukuran sampel yang umum digunakan adalah:

- a. **Rumus Slovin:** Digunakan untuk menentukan jumlah sampel berdasarkan margin of error tertentu. (Yamane, 2020).  
Rumus slovin dapat dituliskan sebagai:

$$n = N / (1 + Ne^2)$$

Keterangan:

- n: minimum jmlh sampel
- N: ukuran/jumlah populasi
- e: tingkat/batas kesalahan (*error*)

Kemudian pada bagian di bawah ini terdapat contoh perhitungan menggunakan rumus slovin.

1. Sebuah lembaga ingin mengkaji pengaruh polusi udara terhadap kondisi paru-paru warga di desa Sukatani. Dari data aparat desa, jumlah warga adalah sekitar 1.000 orang. Untuk menentukan jumlah sampel yang representatif, peneliti menetapkan batas toleransi kesalahan sebesar 10%.

Pembahasan:

- Total populasi (N) = 1.000
- Batas toleransi kesalahan (e) = 10% = 0,1

Dengan menggunakan rumus Slovin, jumlah sampel minimum yang diperlukan dapat dihitung sebagai berikut:

$$n = 1000 / (1 + 1000 \times (0,1)^2)$$

$$n = 1000 / 11$$

$$n = 90,9$$

Sehingga jumlah sampel minimum yang diperlukan adalah 91 orang (dibulatkan ke atas).

- a. **Krejcie & Morgan Table:** Digunakan dalam penelitian kuantitatif untuk menentukan jumlah sampel dari populasi besar. Berikut ini adalah tabelnya:

Table 3.1

Table for Determining Sample Size of a Known Population

N	S	N	S	N	S	N	S	N	S
10	10	100	80	280	162	800	260	2800	338
15	14	110	86	290	165	850	265	3000	341
20	19	120	92	300	169	900	269	3500	346
25	24	130	97	320	175	950	274	4000	351
30	28	140	103	340	181	1000	278	4500	354
35	32	150	108	360	186	1100	285	5000	357
40	36	160	113	380	191	1200	291	6000	361
45	40	170	118	400	196	1300	297	7000	364
50	44	180	123	420	201	1400	302	8000	367
55	48	190	127	440	205	1500	306	9000	368
60	52	200	132	460	210	1600	310	10000	370
65	56	210	136	480	214	1700	313	15000	375
70	59	220	140	500	217	1800	317	20000	377
75	63	230	144	550	226	1900	320	30000	379
80	66	240	148	600	234	2000	322	40000	380
85	70	250	152	650	242	2200	327	50000	381
90	73	260	155	700	248	2400	331	75000	382
95	76	270	159	750	254	2600	335	100000	384

Note: N is Population Size; S is Sample Size

Source: Krejcie &amp; Morgan, 1970

Selain itu, kesalahan dalam sampling juga perlu diperhatikan, yaitu:

1. **Sampling Error:** Kesalahan yang terjadi karena hanya sebagian populasi yang diambil sebagai sampel.
2. **Non-Sampling Error:** Kesalahan yang terjadi karena faktor lain, seperti kesalahan pengukuran atau bias responden. (Bryman, 2021)

### Aplikasi dan Contoh dalam Penelitian

Sebagai contoh, dalam penelitian tentang efektivitas strategi pemasaran digital di kalangan UMKM, populasi targetnya adalah semua UMKM di Indonesia, sedangkan populasi terjangkaunya adalah UMKM di lima kota besar. Dengan menggunakan stratified random sampling, peneliti dapat membagi populasi berdasarkan sektor usaha, lalu mengambil sampel secara acak dari masing-masing sektor untuk mendapatkan hasil yang lebih representatif. (Zikmund, et al., 2022).

### ***Destructive Test (DT) dan Non-Destructive Test (NDT) Sampling***

Di dalam penelitian yang pengambilan sampelnya bersifat NDT maka jumlah sampel yang besar mungkin saja tidak menjadi masalah namun untuk pengambilan sampel dengan cara DT, maka ukuran sampel minimumlah yang harus dicari untuk menghindari biaya sampel yang tinggi mengingat sampel hasil DT biasanya tidak bisa digunakan Kembali.

## **8.4 Kesimpulan**

Memahami konsep populasi dan teknik sampling sangat penting dalam memastikan validitas dan reliabilitas penelitian. Pemilihan metode sampling yang tepat dapat meningkatkan efisiensi penelitian, mengurangi bias, serta memastikan bahwa hasil penelitian dapat digeneralisasi dengan baik ke populasi yang lebih luas. Oleh karena itu, peneliti harus mempertimbangkan dengan cermat karakteristik populasi dan tujuan penelitian sebelum memilih teknik sampling yang akan digunakan.

## DAFTAR PUSTAKA

- Bryman, A. (2021). *Social Research Methods*. Oxford University Press.
- Creswell, J. W. (2021). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE Publications.
- Etikan, I., & Bala, K. (2017). Sampling Methods and Sample Size Calculation. *Biostatistics & Epidemiology*, 2(1), 1-7.
- Harinaldi, Prinsip Prinsip Statistik Untuk Teknik dan Sains. (2002). Penerbit Erlangga. Jakarta.
- Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Administration and Policy in Mental Health and Mental Health Services Research*, 42(5), 533-544.
- Saunders, M., Lewis, P., & Thornhill, A. (2019). *Research Methods for Business Students*. Pearson Education.
- Sugiyono. (2020). *Buku Statistika Untuk Penelitian*. Penerbit Alfabeta.
- Taherdoost, H. (2016). Sampling Methods in Research Methodology; How to Choose a Sampling Technique for Research. *International Journal of Academic Research in Management (IJARM)*, 5(2), 18-27.
- Walpole, Ronald. E. (2000). *Pengantar Statistik*. Penerbit Erlangga. Jakarta.
- Yamane, T. (2020). *Statistics: An Introductory Analysis*. Harper & Row.
- Zikmund, W. G., Babin, B. J., Carr, J. C., & Griffin, M. (2022). *Business Research Methods*. Cengage Learning.
- [https://www.researchgate.net/profile/Syed-Abdul-Bukhari/publication/349118299\\_Sample\\_Size\\_Determination\\_Using\\_Krejcie\\_and\\_Morgan\\_Table/links/60215fb492851c4ed55b6bd8/Sample-Size-Determination-Using-Krejcie-and-Morgan-Table.pdf](https://www.researchgate.net/profile/Syed-Abdul-Bukhari/publication/349118299_Sample_Size_Determination_Using_Krejcie_and_Morgan_Table/links/60215fb492851c4ed55b6bd8/Sample-Size-Determination-Using-Krejcie-and-Morgan-Table.pdf)



# **BAB 9**

## **DISTRIBUSI SAMPLING**

**Oleh Agung Wahyudi Biantoro**

Distribusi sampling adalah distribusi probabilitas dari suatu statistik sampel (misalnya, rata-rata, varians, atau proporsi) yang dihitung dari banyak sampel acak yang diambil dari suatu populasi. Distribusi ini menunjukkan bagaimana nilai-nilai statistik sampel tersebar dan digunakan untuk melakukan inferensi statistik terhadap populasi. Distribusi sampling bertujuan untuk mempelajari karakteristik populasi berdasarkan pada sampel yang diambil dari populasi tersebut secara random (setiap unsur dalam populasi memiliki kesempatan yang sama untuk diikutsertakan ke dalam sampel).

### **9.1 Populasi Dan Sampel**

#### **9.1.1 Populasi**

Populasi adalah totalitas semua nilai yang mungkin, baik hasil menghitung ataupun pengukuran kuantitatif maupun kualitatif daripada karakteristik tertentu mengenai sekumpulan objek yang lengkap (Nawawi, 1985 : 141).

Populasi merupakan seluruh subjek atau objek penelitian yang mempunyai dengan karakteristik tertentu, bersifat homogen yang akan ditetapkan oleh peneliti untuk dipelajari dan akan ditarik kesimpulannya. Bukan hanya objek atau subjek yang dipelajari saja tetapi seluruh karakteristik atau sifat yang dimiliki subjek atau objek tersebut. populasi adalah keseluruhan kelompok individu, objek, atau data yang menjadi fokus dalam suatu penelitian. Populasi bisa terdiri dari orang, benda, kejadian, atau pengukuran yang memiliki karakteristik tertentu yang ingin diteliti.

Menentukan populasi adalah hal yang sangat penting sebelum melakukan penelitian. Dengan menentukan populasi maka kita dapat membatasi penelitian dan tidak melebar ke wilayah-wilayah lain dalam pengambilan sampel, sehingga bisa menghemat biaya, waktu,



dan tenaga. Contoh dalam populasi adalah : Sekelompok mahasiswa Univ. ABC, sekelompok peneliti teknik Universitas XYZ, sejumlah kelompok Majelis taklim yang berlokasi di Bantul, Yogyakarta dan lain-lain.

Namun demikian arti populasi tidak hanya terbatas pada kelompok orang saja. Sedikit orang atau bahkan satu orang saja pun dapat disebut dengan populasi. Menurut Sugiyono (2003: 90) satu orang pun dapat digunakan sebagai populasi, karena satu orang itu mempunyai berbagai karakteristik, misalnya gaya bicaranya, disiplin pribadi, hobi, cara bergaul, kepemimpinannya dan lain-lain. Bila peneliti akan melakukan penelitian tentang kepribadian seorang ulama sekaligus umaro yaitu Menteri Sosial RI di jajaran menteri Kabinet Indonesia Bersatu (KIB), maka kepribadian itu merupakan sampel dari semua karakteristik yang dimiliki menteri tersebut.

Dalam bidang kesehatan khususnya kedokteran, satu orang sering bertindak sebagai populasi. Pada saat perawat kesehatan mengambil darah dari pasien yang akan diperiksa, maka darah yang diambil tersebut adalah populasi. Sedangkan kalau akan diperiksa cukup diambil sebagian darah yang berupa sampel. Data yang diteliti dari sampel tersebut selanjutnya diberlakukan ke seluruh darah yang dimiliki orang tersebut.

Populasi terbagi menjadi 2 jenis, yaitu:

1. **Populasi terbatas** adalah suatu populasi yang unurnya memiliki batasan jelas secara kuantitatif berukuran  $N$  sehingga dapat dihitung dan jumlahnya dapat dinyatakan dalam bilangan.  
Contoh: populasi bank di Jawa Barat, jumlah mahasiswa baru FEB Universitas PQR tahun 2019
2. **Populasi tidak terbatas** adalah suatu populasi yang mengalami proses secara terus menerus sehingga ukuran  $N$  menjadi tidak terbatas perubahan nilainya dan jumlahnya tidak dapat dinyatakan dalam bilangan. Contoh: populasi bintang di langit.

### 9.1.2 Sampel

Sampel adalah sejumlah atau bagian dari populasi yang diteliti, yang terwakili. Sampel merupakan bagian atau cuplikan dari populasi yang terwakili dan akan diteliti atau sebagian jumlah dari karakteristik yang dimiliki oleh populasi yang terwakili. Bila populasi

berjumlah 500 orang, maka 100 orang bisa disebut sebagai sampelnya bila peneliti yakin itu adalah bagian dari populasi yang mewakilinya. Dari jumlah populasi 500 itu, maka sampel juga dapat diambil 500 orang juga, yang mana ini disebut sebagai teknik sampling jenuh atau teknik sensus.

Sampel terbagi menjadi 2 jenis, yaitu:

1. **Sampel Probabilitas** atau *Random Sample* atau Peluang merupakan suatu sampel yang dipilih sedemikian rupa dari populasi sehingga masing-masing anggota populasi memiliki probabilitas atau peluang yang sama untuk dijadikan sampel (random; sistematis).
2. **Sampel Nonprobabilitas** atau *Nonrandom Sample* merupakan suatu sampel yang dipilih sedemikian rupa dari populasi sehingga setiap anggota tidak memiliki probabilitas atau peluang yang sama untuk dijadikan sampel.

## 9.2 Sampling

**Sampling** adalah proses pemilihan sejumlah individu, objek, atau unit dari suatu populasi untuk dijadikan sebagai sampel dalam penelitian. Tujuan sampling adalah untuk mewakili karakteristik populasi tanpa harus meneliti seluruh populasi, sehingga analisis menjadi lebih efisien dan praktis. Dalam mempelajari karakteristik populasi, biasanya dilakukan sampling karena beberapa alasan, diantaranya:

1. Memakan waktu banyak untuk mendapatkan data dari seluruh populasi
2. Biaya untuk meneliti semua bagian dari populasi kemungkinan terlalu tinggi
3. Ketidakmungkinan untuk meneliti semua bagian dari populasi
4. Beberapa percobaan bersifat merusak
5. Hasil dari sampel sudah memadai

### Metode Sampling:

1. ***Simple Random Sampling***  
Setiap anggota dari suatu populasi tersebut memiliki kondisi dan kesempatan yang sama untuk dijadikan sampel.

## 2. *Systematic Random Sampling*

Memilih starting point dari suatu populasi, mengurutkan setiap anggota dari populasi tersebut dengan cara tertentu (*alphabetically, numerically, dan lainnya*), kemudian sampel diambil pada urutan tertentu atau kesekian sampai jumlah sampel mencukupi.

## 3. *Stratified Random Sampling*

Sebuah populasi dibagi menjadi strata-strata atau kelompok-kelompok tertentu, kemudian sampel diambil secara random dari setiap strata atau kelompok.

## 4. *Cluster Sampling*

Sebuah populasi dibagi menjadi beberapa *cluster* atau kelompok berdasarkan geografis, dipilih beberapa *cluster* secara random, kemudian dari setiap cluster yang terpilih diambil sampel secara random.

# 9.3 Distribusi Sampling

**Distribusi Sampling** adalah suatu probabilitas distribusi dari semua rata-rata sampel yang memungkinkan dari sebuah ukuran sampel. Distribusi sampling dilakukan untuk mempelajari karakteristik populasi berdasarkan sampel dari populasi tersebut.

Terdapat 4 jenis distribusi sampling, yaitu:

1. Distribusi Sampling Rata-Rata
2. Distribusi Sampling Proporsi
3. Distribusi Sampling Selisih Rata-Rata
4. Distribusi Sampling Selisih Proporsi

## A. Distribusi Sampling Rata-Rata

Distribusi Sampling rata-rata adalah suatu distribusi probabilitas dari besaran rata-rata hitung yang muncul dari sampel dan probabilitas terjadinya dihubungkan dengan setiap rata-rata hitung sampel. Distribusi sampling rata-rata adalah distribusi probabilitas yang menunjukkan bagaimana nilai rata-rata sampel tersebar ketika kita mengambil banyak sampel acak dari suatu populasi. Distribusi ini menggambarkan variasi rata-rata sampel dibandingkan dengan rata-rata populasi.

Berdasarkan Teorema Limit Tengah (*Central Limit Theorem*), jika ukuran sampel cukup besar, maka distribusi sampling dari rata-rata sampel akan mendekati distribusi normal, terlepas dari bentuk distribusi populasi awal. Selain itu, rata-rata dari distribusi sampling rata-rata akan sama dengan rata-rata populasi, sedangkan simpangan bakunya akan lebih kecil dan dihitung menggunakan standar error. Distribusi sampling rata-rata sangat berguna dalam statistik inferensial karena memungkinkan kita untuk membuat estimasi tentang rata-rata populasi hanya dengan menggunakan sampel terbatas. Berikut adalah rumus yang berkaitan dengan distribusi sampling rata-rata :

**Tabel 9.1.** Rumus distribusi sampling rata-rata

	Jika $\frac{n}{N} \leq 5\%$	Jika $\frac{n}{N} > 5\%$
Rata rata	$\mu_{\bar{x}} = \mu$	$\mu_{\bar{x}} = \mu$
Standar Deviasi	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}$
Nilai Baku	$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$	$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$

Keterangan :

$n$	= ukuran sampel	$N$	= ukuran populasi
$\bar{x}$	= rata-rata sampel	$\mu$	= rata-rata populasi
$s$	= standar deviasi sampel	$\sigma$	= standar deviasi populasi
$\mu_{\bar{x}}$	= rata-rata distribusi sampling	$\sigma_{\bar{x}}$	= standar deviasi distribusi sampling

$$\sqrt{\frac{N-n}{N-1}} = \text{faktor koreksi}$$

Contoh soal:

Dari 200 orang calon mahasiswa baru perguruan Tinggi ABC, yang mengikuti Ujian akhir, nilai mereka berdistribusi normal

dengan rata-rata 70 dan standar deviasi 8. Berapakah peluang dari 25 calon mahasiswa baru tersebut yang mempunyai rata-rata nilai UTBK 68 dan 80?

Diketahui :  $N = 200$        $\mu = 70$

$n = 25$        $\sigma = 8$

Ditanya:  $p(68 < \bar{x} < 80)$  bila  $n = 25$

Jawab :

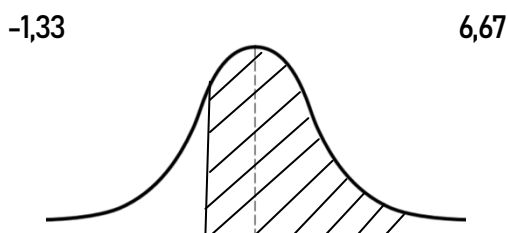
- $\frac{n}{N} = \frac{25}{200} = 0,125$ ; menggunakan faktor koreksi karena  $\frac{n}{N} > 5\%$

- $\mu_{\bar{x}} = \mu = 70$

- $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}} = \frac{8}{\sqrt{25}} \times \sqrt{\frac{200-25}{200-1}} = 1,500418702$

- $z_1 = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{68 - 70}{1,500418702} = -1,33$

- $z_2 = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{80 - 70}{1,500418702} = 6,67$  (lihat tabel z)



Gambar 9.1. Luas daerah menggunakan kurva distribusi z

$$p(68 < \bar{x} < 80) = (\text{luas } z_1 = -1,33) + (\text{luas } z_2 = 6,67) = 0,4082 + 0,500 = 0,9082$$

Jadi, peluang dari 25 orang peserta Ujian akhir yang di observasi ternyata mempunyai rata-rata nilai UTBK 68 dan 80 adalah sebesar 0,9082 atau 90,82%.

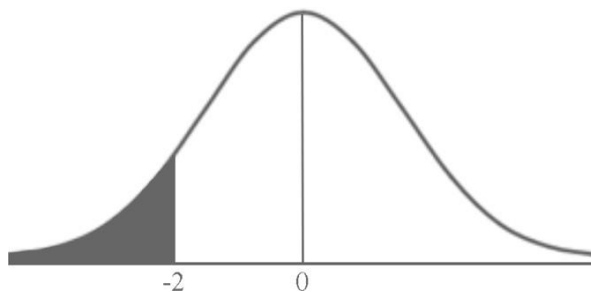
Sebuah manufaktur memproduksi plat besi yang memiliki daya regang rata-rata 500 dan standar deviasi 20 yang penyebarannya mendekati distribusi normal. Sebagai bagian dari pengujian, dipilih secara random 100 plat besi dari 100.000 plat besi yang dihasilkan. Berapa probabilitas didapat rata-rata sampel akan kurang dari 496?

Dik:  $N = 100.000$   
 $n = 100$   
 $\mu = 500$   
 $\sigma = 20$

Dit:  $P(\bar{x} < 496)$

Jawab:

$$\begin{aligned}\mu_{\bar{x}} &= \mu = 500 \\ \frac{n}{N} &= \frac{100}{100.000} = 0,001; \text{ tidak perlu faktor koreksi karena } \frac{n}{N} \leq 5\% \\ \sigma_{\bar{x}} &= \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{100}} = 2 \\ z &= \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{496 - 500}{2} = -2\end{aligned}$$



**Gambar 9.2.** Luas daerah menggunakan kurva distribusi z

$$P(\bar{x} < 496) = \text{luas kiri } 0 - \text{luas } Z = 0,5 - 0,4772 = 0,0228$$

Jadi, probabilitas didapat rata-rata sampel akan kurang dari 496 adalah 2,28%.

## B. Distribusi Sampling Proporsi

Distribusi Sampling proporsi adalah suatu distribusi probabilitas dari yang statistik sampelnya merupakan proporsi sampel.

**Tabel 9.2.** Rumus terkait dengan distribusi sampling proporsi

	Jika $\frac{n}{N} \leq 5\%$	Jika $\frac{n}{N} > 5\%$
Rata - rata	$\mu \frac{x}{n} = \pi$	$\mu \frac{x}{n} = \pi$
Standar Deviasi	$\sigma \frac{x}{n} = \sqrt{\frac{\pi (1 - \pi)}{n}}$	$\sigma \frac{x}{n} = \sqrt{\frac{\pi (1 - \pi)}{n}} \times \sqrt{\frac{N - n}{N - 1}}$
Nilai Baku	$z = \frac{\frac{x}{n} - \mu \frac{x}{n}}{\sigma \frac{x}{n}}$	$z = \frac{\frac{x}{n} - \mu \frac{x}{n}}{\sigma \frac{x}{n}}$

Keterangan :

$n$  = ukuran sampel

$N$  = ukuran populasi

$\bar{x}$  = rata-rata sampel

$\mu$  = rata-rata populasi

$s$  = standar deviasi sampel

$\sigma$  = standar deviasi populasi

$\mu \frac{x}{n}$  = rata-rata distribusi sampling

$\sqrt{\frac{N-n}{N-1}}$  = faktor koreksi

$\sigma \frac{x}{n}$  = standar deviasi distribusi

sampling

Contoh soal:

Dalam suatu tes ditemukan bahwa 25% dari peserta lulus. Diambil 180 peserta sebagai sampel acak dari semua peserta yang mengikuti tes. Berapa probabilitas pelanggan yang lulus kurang dari 15%?

Diketahui:  $n = 180$

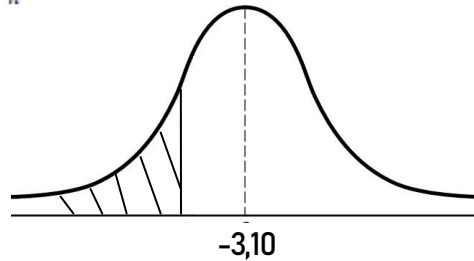
$\pi(\text{lulus}) = 0,2$

Ditanya:  $p\left(\frac{x}{n} < 15\%\right)$  ?

Jawab :

- $\mu \frac{x}{n} = \pi = 0,2$

- $\sigma_{\frac{x}{n}} = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0,25(1-0,25)}{180}} = 0,0322$
- $z = \frac{\frac{x}{n} - \mu}{\sigma_{\frac{x}{n}}} = \frac{0,15 - 0,25}{0,0322} = -3,1055 \text{ (lihat tabel } z \text{)}$



**Gambar 9.3.** Luas daerah menggunakan kurva distribusi z

$$p\left(\frac{x}{n} < 15\%\right) = (\text{luas sisi kiri } z) - (\text{luas } z = -3,10) = 0,5000 - 0,001 = 0,499$$

Jadi, dari 180 peserta yang dijadikan sampel, probabilitas peserta lulus kurang dari 25% adalah sebesar 0,499 atau 49,9%.

### C. Kekeliruan Baku

Kekeliruan baku atau *Standard Error* (SE) adalah ukuran yang menunjukkan seberapa besar variabilitas suatu statistik sampel (seperti rata-rata atau proporsi) dibandingkan dengan parameter populasi sebenarnya. Kekeliruan baku digunakan dalam statistika inferensial untuk mengukur ketidakpastian atau keakuratan dari suatu estimasi sampel terhadap populasi.

Nilai minimum ukuran sampel  $n$  bisa ditentukan apabila ukuran proporsi tiap sampel yang diharapkan terjadi diketahui. Jika dikehendaki selisih rata-rata setiap dua sampel tidak lebih dari  $d$ , maka berlaku:



**Tabel 9.3.** Rumus yang berkaitan dengan kekeliruan baku

Rata -rata	Proporsi
$\frac{\sigma}{\sqrt{n}} \leq d$	$\sqrt{\frac{\pi(1-\pi)}{n}} = \leq d$

**Contoh soal:**

Dari suatu proses produksi seragam sekolah SMA, ternyata 80% hasilnya baik dan 20% sisanya kurang baik. Selama proses berlangsung, diambil sebuah sampel acak terdiri dari 100 baju. Berapa sampel paling sedikit yang dibutuhkan agar jika kita mengambil sampel dengan ukuran tersebut, persentase kerusakannya diharapkan tidak lebih dari 2%?

Diketahui:  $\pi = 20\% = 0,2$

$d = 2\% = 0,02$

Ditanya :  $n$ ?

Jawab :

$$\sqrt{\frac{\pi(1-\pi)}{n}} = \leq d$$

$$\sqrt{\frac{0,2(1-0,2)}{n}} = \leq 0,02$$

$$n \geq 400$$

Jadi, ukuran sampel paling sedikit yang dibutuhkan ketika kita mengambil ukuran sampel yang cukup agar persentase kerusakannya diharapkan akan tidak lebih dari 2% adalah 400 baju.

**D. Distribusi Sampling Selisih Rata-rata**

Distribusi sampling selisih rata-rata adalah distribusi probabilitas yang dapat terjadi dari selisih rata-rata dua sampel yang berbeda berdasarkan pada dua sampel tertentu dari ukuran parameter dua populasinya

Apabila  $n_1, n_2 > 30$  atau ukuran sampel cukup besar, maka distribusi sampling selisih rata-rata sangat mendekati distribusi normal, untuk mengubah ke dalam bentuk normal standar maka diperlukan rumus:

**Tabel 9.4.** Rumus Distribusi sampling selisih rata rata

Rata - rata	$\mu\bar{x}_1 - \bar{x}_2 = \mu_1 - \mu_2$
Standar Deviasi	$\sigma\bar{x}_1 - \bar{x}_2 = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
Nilai Baku	$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma\bar{x}_1 - \bar{x}_2}$

Jika  $\sigma_1^2$  dan  $\sigma_2^2$  tidak diketahui, dapat menggunakan standar deviasi dari sampel.

#### E Distribusi Sampling Selisih Proporsi

Distribusi sampling selisih proporsi adalah distribusi probabilitas yang dapat terjadi dari selisih proporsi dua sampel yang berbeda berdasarkan pada dua sampel tertentu dari ukuran parameter dua populasinya.

**Tabel 9.5.** Rumus Distribusi sampling selisih proporsi

Rata - rata	$\mu \frac{x_1}{n_1} - \frac{x_2}{n_2} = \pi_1 - \pi_2$
Standar Deviasi	$\sigma \frac{x_1}{n_1} - \frac{x_2}{n_2} = \sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}$
Nilai Baku	$z = \frac{\left(\frac{x_1}{n_1} - \frac{x_2}{n_2}\right) - \mu \frac{x_1}{n_1} - \frac{x_2}{n_2}}{\sigma \frac{x_1}{n_1} - \frac{x_2}{n_2}}$

## DAFTAR PUSTAKA

- Biantoro, Agung Wahyudi dan Muh. Kholil. 2020. *Statistika Penelitian : Analisis Manual Dan IBM SPSS*. El Markazi - Waru Sejahtera Press. ISBN : 978-623-6865-73-6
- Elvania Surya Dewi. 2025. *Distribusi sampling*. Diakses tanggal 11 Maret 2025 pukul 14.45 WIB. Sumber : <https://id.scribd.com/document/543561154/Distribusi-Sampling>
- Sugiyono, 2021. *Statistik Penelitian*. Penerbit Alfabeta. Bandung

# **BAB 10**

## **PENGUJIAN HIPOTESIS**

**Oleh Yani Quarta Mondiana**

### **10.1 Pendahuluan**

Dalam kehidupan sehari-hari, sering kali kita dihadapkan pada situasi untuk membuat keputusan berdasarkan informasi yang tersedia. Misalnya, seorang pengusaha ingin mengetahui apakah strategi pemasaran baru yang diterapkan benar-benar meningkatkan penjualan dibandingkan dengan strategi sebelumnya. Seorang dokter mungkin ingin memastikan apakah obat baru yang dikembangkan lebih efektif dibandingkan dengan obat yang telah lama digunakan. Seorang peneliti dapat memiliki hipotesis bahwa pupuk organik tertentu dapat meningkatkan hasil panen lebih baik daripada pupuk kimia. Semua situasi ini memiliki kesamaan: adanya suatu dugaan atau pernyataan yang perlu diuji kebenarannya sebelum dapat dijadikan dasar dalam pengambilan keputusan yang lebih luas.

Dalam statistika, proses menguji suatu pernyataan atau disebut sebagai uji hipotesis. Uji hipotesis adalah salah satu konsep fundamental dalam statistik inferensia yang memungkinkan kita untuk menarik kesimpulan tentang suatu populasi berdasarkan data sampel yang diperoleh. Dengan menggunakan metode ini, kita dapat menentukan apakah suatu perbedaan atau hubungan yang diamati dalam data terjadi secara kebetulan atau memiliki makna statistik yang dapat digeneralisasi.

### **10.2 Konsep Teoritis**

Uji hipotesis merupakan salah satu metode dalam statistik inferensia yang digunakan untuk menarik kesimpulan tentang suatu populasi berdasarkan data sampel. Dalam prosesnya, uji hipotesis melibatkan perumusan dugaan awal yang disebut hipotesis nol ( $H_0$ ) dan dugaan alternatif yang disebut hipotesis alternatif ( $H_1$ ). Hipotesis nol biasanya menyatakan bahwa tidak ada perbedaan atau efek yang

signifikan dalam suatu fenomena, sedangkan hipotesis alternatif menyatakan adanya perbedaan atau efek tertentu yang ingin dibuktikan.

Teori probabilitas terkait erat dengan konsep uji hipotesis, terutama dalam hal menentukan tingkat kesalahan yang dapat diterima dalam pengambilan keputusan. Terdapat 2 jenis kesalahan dalam pengujian hipotesis yaitu kesalahan tipe I—gagal menolak hipotesis nol meskipun benar—dan kesalahan tipe II—adalah dua jenis kesalahan yang dapat terjadi dalam uji hipotesis. Tingkat signifikansi ( $\alpha$ ), biasanya ditetapkan sebesar 5% atau 1%, tergantung pada tingkat kepastian yang diinginkan dalam analisis, digunakan untuk mengontrol kemungkinan terjadinya kesalahan tersebut.

Dalam perhitungan, nilai  $\alpha$  dapat ditentukan sebelumnya; besaran tersebut juga disebut sebagai ukuran daerah kritis yaitu luas daerah penolakan  $H_0$ . Sedangkan nilai  $\beta$  hanya bisa dihitung jika nilai hipotesis alternatif sangat spesifik. Dalam pengujian hipotesis, kita lebih sering berhubungan dengan nilai  $\alpha$  dengan asumsi bahwa nilai  $\alpha$  yang kecil juga mencerminkan nilai  $\beta$  yang kecil. Peluang timbulnya  $\beta$  sulit ditentukan karena penyebaran hipotesis tandingan biasanya tidak diketahui. Jika hal ini terjadi, maka penerimaan  $H_0$  sebagai sesuatu yang dianggap benar, mengandung kesalahan yang tidak diketahui berapa besar peluangnya. Oleh sebab itu akan lebih baik jika dikatakan bahwa : data tidak mendukung untuk menolak  $H_0$ , dari pada dinyatakan dalam kalimat : menerima kebenaran  $H_0$ .

Nlai  $(1-\beta)$  disebut sebagai kuasa pengujian, yang merupakan peluang menerima hipotesis tandingan, apabila hipotesis tersebut sebetulnya benar. Antara kebenaran hipotesis yang disusun dan kemungkinan tindakan yang diambil berdasarkan hasil pengujian disajikan pada tabel berikut :

Keputusan yang diambil	Keadaan yang sebenarnya	
	$H_0$ benar	$H_1$ benar
Menerima $H_0$	$1 - \alpha$	$\beta$
Menolak $H_0$ / Menerima $H_1$	$\alpha$	$1 - \beta$

Untuk dapat mengendalikan peluang terjadinya salah jenis pertama, maka hipotesis nol dispesifikasikan sebagai nilai tunggal. Kesimpulan yang diambil apabila  $H_0$  ditolak, sangat tergantung pada penggunaan uji satu sisi atau dua sisi pada hipotesis alternatif.

Uji satu sisi (one-tailed test) dimaksudkan bahwa pernyataan pada  $H_1$  bisa lebih besar (uji satu sisi sebelah kanan) atau lebih kecil dari pada pernyataan pada  $H_0$  (uji satu sisi sebelah kiri). Pada uji satu sisi sebelah kanan, daerah penolakan  $H_0$  berada di wilayah bagian kanan dari suatu sebaran dengan luasan sebesar  $\alpha$ . Sebaliknya pada uji satu sisi sebelah kiri maka wilayah penolakan  $H_0$  berada di sebelah kiri dengan luasan sebesar  $\alpha$ . Pengajuan  $H_0$  dan  $H_1$  dalam uji satu arah (=satu sisi) adalah sebagai berikut:

$H_0$  : ditulis menggunakan tanda =

$H_1$  : ditulis menggunakan tanda lebih besar ( $>$ ) atau lebih kecil ( $<$ )

Pada kasus tertentu dapat dilakukan uji dua sisi yaitu apabila pernyataan pada  $H_1$  menunjukkan perbedaan dengan pernyataan pada  $H_0$ . Pada uji dua sisi (two-tailed test), wilayah penolakan  $H_0$  terdapat pada kedua sisi sebaran, masing-masing dengan luasan sebesar  $\alpha/2$ . Pengajuan  $H_0$  dan  $H_1$  dalam uji dua arah adalah sebagai berikut :

$H_0$  : ditulis dalam bentuk persamaan (menggunakan tanda =)

$H_1$  : ditulis dengan menggunakan tanda  $\neq$

Misalkan pada kasus berikut : ingin menguji rata-rata nilai statistik mahasiswa laki laki dan perempuan; maka pada pengujian hipotesis dapat ditulis secara singkat misalnya :

$H_0$  :  $\mu_{\text{laki laki}} = \mu_{\text{perempuan}}$

$H_1$  :  $\mu_{\text{laki laki}} \neq \mu_{\text{perempuan}}$

Wilayah penolakan  $H_0$  disebut juga sebagai daerah atau wilayah kritis; sedangkan wilayah sebesar  $(1-\alpha)$  adalah daerah penerimaan  $H_0$ . Batas antara wilayah  $H_0$  dan  $H_1$  adalah nilai kritis yang dapat dilihat pada tabel statistika. Penentuan digunakannya uji satu sisi atau dua sisi tergantung pada keputusan yang diambil apabila  $H_0$  ditolak.

Suatu hasil pengujian hipotesis dikatakan nyata apabila hipotesis nol ditolak pada taraf nyata 5% dan dikatakan sangat nyata

jika hipotesis nol ditolak pada taraf nyata 1%. Apabila hipotesis nol diterima, maka dikatakan bahwa uji tidak signifikan secara statistika.

Salah satu cara pengujian hipotesis yaitu melalui statistik uji (yang kemudian masih harus dibandingkan dengan nilai kritis).

### 10.3 Kriteria pengujian Hipotesis Melalui Statistik Uji

Kesimpulan yang diambil pada pengujian hipotesis yang didasarkan pada statistik uji mudah untuk dipelajari serta dipahami. Prosedur menguji hipotesis dengan statistik uji adalah sebagai berikut:

1. Mendefinisikan formula terkait dengan parameter yang hendak diuji. Artinya kita menentukan  $H_0$ .
2. Menentukan pernyataan yang mungkin akan diterima jika ditolak, memilih uji dua sisi atau uji satu sisi. Langkah ini adalah menentukan  $H_1$ .
3. Memilih taraf nyata  $\alpha$
4. Menentukan uji statistik berdasarkan sebaran yang sesuai
5. Menentukan nilai kritis yang merupakan batas antara wilayah  $H_0$  dan wilayah  $H_1$
6. Menghitung statistik uji berdasarkan data terkumpul
7. Memeriksa keberadaan statistik uji apakah jatuh di wilayah  $H_0$  ataukah di wilayah  $H_1$ .
8. Membuat keputusan secara statistika (menerima atau menolak  $H_0$ ) dengan memperhatikan wilayah kritis. Jika statistik uji berada di daerah penerimaan maka diputuskan untuk menerima  $H_0$ , artinya belum cukup bukti untuk menerima hipotesis alternatif karena kemungkinan ukuran contoh kurang besar. Sebaliknya jika statistik uji berada di wilayah kritis maka diputuskan bahwa hipotesis awal ditolak pada taraf nyata yang telah ditentukan.
9. Membuat kesimpulan yaitu interpretasi atau ekspresi dari keputusan yang diambil, berdasarkan permasalahannya.

### 10.4 Uji Hipotesis Terhadap Rata-rata Populasi

Dalam pengujian hipotesis terhadap rata-rata populasi yang sama dengan nilai tertentu, maupun menguji kesamaan rata-rata dari

dua populasi, pengetahuan tentang ragam populasi penting untuk menentukan prosedur perhitungan mana yang hendak digunakan. Jika contoh acak diambil dari suatu populasi yang menyebar normal dengan rata-rata  $\mu$  dan simpangan baku  $\sigma$ , maka  $\bar{X} \approx N(\mu, \frac{\sigma^2}{n})$ ; dengan demikian apabila populasi berdistribusi normal dan ragam populasi diketahui maka besaran  $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$  akan mengikuti distribusi  $N(0,1)$  sehingga sebaran normal baku  $Z$  dapat digunakan untuk menentukan titik kritis. Namun jika ragam populasi tidak diketahui dan ukuran contoh relatif kecil maka ragam populasi diduga dengan ragam contoh dan

$$t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}} \text{ mengikuti sebaran } t_{(db)}$$

maka titik kritis dapat ditentukan dengan melihat tabel distribusi  $t$ . Distribusi ini bentuknya setangkup juga tetapi tinggi rendahnya puncak kurva sangat tergantung pada derajat bebas yang nilainya adalah  $(n-1)$ . Semakin besar ukuran contoh, sebaran  $t$  akan semakin mendekati sebaran normal baku. Kriteria pengujian adalah membandingkan nilai statistik uji (bisa  $z$  hitung atau  $t$  hitung) dengan titik kritis (nilai  $z$  tabel atau  $t$  tabel); apabila statistik uji berada di daerah penerimaan  $H_0$  maka keputusannya adalah menerima  $H_0$  (ingat bahwa titik kritis adalah nilai yang menjadi batas daerah penerimaan dan penolakan hipotesis). Atau menggunakan kriteria uji : jika  $t\text{-hitung} \leq t\text{-tabel}$  maka  $H_0$  diterima dan sebaliknya. Rumus untuk pengujian hipotesis terhadap satu rata-rata populasi dapat dirangkum seperti disajikan pada tabel 101.



**Tabel 10.1.** Kriteria pengujian hipotesis terhadap satu rata-rata populasi

$H_0$	Statistik uji	$H_1$	Wlayah Kritis
$\mu = \mu_0$  contoh besar $n \geq 30$ atau $\sigma$ diketahui	$z_{hitung} = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$ $\sigma$ dapat diganti dengan $s$	$\mu < \mu_0$ $\mu > \mu_0$ $\mu \neq \mu_0$	$z < -z_\alpha$ $z > z_\alpha$ $z < -z_{\alpha/2}$ dan $z > z_{\alpha/2}$
$\mu = \mu_0$  contoh kecil $n < 30$ dan $\sigma$ tidak diketahui	$t_{hitung} = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$	$\mu < \mu_0$ $\mu > \mu_0$ $\mu \neq \mu_0$	$t < -t_{(\alpha; db)}$ $t > t_{(\alpha; db)}$ $t < -t_{(\alpha; db)}$ dan $t > t_{(\alpha/2; db)}$ $db = n-1$

Contoh :

Seorang peneliti ingin mengetahui apakah produktivitas padi di suatu daerah lebih rendah dari standar nasional, yaitu 6 ton per hektar. Ia mengambil sampel acak sebanyak 20 petani dan mencatat hasil panen mereka. Dari sampel tersebut, diperoleh rata-rata produktivitas sebesar 5,7 ton per hektar dengan simpangan baku 0,8 ton. Pada taraf nyata 5%, apa kesimpulan anda ?

Untuk menjawab soal ini, baca soalnya dengan benar tentang simpangan baku tersebut berasal dari perhitungan sampel ( $s$ ) ataukah dari populasi yang sudah diketahui ( $\sigma$ ). Dari kalimat tersebut maka nilai 0.8 adalah simpangan baku sampel karena merujuk pada kalimat yang mendahuluinya.

$$H_0 : \mu = 6$$

$$H_1 : \mu < 6$$

$$\alpha = 0.05$$

$$t_{hitung} = \frac{\bar{x} - \mu_0}{s / \sqrt{n}} = \frac{5.7 - 6}{0.8 / \sqrt{20}} = -1.68$$

$$\text{Titik kritis : } t_{(0.025, 19)} = -1.729$$

Kriteria pengambilan keputusan melalui statistik uji

Karena  $t_{hitung} < \text{nilai kritis}$  artinya titik tersebut berada di daerah kekuasaan  $H_0$  maka keputusannya menerima  $H_0$ , artinya Dengan tingkat signifikansi 5%, tidak ada cukup bukti untuk menyatakan bahwa produktivitas padi di daerah tersebut lebih rendah dari standar nasional.

## 10.5 Uji Hipotesis terhadap ragam populasi

Uji terhadap ragam populasi dimaksudkan membandingkan keragaman populasi dengan nilai tertentu. Apabila  $s^2$  adalah ragam sampel acak berukuran  $n$  yang ditarik dari populasi yang berdistribusi normal, maka statistik

$\frac{(n-1)s^2}{\sigma^2} \approx \chi^2_{n-1}$ . Sehingga untuk menguji ragam populasi digunakan sebaran khi-kuadrat  $\chi^2_{n-1}$ .

Statistik uji serta wilayah kritis dapat dilihat pada tabel berikut :

**Tabel 10.2.** Statistik Uji dan Wilayah Kritis Uji Ragam Populasi

$H_0$	Statistik uji	$H_1$	Daerah Kritis
$\sigma^2 = \sigma_0^2$	$\chi^2_{hitung} = \frac{(n-1)s^2}{\sigma_0^2}$	$\sigma^2 < \sigma_0^2$ $\sigma^2 > \sigma_0^2$ $\sigma^2 \neq \sigma_0^2$	$\chi^2 < \chi^2_{1-\alpha}$ $\chi^2 > \chi^2_{\alpha}$ $\chi^2 < \chi^2_{1-\alpha/2}$ $\chi^2 > \chi^2_{\alpha/2}$

dan

Contoh :

Seorang peneliti ingin mengetahui apakah variasi produksi jagung per hektar di suatu daerah berebda dari standar nasional. Menurut data nasional, ragam produksi jagung per hektar seharusnya adalah 0,64 ton<sup>2</sup>. Untuk menguji hal ini, peneliti mengambil sampel 20 petani, dan diperoleh simpangan baku sampel sebesar 0,9 ton per hektar.

$H_0 : \sigma^2 = 0.64$  lawan  $H_1 : \sigma^2 \neq 0.64$

a. Melalui statistik uji :  $\chi^2_{hitung} = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(20-1)0.9^2}{0.64} = 24.05$

Titik kritis  $\chi^2_{0.025,5} = 32.85$  dan  $\chi^2_{0.975,5} = 10.12$

Pada tingkat signifikansi 5%, tidak ada cukup bukti untuk menyatakan bahwa variasi produksi jagung di daerah tersebut berbeda secara signifikan dari standar nasional. Dengan kata lain, ragam produksi jagung masih sesuai dengan standar nasional.

## 10.6 Uji Hipotesis terhadap selisih rata-rata populasi

Apabila kita bekerja dengan dua populasi maka yang dibahas adalah kesamaan atau perbedaan dari rata-rata kedua populasi. Jika diasumsikan setiap populasi mengikuti sebaran normal dengan rata-rata  $\mu_1$  dan ragam  $\sigma_1^2$  untuk populasi 1 serta rata-rata  $\mu_2$  dan ragam  $\sigma_2^2$  untuk populasi 2, dan ukuran sampel yang diambil secara acak dari populasi 1 dan populasi 2 adalah  $n_1$  dan  $n_2$ , maka selisih rata-rata sampel akan berdistribusi normal dan berlaku :

$$(\bar{x}_1 - \bar{x}_2) \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$$

Sehingga : 
$$\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}} \sim N(0,1)$$

Hipotesis yang mendasari pengujian adalah :

$$H_0 : \mu_1 = \mu_2 \text{ atau } H_0 : \mu_1 - \mu_2 = 0$$

Dengan beberapa hipotesis alternatif yaitu :

$$H_1 : \mu_1 \neq \mu_2 \text{ atau } H_1 : \mu_1 - \mu_2 \neq 0 \text{ (uji dua sisi)}$$

$$H_1 : \mu_1 < \mu_2 \text{ atau } H_1 : \mu_1 - \mu_2 < 0 \text{ (uji satu sisi sebelah kiri)}$$

$$H_1 : \mu_1 > \mu_2 \text{ atau } H_1 : \mu_1 - \mu_2 > 0 \text{ (uji satu sisi sebelah kanan)}$$

Terdapat dua kasus yang mungkin terjadi yaitu

1. Ragam kedua populasi diketahui, atau mungkin tidak diketahui tetapi dan ukuran sampel besar
2. Ragam kedua populasi tidak diketahui tetapi diasumsikan sama. Jika hal kedua yang terjadi dan ukuran contoh tidak terlalu besar, maka ragam contoh dapat digabung menjadi ragam gabungan (pooled variance atau disingkat  $s_{gab}$  atau  $s_p^2$ ).

Statistik uji serta daerah kritis dapat dilihat pada tabel 10.5.

Sebagai contoh, perhatikan soal berikut

1. Seorang peneliti ingin membandingkan produktivitas padi antara Wilayah A dan Wilayah B. Ia ingin mengetahui apakah ada perbedaan signifikan dalam rata-rata produksi padi di kedua wilayah.

	Wilayah A	Wilayah B
Rata-rata daya tahan	$\bar{x}_1 = 6.2$	$\bar{x}_2 = 5.8$
Simpangan baku	$s_1^2 = 0.8$	$s_2^2 = 0.7$
Ukuran sampel	$n_1 = 30$	$n_2 = 30$

Dalam hal ini yang ingin diuji adalah : apakah ada perbedaan rata-rata produksi padi di wilayah A dan wilayah B

$$H_0 : \mu_A = \mu_B$$

$$H_1 : \mu_A \neq \mu_B$$

**Tabel 10.3.** Statistik uji dan Daerah Kritis pada Uji hipotesis terhadap Selisih rata-rata 2 populasi.

$H_0$	Nilai Uji Statistik	$H_1$	Daerah Kritis
$ \mu_1 - \mu_2  = d_0$  $n_1 \geq 30$ $n_2 \geq 30$	$z = \frac{ \bar{x}_1 - \bar{x}_2  - d_0}{\sqrt{(\sigma_1^2 / n_1) + (\sigma_2^2 / n_2)}}$  Jika $\sigma_1^2$ dan $\sigma_2^2$ tidak diketahui, diduga dengan $s_1^2$ dan $s_2^2$	$ \mu_1 - \mu_2  < d_0$  $ \mu_1 - \mu_2  > d_0$  $ \mu_1 - \mu_2  \neq d_0$	$z < -z_\alpha$  $z > z_\alpha$  $z < -z_{\alpha/2}$ dan $z > z_{\alpha/2}$

$ \mu_1 - \mu_2  = d_0$  $n_1 < 30$ $n_2 < 30$	$t = \frac{ \bar{x}_1 - \bar{x}_2  - d_0}{\sqrt{s_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$ $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}$	$ \mu_1 - \mu_2  < d_0$ $ \mu_1 - \mu_2  > d_0$ $ \mu_1 - \mu_2  \neq d_0$	$t < -t_\alpha$ $t > t_\alpha$  $t < -t_{(db, \alpha/2)}$ dan $t > t_{(db, \alpha/2)}$  <b>db</b> = $n_1 + n_2 - 2$
---	---	--	---

Statistik uji :

$$z = \frac{|\bar{x}_1 - \bar{x}_2| - d_0}{\sqrt{(s_1^2 / n_1) + (s_2^2 / n_2)}} = \frac{|6.2 - 5.8| - 0}{\sqrt{(0.8^2 / 30) + (0.7^2 / 30)}} = 2.06$$

Derajat bebas  $t_{0.05(30+30-2)} = 2.001$

Keputusan: karena t hitung ada di daerah kritis atau wilayah penolakan  $H_0$  maka  $H_0$  ditolak , artinya ada perbedaan rata-rata produktivitas padi antara wilayah A dan B berbeda. (Catatan : karena ragam populasi tidak diketahui, dan ukuran contoh yang diambil dari masing-masing populasi berukuran besar, maka ragam populasi diduga dengan ragam sampel).

2. Sekarang coba anda bedakan dengan contoh berikut :

	Sampel dari populasi 1	Sampel dari populasi 2
rata-rata kerusakan	$\bar{x} = 20$	$\bar{x}_2 = 12$
ragam	$s_1^2 = 3.9$	$s_2^2 = 0.72$
ukuran sampel	$n_1 = 13$	$n_2 = 12$

Jika kemudian muncul kasus yaitu bahwa dari hasil uji kesamaan ragam ternyata diputuskan bahwa **ragam populasi ternyata**

berbeda, maka besaran  $\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{(s_1^2/n_1) + (s_2^2/n_2)}}$  akan mengikuti distribusi

yang mendekati distribusi t dengan derajat bebas efektif atau derajat bebas kira-kira yang harus dihitung lebih dahulu dengan rumus :

$$db_{efektif} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

Bagaimana mengetahui bahwa ragam kedua populasi sama atau berbeda, terlebih dahulu dilakukan uji terhadap kedua ragam tersebut, dengan hipotesis bahwa ragam dari populasi satu lebih besar daripada ragam populasi kedua. Hipotesis yang mendasari pengujian adalah :

$$H_0 : \sigma_1^2 = \sigma_2^2 \text{ lawan } H_1 : \sigma_1^2 > \sigma_2^2$$

Statistik uji :  $F_{hitung} = \frac{s_{besar}^2}{s_{kecil}^2}$  akan berdistribusi F

pada  $\alpha$  tertentu dengan  $v_1$  (derajat bebas pembilang),  $v_2$  (derajat bebas penyebut). Kriteria pengujian adalah jika  $F_{hitung} \leq F_{tabel}$  maka hipotesis nol diterima dan sebaliknya jika  $F_{hitung} > F_{tabel}$  maka hipotesis nol ditolak. Pada contoh soal di atas diperoleh  $F_{hitung} = \frac{0.64}{0.49} = 1.306$ , sedangkan dari tabel diperoleh  $F_{0.05, 29, 29} = 2.39$  maka diputuskan untuk menolak menerima  $H_0$ , artinya perbedaan ragam antara kedua populasi tidak signifikan.

## DAFTAR PUSTAKA

- Shayid, M.A. 2013. Applied Statistics. [www.Bookbon.com](http://www.Bookbon.com)
- Mondiana, Y.Q & Wardhani, S. 2021. Statistika Kehutanan. Malang :  
Tunggal Mandiri Publishing
- Wardhani, N.W.S. 2010. Dasar Dasar Statistika Untuk Ilmu Pertanian.  
Malang : Tunggal Mandiri Publishing.
- Walpole, R.E 1990. Pengantar Statistika. Edisi 3. Jakarta : Gramedia

# BAB 11

## ANALISIS REGRESI DAN KORELASI

Oleh Sawarni Hasibuan

### 11.1 Pendahuluan

Analisis regresi dan korelasi adalah dua teknik statistik yang umum diterapkan dalam berbagai disiplin penelitian, termasuk ekonomi, teknik, ilmu sosial, dan sains. Regresi digunakan untuk memahami hubungan antara satu atau lebih variabel independen dengan variabel dependen, sementara korelasi digunakan untuk mengukur sejauh mana dua variabel berkaitan satu sama lain. Dengan teknik ini, peneliti dapat mengidentifikasi pola hubungan, memprediksi nilai masa depan, dan membuat keputusan berbasis data.

Dalam era big data dan kecerdasan buatan, analisis regresi dan korelasi semakin penting dalam proses pengambilan keputusan. Berbagai industri, seperti manufaktur, kesehatan, dan keuangan, memanfaatkan teknik ini untuk mengoptimalkan kinerja dan efisiensi operasional. Pemahaman yang baik tentang analisis regresi dan korelasi memungkinkan para profesional menginterpretasikan data dengan lebih akurat serta membuat strategi yang lebih tepat berdasarkan hasil analisis statistik.

### 11.2 Teori Dasar Analisis Regresi

Analisis regresi adalah metode statistik yang digunakan untuk memahami hubungan antara variabel dependen (tergantung) dan satu atau lebih variabel independen (bebas). Teknik ini memungkinkan peneliti untuk mengembangkan model prediksi yang dapat digunakan dalam berbagai bidang, seperti ekonomi, sains, dan rekayasa.

Terdapat beberapa jenis analisis regresi yang umum digunakan (Montgomery et al, 2021), diantaranya adalah sebagai berikut:

1. **Regresi Linear Sederhana:** Model regresi yang hanya melibatkan satu variabel independen untuk memprediksi variabel dependen.



2. **Regresi Linear Berganda:** Model regresi yang melibatkan lebih dari satu variabel independen.
3. **Regresi Non-Linear:** Model regresi di mana hubungan antara variabel tidak dapat direpresentasikan dalam bentuk garis lurus.
4. **Regresi Logistik:** Model regresi yang digunakan ketika variabel dependen bersifat kategorikal, seperti klasifikasi biner.

Dalam penerapan analisis regresi, terdapat beberapa asumsi yang harus dipenuhi agar hasil analisis valid dan dapat diinterpretasikan dengan benar:

1. **Linearitas:** Hubungan antara variabel independen dan dependen harus bersifat linear.
2. **Independensi:** Observasi dalam dataset harus independen satu sama lain.
3. **Homoskedastisitas:** Varians dari residual harus konstan di seluruh rentang prediktor.
4. **Normalitas:** Residual harus terdistribusi secara normal.
5. **Tidak ada multikolinearitas:** Variabel independen tidak boleh memiliki korelasi yang tinggi satu sama lain.

### 11.2.1 Metode Estimasi Parameter

Model regresi digunakan untuk membuat taksiran mengenai variabel dependen disebut persamaan regresi estimasi, yaitu suatu formula matematis yang menunjukkan hubungan keterkaitan antara satu atau beberapa variabel yang nilainya sudah diketahui dengan suatu variabel yang nilainya belum diketahui.

Dalam menggunakan model regresi perlu diyakini terlebih dahulu bahwa variabel yang digunakan dalam membangun model memiliki keterkaitan secara teoritis atau dapat diestimasi sebelumnya. Hal ini dikarenakan hubungan antar variabel dalam model regresi merupakan hubungan sebab akibat (*causal relationship*). Variabel di dalam model regresi dibedakan menjadi dua yaitu variabel bebas (*independen variabel*) dan variabel terikat (*dependen variabel*). Variabel bebas adalah variabel yang nilainya dapat mempengaruhi variabel lain, dan variabel terikat adalah variabel yang nilainya dipengaruhi oleh variabel lain.

Regresi linier sederhana merupakan pemodelan regresi yang digunakan untuk menggambarkan hubungan antara satu variabel bebas dengan satu variabel terikat. Bentuk umum persamaan regresi linier sederhana adalah sebagai berikut:

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (1)$$

Di mana:

$Y$  adalah variabel dependen (terikat/respons),

$X$  adalah variabel independen (bebas),

$\beta_0$  adalah intersep (titik potong sumbu  $Y$  jika  $X=0$ )

$\beta_1$  adalah kemiringan/gradient

Persamaan regresi linier sederhana paling baik adalah dengan metode *ordinary least square* (OLS) dengan menggunakan persamaan berikut untuk mencari intersep dan kemiringan persamaan  $Y = \alpha + \beta X$ :

$$\beta = \frac{\sum_1^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_1^n (x_i - \bar{x})^2} \quad (2)$$

$$\beta = r_{xy} \frac{s_y}{s_x} \quad (3)$$

$$\alpha = \bar{y} - \beta \bar{x} \quad (4)$$

Dimana  $\bar{x}$  adalah rata-rata, dan  $s_x$  adalah simpangan baku dari set data kelompok pertama, di mana setiap titik data direpresentasikan oleh  $x_i$ . Demikian pula,  $\bar{y}$  adalah rata-rata, dan  $s_y$  adalah simpangan baku dari set data kelompok kedua;  $n$  adalah jumlah titik data dalam set data.

Model regresi linier berganda merupakan pengembangan dari model regresi linier sederhana. Pada regresi linier berganda jumlah variabel bebasnya lebih dari satu dan satu variabel terikat. Dengan bertambahnya variabel variabel bebas maka bentuk umum dari persamaan regresi linier berganda yang mencakup  $\geq 2$  variabel bebas adalah sebagai berikut:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \quad (5)$$

Apabila menggunakan data sampel, maka model estimasi dari Persamaan (2) dapat dituliskan sebagai berikut:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + e \quad (6)$$

Berdasarkan Persamaan (3) maka bentuk matriks dari model regresi linier berganda dapat dituliskan sebagai berikut:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 1 & X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{kn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

atau dapat ditulis dengan  $Y = X\beta + e$ , dengan:

$Y$  adalah vektor pengamatan berukuran  $n \times 1$

$X$  adalah matriks variabel bebas ukuran  $n \times k$

$\beta$  adalah vektor parameter yang akan ditaksir berukuran  $k \times 1$

$e$  adalah vektor random error berukuran  $n \times 1$

Secara teoritis penggunaan analisis regresi linier berganda akan menghasilkan nilai estimasi parameter yang valid bila terpenuhinya asumsi klasik.

## 11.2.2 Asumsi Klasik

Asumsi klasik adalah persyaratan perhitungan statistik yang harus dipenuhi pada sebuah analisis regresi linear yang berdasarkan pada *ordinary least square* (OLS). Perhitungan analisis yang bukan berdasarkan OLS tidak memerlukan uji asumsi klasik.

Uji asumsi klasik merupakan syarat wajib harus dilakukan sebelum menguji hipotesis. Namun, apabila hasil uji asumsi klasik tidak sesuai dengan harapan atau syarat maka tidak serta merta hasil analisis atau uji hipotesisnya salah atau invalid. Jadi, jika terjadi hasil uji yang tidak sesuai, ada baiknya tetap mempertahankan data yang ada.

Berikut penjelasan asumsi klasik dalam analisis regresi:

### 1. Uji Normalitas

Uji normalitas bertujuan untuk menguji apakah dalam model regresi variabel terikat dan variabel bebas mempunyai distribusi normal atau tidak, nilai residualnya mempunyai distribusi normal atau tidak. Model regresi yang baik adalah memiliki nilai residual normal atau mendekati normal. Uji normalitas yang digunakan dalam penelitian umumnya adalah *Kolmogorov Smirnov* yaitu dengan kriteria jika signifikan *Kolmogorov Smirnov*  $< 0,05$  maka data tidak normal, sebaliknya jika signifikan  $> 0,05$  maka data normal.

Hipotesis statistik yang diuji dinyatakan sebagai berikut.

$H_0$  : Sampel berdistribusi normal

$H_1$  : Sampel tidak berdistribusi normal

Kriteria pengujiannya adalah tolak  $H_0$  jika signifikansi kurang dari  $\alpha = 0,05$

Jika diperoleh nilai signifikansi mendekati 0,05 pada penelitian, mis. 0,049, maka bisa melakukan beberapa hal yaitu: 1) mengeliminir outlier; 2) menambahkan data observasi, atau 3) melakukan transformasi data.

### 2. Uji Multikolinieritas

Uji multikolinieritas bertujuan untuk menguji apakah di dalam model regresi ditemukan adanya korelasi antara variabel bebas. Model regresi yang baik seharusnya tidak terjadi korelasi diantara variabel bebas. Untuk mendeteksi multikolinearitas di dalam regresi dapat dilihat dari nilai *Variance Inflation Factor (VIF)* dan nilai *Tolerance*. Jika  $VIF < 10$  dan *Tolerance*  $> 0,1$  maka tidak terjadi multikolinieritas, tetapi jika  $VIF > 10$  dan *Tolerance*  $< 0,1$  maka terjadi multikolinearitas.

Jadi dapat disimpulkan semakin tinggi hasil nilai uji VIF, maka semakin serius atau berat permasalahan mengenai multikolinearitasnya atau korelasi antar variabel independen pada sebuah penelitian.

### 3. Uji Heteroskedastisitas

Uji heteroskedastisitas bertujuan untuk melihat apakah dalam model regresi terjadi ketidaksamaan varian residual satu pengamatan ke pengamatan lain. Jika residual satu pengamatan ke pengamatan lain tetap, maka disebut homokedastisitas dan jika berbeda disebut heteroskedastisitas. Model regresi yang baik adalah yang homokedastisitas atau yang tidak terjadi heteroskedastisitas.

Untuk mendeteksi adanya heteroskedastisitas dapat menggunakan uji Glejser. Dalam uji ini, apabila hasilnya  $\text{sig.} > 0,05$  maka tidak terdapat gejala heterokedastisitas, model yang baik ialah tidak terjadi heterokedastisitas.

Proses uji heteroskedastisitas lainnya dapat dilakukan dengan metode scatterplot yakni dengan membuat plot nilai prediktif atau nilai  $Z_{\text{PRED}}$  dengan nilai sisa atau nilai  $S_{\text{RESID}}$ . Model regresi berganda yang baik yakni ketika grafik tidak menunjukkan pola tertentu, seperti menyempit dan melebar, berkumpul di tengah, atau melebar dan menyempit.

Jika hasil uji heteroskedastisitas melanggar kondisi homoskedastisitas, maka beberapa langkah yang bisa dilakukan diantaranya adalah:

- a. Melakukan transformasi data menjadi bentuk logaritmik, dan
- b. Semua variabel yang ada dalam penelitian dapat dibagi dengan variabel yang dianggap mengalami gangguan.

### 4. Uji Autokorelasi

Uji autokorelasi merupakan keadaan dimana pada model regresi ada korelasi antara residual pada periode  $t$  dengan residual pada periode sebelumnya ( $t-1$ ). Model regresi yang baik adalah yang tidak adanya autokorelasi.

Autokorelasi sering terjadi pada data *time series* dan dapat juga terjadi pada data *cross section* tetapi jarang (Gozali, 2017). Hal yang dilakukan untuk mendeteksi adanya autokorelasi dalam model regresi linier berganda adalah menggunakan metode Durbin-Watson. Durbin-Watson telah berhasil mengembangkan suatu metode yang digunakan untuk mendeteksi adanya masalah autokorelasi dalam model regresi linier berganda menggunakan

pengujian hipotesis dengan statistik uji yang cukup populer. Hipotesis untuk uji Durbin-Watson adalah sebagai berikut:

$H_0 : \rho = 0$  (tidak terdapat autokorelasi)

$H_1 : \rho \neq 0$  (terdapat autokorelasi)

**Statistik uji:** 
$$d = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n (e_i)^2} \quad (7)$$

Durbin-Watson berhasil menurunkan nilai kritis batas bawah ( $d_L$ ) dan batas atas ( $d_U$ ) sehingga jika nilai  $d$  hitung dari persamaan (4) terletak di luar nilai kritis ini, maka ada atau tidaknya autokorelasi baik positif atau negatif dapat diketahui. Deteksi autokorelasi pada model regresi linier berganda dengan metode Durbin-Watson adalah seperti pada Tabel 11.1 (Turner, 2021).

**Tabel 11.1 . Uji Statistik Durbin-Watson**

$0 < d < d_U$	Menolak $H_0$ ; ada autokorelasi positif
$d_L \leq d \leq d_U$	Daerah keragu-raguan, tidak ada keputusan
$d_U \leq d \leq 4 - d_U$	Menerima $H_0$ , tidak ada autokorelasi positif/negatif
$4 - d_U \leq d \leq 4 - d_L$	Daerah keragu-raguan, tidak ada keputusan
$4 - d_L \leq d \leq 4$	Menolak $H_0$ ; ada autokorelasi negatif

Selain dengan uji Durbin-Watson dapat juga digunakan uji Run Test. Namun, apabila data observasi lebih dari 100 data, maka lebih baik menggunakan uji Lagrange Multiplier.

Beberapa langkah yang bias dilakukan jika dari hasil uji statistic terjadi permasalahan autokorelasi:

- (1) Menambahkan variabel bebas dari variabel tidak bebas sehingga data observasi akan berkurang satu atau
- (2) Mengubah model regresi berganda menjadi persamaan perbedaan umum.

## 5. Uji Linieritas

Uji Linieritas bertujuan untuk menguji apakah spesifikasi model regresi berganda yang digunakan yaitu studi empiris memiliki hubungan linier atau tidak. Uji ini termasuk jarang digunakan dibandingkan dengan keempat uji lainnya.

Uji linearitas pada model regresi berganda dapat menggunakan uji Durbin-Watson, tes pengali Lagrange, dan tes Ramsey. Selain itu cara untuk mendeteksi apakah spesifikasi model linier dapat dengan memperhatikan uji F.

### 11.2.3 Uji Hipotesis

Pengujian hipotesis dimaksudkan untuk melihat apakah suatu hipotesis yang diajukan ditolak atau dapat diterima. Hipotesis merupakan asumsi atau pernyataan yang mungkin benar atau salah mengenai suatu populasi. Dengan mengamati seluruh populasi, maka suatu hipotesis akan dapat diketahui apakah suatu penelitian itu benar atau salah. Untuk keperluan praktis, pengambilan sampel secara acak dari populasi akan sangat membantu.

Dalam pengujian hipotesis terdapat asumsi istilah hipotesis nol. Hipotesis nol merupakan hipotesis yang akan diuji, dinyatakan oleh  $H_0$  dan penolakan  $H_0$  dimaknai dengan penerimaan hipotesis alternatif yang dinyatakan oleh  $H_1$ . Jika telah ditentukan koefisien determinasi ( $R^2$ ), maka selanjutnya dilakukan uji signifikan hipotesis yang diajukan. Uji ini dapat menggunakan uji t, uji F, uji z, atau uji Chi Kuadrat. Dengan uji signifikansi ini dapat diketahui apakah variabel bebas/prediktor/independent (X) berpengaruh secara signifikan terhadap variabel tak bebas/response/dependent (Y). Arti dari signifikan adalah bahwa pengaruh antar variabel berlaku bagi seluruh populasi.

### 11.2.4 Uji Parsial (Uji t)

Pengujian koefisien regresi secara parsial bertujuan mengetahui apakah persamaan model regresi yang terbentuk secara parsial variabel-variabel bebasnya ( $X_1$  dan  $X_2$ ) berpengaruh signifikan terhadap variabel tak bebas (Y).

Hipotesisnya dapat dituliskan sebagai berikut:

1.  $H_0$ : Tidak ada pengaruh  $X_1$ ,  $X_2$  secara parsial terhadap  $Y$
2.  $H_a$ : Ada pengaruh  $X_1$ ,  $X_2$  secara parsial terhadap  $Y$

Kriteria pengambilan keputusan:

1.  $H_0$  diterima bila signifikansi  $> 0,05$  (tidak berpengaruh)
2.  $H_0$  ditolak bila signifikansi  $< 0,05$  (berpengaruh)

### **11.2.5 Uji Simultan (Uji F)**

Uji F digunakan untuk mengetahui apakah variabel bebas secara bersama-sama berpengaruh secara signifikan terhadap variabel tidak bebas.

Hipotesisnya dapat dituliskan sebagai berikut:

1.  $H_0$ : Tidak ada pengaruh  $X_1$ ,  $X_2$  secara bersama-sama terhadap  $Y$
2.  $H_a$ : Ada pengaruh  $X_1$ ,  $X_2$  secara bersama-sama terhadap  $Y$
3. Kriteria pengambilan keputusan:
4.  $H_0$  diterima bila Signifikansi  $> 0,05$  (tidak berpengaruh)
5.  $H_0$  ditolak bila Signifikansi  $< 0,05$  (berpengaruh)

### **11.2.5 Analisis Determinasi (R-square)**

Analisis determinasi merupakan ukuran yang menunjukkan seberapa besar variabel  $X$  memberikan kontribusi terhadap variabel  $Y$ . Analisis ini digunakan untuk mengetahui prosentase sumbangan pengaruh variabel independen secara serentak terhadap variabel dependen.

## **11.3 Teori Dasar Analisis Korelasi**

Analisis korelasi adalah teknik statistik yang digunakan untuk mengukur hubungan antara dua atau lebih variabel (Gogtay & Thatte, 2017). hair et al, 2006). Hubungan ini dapat bersifat positif, negatif, atau tidak memiliki hubungan sama sekali. Korelasi sering digunakan dalam berbagai bidang, termasuk ekonomi, psikologi, teknik, dan ilmu sosial, untuk memahami keterkaitan antar variabel dan membuat prediksi yang lebih akurat.

Korelasi menggambarkan derajat hubungan linier antara dua variabel dengan range nilai antara  $-1$  hingga  $+1$ . Korelasi positif



menunjukkan bahwa ketika satu variabel meningkat, variabel lainnya juga cenderung meningkat. Sebaliknya, korelasi negatif berarti bahwa kenaikan satu variabel diikuti oleh penurunan variabel lainnya. Korelasi nol menunjukkan bahwa tidak ada hubungan linier yang signifikan antara variabel-variabel tersebut.

Terdapat dua kelompok metode untuk analisis korelasi.

1. *Korelasi Parametrik*: korelasi Pearson statis ( $r$ ) digunakan untuk data numeric untuk mengukur hubungan antara dua variabel dan dikenal sebagai uji korelasi parametrik.

Koefisien korelasi Pearson diperoleh menggunakan persamaan berikut ini:

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}} \quad \square \square \square \square$$

Beberapa asumsi dasar pada korelasi *Pearson* adalah variabel memiliki sebaran normal, hubungan linier antara variabel bersifat linier, dan tidak ada pencilan yang ekstrem. Jika asumsi-asumsi ini tidak terpenuhi, korelasi *Spearman* atau *Kendall* lebih disarankan.

Skor koefisien korelasi ( $r$ ) digunakan untuk mengukur kekuatan hubungan antara variabel, berikut interpretasinya:

- $r = 0,00 - 0,19$  Hubungan antar variabel sangat lemah
- $r = 0,20 - 0,39$  Hubungan antar variabel lemah
- $r = 0,40 - 0,59$  Hubungan antar variabel sedang
- $r = 0,60 - 0,79$  Hubungan antar variabel kuat
- $r = 0,80 - 1,00$  Hubungan antar variabel sangat kuat

2. *Korelasi Non-Parametrik*: korelasi *Spearman* ( $\rho$ ) dan Kendall ( $\tau$ ) adalah koefisien korelasi berbasis peringkat, dikenal sebagai korelasi non-parametrik digunakan untuk data kategoris. Perhitungan koefisien korelasi Spearman menggunakan persamaan berikut:

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

Dimana

$\rho$  = koefisien korelasi *Spearman*

$n$  = jumlah pasangan data

$d_i$  = perbedaan peringkat yang diberikan pada nilai dua variabel untuk setiap item data

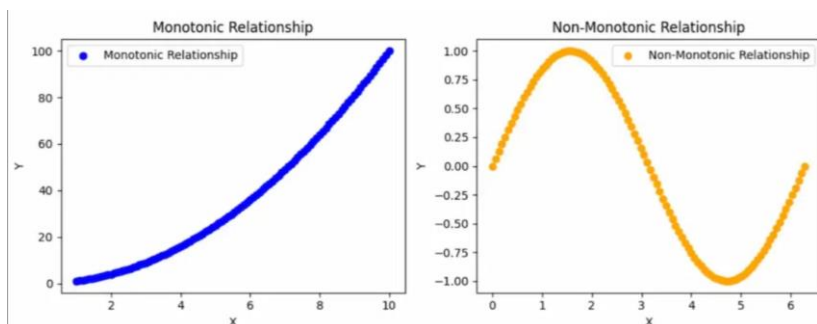
Karakteristik dari korelasi Spearman:

1. Nilai korelasi Spearman ( $r_s$ ) antara -1 (asosiasi negatif) dan 1 (asosiasi positif).
  2. Nilai  $r_s = 0$  berarti tidak ada asosiasi.
  3. Dapat digunakan ketika asosiasi tidak linier.
  4. Dapat diterapkan pada variabel ordinal.
- Korelasi *Kendall* sebagai alternatif korelasi Spearman lebih stabil terhadap data kecil atau dengan banyak peringkat yang sama.

Keunggulan korelasi peringkat Spearman adalah sebagai berikut:

1. Metode Spearman lebih mudah dipahami.
2. Lebih unggul untuk menghitung pengamatan kualitatif seperti kecerdasan orang, penampilan fisik, dll.
3. Cocok ketika deret hanya memberikan urutan preferensi dan bukan nilai aktual variabel.
4. Tahan terhadap outlier yang ada dalam data Dirancang untuk menangkap hubungan monotonik antara variabel.
5. Hubungan monotonik mengukur efek perubahan dalam satu variabel terhadap variabel lain

Hubungan monotonik adalah hubungan matematis antara dua variabel yang arah hubungannya (meningkat atau menurun) tetap konsisten. Sementara hubungan nonmonotonik adalah hubungan matematis antara dua variabel yang arah hubungannya tidak meningkat atau menurun secara konsisten.



**Gambar 11.1.** Model korelasi monotonik dan nonmonotonik.

Kekurangan Korelasi Peringkat Spearman adalah sebagai berikut:

1. Tidak berlaku dalam kasus data yang dikelompokkan.
2. Hanya dapat menangani sejumlah observasi atau item yang terbatas.
3. Mengabaikan hubungan non-monotonik antara variabel, misalnya tidak menangkap jenis hubungan lain, seperti asosiasi kurvilinear atau nonlinier antara variabel.
4. Hanya mempertimbangkan peringkat titik data dan mengabaikan besarnya perbedaan aktual antara nilai variabel.
5. Mengonversi data menjadi peringkat untuk korelasi peringkat Spearman membuang nilai asli variabel dan menggantinya dengan peringkatnya masing-masing. Transformasi ini dapat mengakibatkan hilangnya informasi dalam data, terutama jika variabel data memiliki besaran atau satuan yang bermakna.

Korelasi tidak selalu berarti kausalitas. Meskipun dua variabel mungkin memiliki hubungan yang signifikan, ini tidak berarti bahwa satu variabel menyebabkan perubahan pada variabel lainnya. Untuk menentukan hubungan kausal, diperlukan analisis regresi atau eksperimen terkontrol.

Dalam era big data, analisis korelasi digunakan untuk menemukan pola tersembunyi dalam jumlah data yang besar. Teknik *machine learning* sering mengandalkan korelasi untuk mengidentifikasi variabel-variabel yang berpengaruh dalam suatu model prediktif (Hair et al., 2014)

Korelasi memiliki beberapa keterbatasan, termasuk sensitivitas terhadap penciran dan kemungkinan munculnya korelasi semu akibat variabel yang tidak diamati. Oleh karena itu, korelasi harus selalu dianalisis dengan hati-hati dan dilengkapi dengan metode statistik lainnya.

Walaupun analisis korelasi sangat berguna untuk mengukur hubungan antara variabel, tetapi harus digunakan dengan memahami keterbatasannya. Meskipun korelasi dapat memberikan wawasan tentang keterkaitan antar variabel, analisis tambahan seperti regresi dan eksperimen lebih disarankan untuk mengidentifikasi hubungan sebab-akibat secara lebih akurat.

## 11.4 Studi Kasus

### Kasus 1 Regresi Linier Sederhana

Bagian marketing dapat menggunakan model regresi linier sederhana untuk mengevaluasi pengaruh belanja iklan ( $X_1$ ) sebagai variabel bebas terhadap pendapatan ( $X_2$ ) sebagai variabel tidak bebas/respon. Model regresinya dapat dituliskan sebagai berikut:

$$\text{Pendapatan} = \beta_0 + \beta_1(\text{belanja iklan})$$

Koefisien  $\beta_0$  menjelaskan proyeksi total pendapatan saat belanja iklan adalah nol.

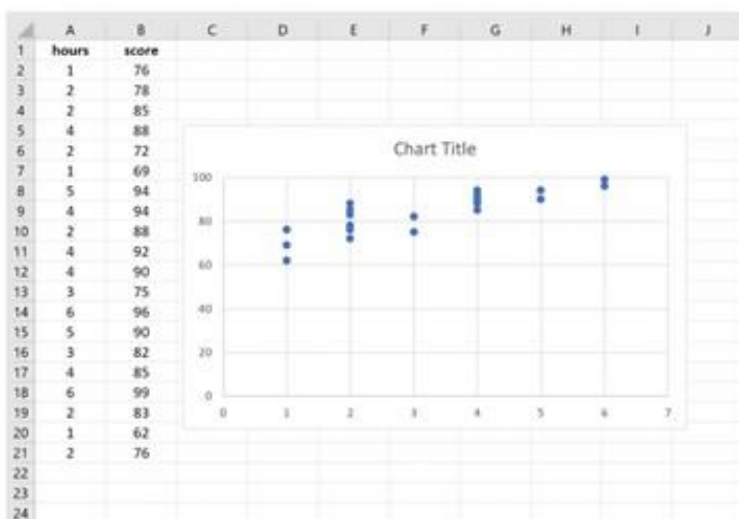
Jika  $\beta_1$  negatif, maka semakin banyak pengeluaran iklan akan menyebabkan pendapatan menurun; jika  $\beta_1$  mendekati nol maka belanja iklan pengaruhnya kecil terhadap peningkatan pendapatan; jika  $\beta_1$  positif maka semakin banyak belanja iklan berpengaruh pada peningkatan pendapatan. Konsekuensinya, perusahaan dapat memutuskan untuk mengurangi atau meningkatkan belanja iklan berdasarkan nilai  $\beta_1$ . Koefisien  $\beta_1$  mewakili perubahan rata-rata total pendapatan jika belanja iklan ditingkatkan sebesar satu unit (mis. per Rp. 1 juta).

### Penyelesaian Regresi Linier Sederhana di Excell

**Langkah 1: Masukkan data.** Masukkan data berikut untuk jumlah jam belajar dan nilai hasil ujian 20 siswa:

	A	B	C	D	E
1	hours	score			
2	1	76			
3	2	78			
4	2	85			
5	4	88			
6	2	72			
7	1	69			
8	5	94			
9	4	94			
10	2	88			
11	4	92			
12	4	90			
13	3	75			
14	6	96			
15	5	90			
16	3	82			
17	4	85			
18	6	99			
19	2	83			
20	1	62			
21	2	76			

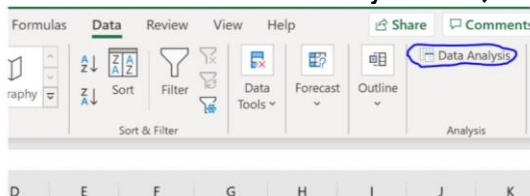
**Langkah 2: Visualisasikan data.** Sebelum melakukan regresi linier sederhana, sebaiknya dibuat scatter diagram data untuk memastikan ada hubungan linier antara jam belajar dan nilai ujian.



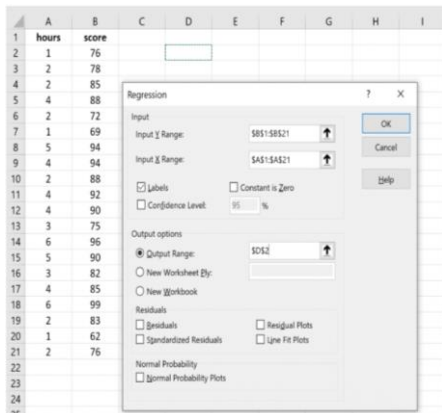
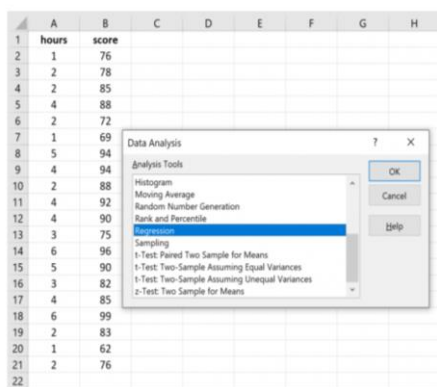
Jumlah jam belajar ditunjukkan pada sumbu-x dan nilai ujian ditunjukkan pada sumbu-y. Kita dapat lihat ada hubungan linier antara kedua variabel – jam belajar yang lebih lama dengan nilai ujian yang lebih tinggi. Untuk mengukur hubungan antara kedua variabel tersebut, dapat dicari dengan regresi linier sederhana.

### ***Langkah 3: Lakukan regresi linier sederhana.***

Selanjutnya klik Analisis Data dan pilih Regresi dan klik OK (add in terlebih dahulu Tool Pack Analysis Data)



Lalu input data Y range untuk variabel terikat; input data X range untuk variabel bebas.



Kemudian centang kotak di samping Label agar Excel mengetahui nama variabel dalam rentang input. Kemudian klik OK. Output berikut akan muncul secara otomatis:

D	E	F	G	H	I	J	K	L
SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.8528							
R Square	0.7273							
Adjusted R Square	0.7121							
Standard Error	5.2805							
Observations	20							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	1	1338.2906	1338.2906	47.9952	0.0000			
Residual	18	501.9094	27.8839					
Total	19	1840.2000						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	67.1617	2.6633	25.2178	0.0000	61.5664	72.7570	61.5664	72.7570
hours	5.2503	0.7578	6.9279	0.0000	3.6581	6.8424	3.6581	6.8424

#### Langkah 4: Menafsirkan output.

1. R-square (koefisien determinasi) = 0,7273 menjelaskan proporsi varians dalam variabel terikat (Y) yang dapat dijelaskan oleh variabel bebas (X); 72,73% variasi nilai ujian dapat dijelaskan oleh jumlah jam belajar. Kesalahan standar: 5,2805. Ini adalah jarak rata-rata jatuhnya nilai yang diamati dari garis regresi. Dalam contoh ini, nilai yang diamati jatuh rata-rata 5,2805 unit dari garis regresi.
2. Nilai statistik F keseluruhan untuk model regresi = 47,9952; statistik F adalah perbandingan MS regresi dengan MS residual (MS regresi/MS residual)
3. Signifikansi F: 0,0000. Ini adalah nilai-p yang dikaitkan dengan statistik F secara keseluruhan. Nilai ini menjelaskan apakah model regresi signifikan secara statistik atau tidak. Dalam kasus ini nilai-p < 0,05, artinya ada hubungan yang signifikan secara statistik antara jam belajar dan nilai ujian yang diterima.
4. Dalam kasus ini persamaan regresi ini diestimasi dengan persamaan berikut:

$$\text{Nilai Ujian} = 67,16 + 5,2503 * (\text{jam})$$

Tafsirnya bahwa untuk setiap jam tambahan jam belajar akan dapat meningkatkan nilai ujian secara rata-rata sebesar 5,2503. Sementara koefisien intersep adalah proyeksi bahwa nilai ujian siswa yang tidak belajar (nol jam) sebesar 67,16.

Kita dapat menggunakan persamaan regresi estimasi ini untuk menghitung nilai ujian yang diharapkan untuk seorang siswa, berdasarkan jumlah jam belajar mereka.

Mis, seorang siswa yang belajar 3 jam diperkirakan memperoleh nilai ujian sebesar 82,91:

$$\text{Nilai ujian} = 67,16 + 5,2503 \cdot (3) = 82,91$$

## Penyelesaian Regresi Linier Berganda di Excell

Misalkan kita ingin mengetahui apakah jumlah jam belajar dan jumlah ujian persiapan yang diambil memengaruhi skor siswa pada ujian masuk PT tertentu.

Dengan menggunakan regresi linier berganda dapat dianalisis apakah penggunaan jam belajar ( $X_1$ ) dan ujian persiapan yang diambil ( $X$ ) sebagai variabel bebas mempengaruhi nilai ujian ( $Y$ ) sebagai variabel terikat/respons.

Langkah-langkah yang diperlukan di Excel adalah sebagai berikut:

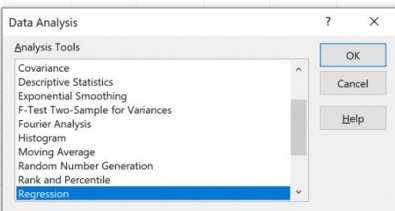
### ***Langkah 1: Masukkan data.***

Menginput data jumlah jam belajar, ujian persiapan yang diambil, dan nilai ujian yang diterima untuk 20 siswa:

### ***Langkah 2: Lakukan regresi linier berganda.***

Di pojok kanan atas di Excel, buka tab Data dan klik Analisis Data. Setelah jendela Analisis Data di klik lalu pilih Regresi dan klik OK.

	A	B	C	D	E	F	G	H	I	J
1	hours	prep_exams	score							
2	1	1	76							
3	2	3	78							
4	2	3	85							
5	4	5	88							
6	2	2	72							
7	1	2	69							
8	5	1	94							
9	4	1	94							
10	2	0	88							
11	4	3	92							
12	4	4	90							
13	3	3	75							
14	6	2	96							
15	5	4	90							
16	3	4	82							
17	4	4	85							
18	6	5	99							
19	2	1	83							
20	1	0	62							
21	2	1	76							





Untuk Input Y Range, isi array nilai untuk variabel terikat. Untuk Input X Range, isi array nilai untuk dua variabel bebas. Lalu centang kotak di samping Label agar Excel mengenali nama variabel dalam rentang input. Untuk Output Range, pilih sel tempat Anda ingin output regresi muncul. Lalu klik OK

	A	B	C	D	E	F	G	H	I	J
1	hours	prep_exams	score							
2	1	1	76							
3	2	3	78							
4	2	3	85							
5	4	5	88							
6	2	2	72							
7	1	2	69							
8	5	1	94							
9	4	1	94							
10	2	0	88							
11	4	3	92							
12	4	4	90							
13	3	3	75							
14	6	2	96							
15	5	4	90							
16	3	4	82							
17	4	4	85							
18	6	5	99							
19	2	1	83							
20	1	0	62							
21	2	1	76							
22										
23										
24										

Output berikut akan muncul secara otomatis:

	D	E	F	G	H	I	J	K
SUMMARY OUTPUT								

### Langkah 3: Menafsirkan output.

Cara menafsirkan angka yang paling relevan dalam output:

1. R-square (koefisien determinasi) = 0,734 menjelaskan proporsi varians dalam variabel terikat (Y) yang dapat dapat dijelaskan oleh

- variabel bebas ( $X_1$  dan  $X_2$ ); 73,4% variasi nilai ujian dapat dijelaskan oleh jumlah jam belajar dan jumlah ujian persiapan yang diambil.
2. Kesalahan standar: 5,366. Ini adalah jarak rata-rata jatuhnya nilai yang diamati dari garis regresi. Dalam contoh ini, nilai yang diamati jatuh rata-rata 5,366 satuan dari garis regresi.
  3. Nilai statistik F keseluruhan untuk model regresi = 23,46; Signifikansi F: 0,00. Dalam kasus ini nilai- $p < 0,05$ , artinya ada hubungan yang signifikan secara statistik antara jam belajar dan ujian persiapan dengan nilai ujian yang diterima.
  4. Nilai- $P$ . Nilai- $p$  parsial menjelaskan apakah setiap variabel bebas signifikan secara statistik atau tidak. Jam belajar signifikan secara statistik ( $p=0,00$ ) sementara ujian persiapan ( $p=0,52$ ) tidak signifikan secara statistik pada  $\alpha=0,05$ . Karena ujian persiapan tidak signifikan secara statistik, maka mungkin diputuskan tidak dimasukkan dalam model.
  5. Koefisien. Koefisien untuk setiap variabel bebas menjelaskan perubahan rata-rata yang diharapkan pada variabel terikat, dengan asumsi variabel bebas lainnya konstan. Misalnya, untuk setiap tambahan jam yang dihabiskan untuk belajar, nilai ujian rata-rata diharapkan meningkat sebesar 5,56, dengan asumsi bahwa ujian persiapan yang diambil tetap konstan. Misalkan untuk 2 siswa A dan B sama-sama mengikuti ujian persiapan yang sama tetapi siswa A belajar satu jam lebih lama, maka siswa A diharapkan memperoleh skor yang 5,56 poin lebih tinggi daripada siswa B.
  6. Koefisien dapat diinterpretasikan sebagai intersep berarti bahwa skor ujian yang diharapkan untuk siswa yang belajar nol jam dan tidak mengikuti ujian persiapan adalah 67,67.
  7. Pada kasus ini persamaan regresi linier berganda diestimasi dengan persamaan berikut:

$$\text{Skor Ujian} = 67,67 + 5,56 * (\text{jam}) - 0,60 * (\text{ujian persiapan})$$

Persamaan regresi estimasi ini dapat digunakan untuk menghitung proyeksi skor ujian untuk seorang siswa, berdasarkan jumlah jam belajar dan ujian persiapan yang diikuti. Misalnya, seorang siswa yang belajar 3 jam dan mengikuti 1 ujian persiapan diharapkan mendapatkan skor 83,75:

$$\text{Nilai ujian} = 67,67 + 5,56 \cdot (3) - 0,60 \cdot (1) = 83,75$$

Perlu diingat bahwa karena ujian persiapan yang diambil tidak signifikan secara statistik ( $p = 0,52$ ), dapat diputuskan untuk mengeliminirnya karena tidak memperbaiki apa-apa pada keseluruhan model. Dalam kasus ini, cukup dengan regresi linier sederhana dengan hanya menggunakan jam belajar sebagai variabel bebas.

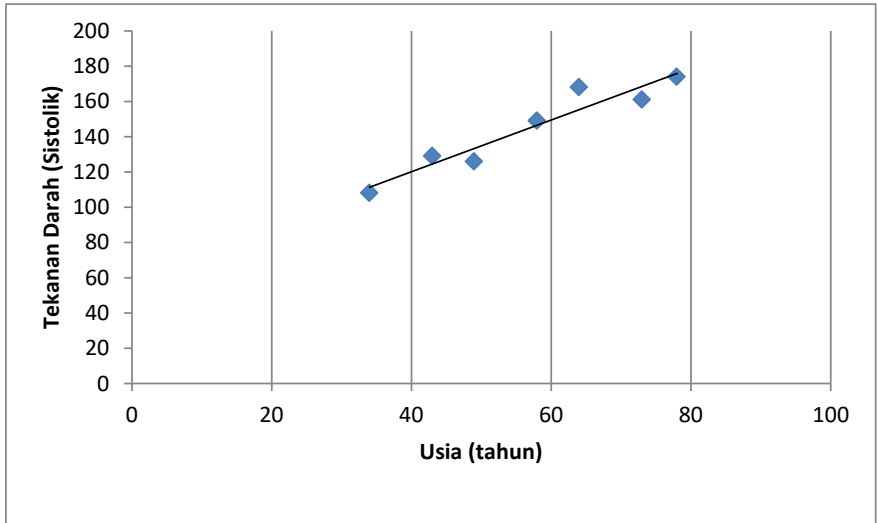
**Kasus 2 Korelasi Pearson**

Jika diambil data data usia, berat badan, dan tekanan darah dari 7 pasien yang berkunjung ke Puskesmas sebagai berikut:

Pasien	Usia (th)	Bobot Badan (kg)	Systolic Pressure
P1	34	45	108
P2	43	44	129
P3	49	56	126
P4	58	57	149
P5	64	65	168
P6	73	63	161
P7	78	55	174

Untuk kasus ini, kita ingin melihat apakah terdapat hubungan linier antara usia dengan tekanan darah sistolik? Taraf nyata yang digunakan adalah 5%.

Eksplorasi diagram pencar (*scatterplot*) variabel usia (tahun) dengan tekanan darah sistolik (mmHg) dapat dilihat pada Gambar 7.3. Berdasarkan *scatterplot*, tampak bahwa sebaran titik-titik mengikuti pola linier dengan kemiringan positif, yang berarti terdapat hubungan yang sejalan antara usia dengan tekanan darah sistolik.



**Gambar 11.2.** Scatter diagram usia vs tekanan darah sistolik.

Dengan demikian, bisa menggunakan koefisien korelasi untuk menentukan apakah hubungan linier kedua variabel tersebut bermakna atau tidak. Apabila pola hubungannya tidak linier, kita tidak tepat menggunakan koefisien korelasi karena nilai  $r$  hanya untuk mengukur kekuatan dan arah hubungan linier antara kedua variabel kuantitatif.

### **Asumsi**

Kedua data berasal dari data kuantitatif. Selanjutnya apakah sebaran kedua variabel berdistribusi normal?

### **Uji Formal**

$H_0$  Data berdistribusi normal

$H_1$  Data tidak berdistribusi normal

**Tabel 11.2. Hasil uji normalitas**

Tests of Normality						
	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Usia	.127	7	.200 <sup>*</sup>	.971	7	.907
Tekanan Darah Sistolik	.171	7	.200 <sup>*</sup>	.939	7	.630

a. Lilliefors Significance Correction

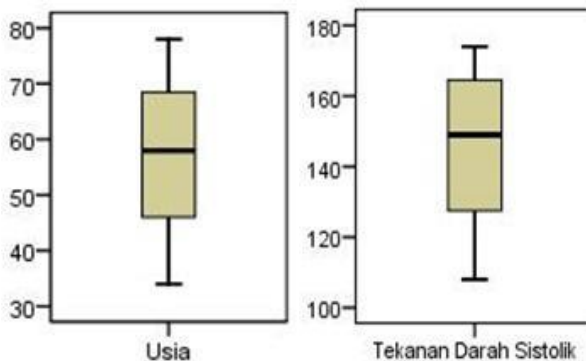
\*. This is a lower bound of the true significance.

## Interpretasi

1. Apabila nilai sig (p-value)  $\leq 0.05$ , maka Tolak  $H_0$  yang berarti data tidak berdistribusi normal
2. Apabila nilai sig (p-value)  $> 0.05$ , maka Terima  $H_0$  yang berarti data berdistribusi normal

Pada kasus di atas, nilai p-value untuk kedua variabel  $> 0.05$ , sehingga kita bisa menyimpulkan bahwa data berdistribusi normal.

Tampak bahwa uji normalitas untuk kedua variabel tersebut memenuhi persyaratan, sebarannya mengikuti distribusi normal, baik dengan menggunakan Uji Kolmogorov-Smirnov ataupun Shapiro-Wilk.



**Gambar 11.3. Box plot sebaran usia dan tekanan darah sistolik**

Secara grafis box plot juga tampak bahwa kedua variabel tersebut berdistribusi normal.

Perhitungan nilai koefisien korelasi ( $r$ ) Pearson menggunakan persamaan (8), dengan memperhatikan hasil perhitungan pada tabel berikut:

Pasien	Usia (X)	Tekanan Darah	X <sup>2</sup>	Y <sup>2</sup>	XY
P1	34	108	1156	11664	3672
P2	43	129	1849	16641	5547
P3	49	126	2401	15876	6174
P4	58	149	3364	22201	8642
P5	64	168	4096	28224	10752
P6	73	161	5329	25921	11753
P7	78	174	6084	30276	13572

$$r_{xy} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}} = \frac{n \sum (X_i Y_i) - (\sum X_i) (\sum Y_i)}{\sqrt{n \sum X_i^2 - (\sum X_i)^2} \sqrt{n \sum Y_i^2 - (\sum Y_i)^2}}$$

$$r_{xy} = \frac{7(60112) - (399 \cdot 1015)}{\sqrt{7 \cdot (24279 - (399)^2) \cdot 7(150803) - (1015)^2}} = \frac{15799}{103.69 \cdot 150.36} = 0,9561$$

Kesimpulan:

Nilai  $|r_{xy}|$  hitung = 0.9561 > nilai tabel kritis  $r$  untuk  $n = 7$  sebesar = 0.754 sehingga bisa disimpulkan terdapat hubungan linier antara usia dengan tekanan darah sistolik.

Output SPSS			
		Height	Weight
Height	Pearson Correlation	1	.956**
	Sig. (2-tailed)		.001
	N	7	7
Weight	Pearson Correlation	.956**	1
	Sig. (2-tailed)	.001	
	N	7	7
**. Correlation is significant at the 0.01 level (2-tailed).			

Korelasi antara usia dengan tekanan darah sistolik:  $r(7) = 0.956$ ;  $p < 0.01$

Nilai koefisien determinasi ( $r^2$ ) = 0.91 menyatakan proporsi keragaman tekanan darah sistolik yang dapat dijelaskan oleh

hubungan linier antara variabel usia dan tekanan darah sistolik. Berdasarkan hasil analisis, kita yakin 95% bahwa sekitar 91% variasi tinggi rendahnya tekanan darah sistolik ditentukan oleh usia seseorang.

**Kasus 3 Korelasi *Spearman***

Sebuah perusahaan sedang melakukan rekrutmen pegawai. Pimpinan perusahaan ingin mengetahui apakah ada hubungan antara nilai ujian tertulis dengan jumlah barang yang dijual oleh masing-masing salesman yang baru direkrut tersebut. Data berikut adalah data mengenai ranking nilai ujian tertulis dan ranking hasil penjualan dari sampel 10 orang salesman yang baru direkrut.

Pegawai	Ranking Nilai Ujian Tertulis	Ranking Jumlah Penjualan
P1	5	3
P2	6	7
P3	8	5
P4	3	1
P5	2	6
P6	7	8
P7	1	2
P8	4	9
P9	10	4
P10	9	10

**Hipotesis:**

- Ho

Tidak ada hubungan antara ranking nilai ujian tertulis salesman dengan rangking jumlah penjualan oleh salesman di tingkat populasi
- H1

Ada hubungan antara ranking nilai ujian tertulis salesman dengan rangking jumlah penjualan oleh salesman di tingkat populasi

Tingkat Signifikansi:  $\alpha=5\%$

Perhitungan nilai korelasi *Spearman* menggunakan persamaan (9) diperoleh sebagai berikut:

$$\rho(\text{rho}) = 1 - \frac{6 \cdot 98}{10(9)^2} = 0,41$$

Pegawai	Ranking Nilai Ujian Tertulis	Ranking Jumlah Penjualan	$d_i$	$d_i^2$
P1	5	3	2	4
P2	6	7	-1	1
P3	8	5	3	9
P4	3	1	2	4
P5	2	6	-4	16
P6	7	8	-1	1
P7	1	2	-1	1
P8	4	9	-5	25
P9	10	4	6	36
P10	9	10	-1	1
			$\Sigma$	98

Dari tabel korelasi Spearman diperoleh nilai  $r_s(0,05;10)$  sebesar 0,648

Karena nilai  $r_{s\text{-hit}} < r_{s\text{-tabel}}$  ( $0,41 < 0,648$ ) maka diputuskan gagal menolak  $H_0$ .

#### Kesimpulan:

Pada tingkat signifikansi 5% ( $\alpha=0,05$ ), belum cukup bukti untuk mengatakan terdapat hubungan antara ranking nilai ujian tertulis dengan ranking jumlah penjualan.

#### Kasus 5 Korelasi Kendall

Jika data yang sama diolah dengan menggunakan korelasi Non Parametrik Kendall dan Spearman diperoleh kesimpulan yang sama, seperti dapat dilihat berikut ini:

Correlations				
			Height	Weight
Kendall's tau_b	Height	Correlation Coefficient	1.000	.810*
		Sig. (2-tailed)	.	.011
		N	7	7
	Weight	Correlation Coefficient	.810*	1.000
		Sig. (2-tailed)	.011	.



		N	7	7
Spearman's rho	Height	Correlation Coefficient	1.000	.929**
		Sig. (2-tailed)	.	.003
		N	7	7
	Weight	Correlation Coefficient	.929**	1.000
		Sig. (2-tailed)	.003	.
		N	7	7
*. Correlation is significant at the 0.05 level (2-tailed).				
**. Correlation is significant at the 0.01 level (2-tailed).				

- Korelasi Kendall's antara usia dengan tekanan darah sistolik:  $r(7) = 0.810$  dengan signifikasi  $p < 0.05$
- Korelasi Spearman's antara usia dengan tekanan darah sistolik:  $r(7) = 0.929$  dengan signifikasi  $p < 0.05$

### 11.5 Penutup

Korelasi dan regresi merupakan *tools* yang ampuh untuk memahami dan memprediksi hubungan antar variabel. Analisis korelasi membantu memahami arah dan kekuatan suatu hubungan, dengan regresi memungkinkan kita membuat model dan prediksi berdasarkan hubungan tersebut.

Tabel berikut merangkum perbedaan utama antara korelasi dan regresi.

Korelasi	Regresi
Korelasi digunakan untuk mengetahui apakah diantara variabel-variabel terdapat hubungan atau tidak.	Regresi digunakan untuk menggambarkan secara numerik bagaimana variabel terikat berubah seiring perubahan variabel bebas
Variabel dapat digunakan secara bergantian	Variabel-variabelnya tidak dapat digunakan secara bergantian
Kekuatan hubungan antara variabel pada korelasi menggunakan nilai numerik	Regresi digunakan untuk menunjukkan seberapa besar dampak perubahan satuan

<b>Korelasi</b>	<b>Regresi</b>
bertanda (+ atau -)	variabel bebas terhadap variabel terikat.
Koefisien Pearson merupakan ukuran korelasi yang terbaik	Metode kuadrat terkecil (ordinary least square) adalah metode terbaik untuk menentukan garis regresi.

## DAFTAR PUSTAKA

- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis*. John Wiley & Sons.
- Pandey, S. (2020). Principles of correlation and regression analysis. *Journal of the practice of cardiovascular sciences*, 6(1), 7-11.
- Shrestha, N. (2020). Detecting multicollinearity in regression analysis. *American Journal of Applied Mathematics and Statistics*, 8(2), 39-42.
- Gogtay, N. J., & Thatte, U. M. (2017). Principles of correlation analysis. *Journal of the Association of Physicians of India*, 65(3), 78-81.

# BAB 12

## ANALISIS VARIANS (ANAVA)

Oleh Daniel Williams Fointuna

### 12.1 Pendahuluan

Pada Bab 10 tentang pengujian hipotesis telah dibahas mengenai uji  $t$  untuk kelompok berbeda dan berpasangan. Uji  $t$  dapat digunakan untuk menguji perbedaan rata-rata suatu variabel terikat yang berskala kontinu atau ordinal antara dua kelompok sampel. Dengan kata lain, Uji  $t$  digunakan untuk membandingkan rata-rata satu variabel terikat yang diukur dengan skala kontinu atau ordinal dan satu variabel bebas yang berskala kategorikal yang hanya terdiri atas *dua* kategori. Contohnya, suatu penelitian ingin melakukan uji  $t$  dua sampel untuk mengetahui perbedaan hasil belajar matematika siswa yang dibagi menjadi dua kelompok berdasarkan desain penelitian yang digunakan, yaitu kelompok kontrol dan kelompok eksperimen.

Namun, uji  $t$  terbatas hanya pada kondisi dimana hanya terdapat satu variabel bebas yang terdiri atas dua kategori saja (misalnya, dua kelompok eksperimental). Dalam kebanyakan kasus, peneliti ingin membandingkan perbedaan rata-rata variabel terikat berskala kontinu dengan dua, tiga, atau lebih variabel bebas berskala kategorikal yang masing-masing memiliki dua atau lebih kategori yang berbeda. Dalam kondisi ini, uji  $t$  tidak dapat digunakan. Sebaliknya, suatu uji yang dapat digunakan adalah analisis varians (ANAVA). Teknik ini memungkinkan peneliti untuk menganalisis kasus dimana terdapat beberapa variabel bebas (faktor) berskala kategorikal. Hasil ANAVA memberitahukan sejauh mana variabel-variabel bebas berinteraksi satu sama lain dan pengaruh interaksi tersebut terhadap variabel terikat.

Kemudian muncul suatu pertanyaan, mengapa tidak melakukan uji  $t$  berulang untuk membandingkan rata-rata variabel terikat antara setiap kombinasi dua kategori variabel bebas? Misalkan, dalam suatu kasus terdapat tiga kategori/kelompok eksperimental dan peneliti tertarik untuk menguji perbedaan antara

ketiga kelompok ini. Jika menggunakan uji  $t$  berulang terhadap setiap kombinasi pasangan, maka harus dilakukan tiga uji secara terpisah, yaitu untuk membandingkan kelompok 1 dan 2, kelompok 1 dan 3, serta kelompok 2 dan 3. Jika digunakan taraf signifikansi 0,05, maka untuk setiap uji  $t$  yang dilakukan, peluang terjadi kesalahan dalam menolak  $H_0$ , atau yang biasa disebut kesalahan tipe I, yaitu hanya 5%. Karenanya, peluang tidak terjadi kesalahan tipe I, yaitu 95%. Jika diasumsikan ketiga tes saling bebas sehingga peluangnya dapat dikalikan, maka peluang tidak terjadi kesalahan tipe I pada ketiga tes secara bersamaan, yaitu  $(0,95)^3 = 0,95 \times 0,95 \times 0,95 = 0,857$ . Berdasarkan hasil ini, peluang terjadi kesalahan tipe I pada ketiga tes secara bersamaan, yaitu  $1 - 0,857 = 0,143$  atau 14,3%. Nilai tersebut meningkat dari 5% ke 14,3% dan merupakan nilai yang cukup besar untuk digunakan sebagai kriteria penerimaan hipotesis. Oleh karena itu, dengan alasan inilah ANAVA lebih tepat untuk digunakan daripada uji  $t$  berulang untuk meminimalisir peluang terjadinya kesalahan tipe I.

ANAVA digunakan untuk menguji hipotesis nol bahwa sampel pada  $i$  kelompok diambil secara acak dari populasi-populasi yang memiliki rata-rata yang sama. Secara matematis, hipotesis nol dapat ditulis

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \dots = \mu_{i-1} = \mu_i$$

Hipotesis alternatifnya ( $H_a$ ) menyatakan bahwa setidaknya terdapat dua kelompok sampel yang diambil secara acak dari populasi yang memiliki rata-rata berbeda. Penolakan terhadap  $H_0$  mengindikasikan bahwa terdapat paling sedikit sepasang ketidaksetaraan rata-rata yang terjadi pada kelompok-kelompok sampel tersebut.

## 12.2 Analisis Varians Satu Jalur (ANAVA Satu Jalur)

ANAVA satu jalur merupakan teknik analisis varians yang digunakan untuk menguji perbedaan rata-rata suatu variabel terikat berskala kontinu atau ordinal antara tiga atau lebih kategori (kelompok) atau tingkatan satu variabel bebas (faktor) berskala kategorikal. Tentunya ANAVA satu jalur juga dapat digunakan untuk

menguji perbedaan rata-ran antara dua kelompok dan hasilnya akan sama dengan uji  $t$  dua sampel yang telah dibahas di Bab 10. ANAVA satu jalur menguji perbedaan rata-ran dengan menguji varians. Oleh karena itu, uji ini didasarkan pada bukti matematis bahwa data sampel dapat dipakai untuk menghasilkan dua taksiran bebas dari varians populasi, yaitu:

1. Taksiran varians *dalam* kelompok sampel, yaitu taksiran yang didasarkan perbedaan antara nilai di suatu kelompok sampel dan nilai lainnya pada kelompok tersebut.
2. Taksiran varians *antar* kelompok sampel, yaitu taksiran yang didasarkan pada perbedaan rata-ran antara suatu kelompok sampel dan kelompok lainnya.

Jika semua sampel berasal dari suatu populasi yang sama dan berdistribusi normal (atau berasal dari populasi-populasi berdistribusi normal dengan rata-ran dan varians yang sama), dapat dibuktikan secara matematis bahwa taksiran varians antar kelompok sampel dan varians dalam kelompok sampel akan mendekati sama satu dengan yang lain dan sama dengan varians populasi ( $\sigma^2$ ). Semakin besar taksiran varians antar kelompok dibandingkan dengan taksiran varians dalam kelompok, semakin besar pula kemungkinan kelompok-kelompok sampel tersebut bukan berasal dari populasi dengan rata-ran yang sama. Perbandingan antara taksiran varians antar kelompok dan varians dalam kelompok sampel dinyatakan dengan rasio  $F$  berikut.

$$F = \frac{\text{taksiran varians antar kelompok}}{\text{taksiran varians dalam kelompok}}$$

Jika rasio  $F$  mendekati 1,0, maka  $H_0$  benar dan taksiran varians antar kelompok sampel mendekati sama dengan taksiran varians dalam kelompok sampel. Demikian pula, rasio  $F$  kurang dari 1,0 mengindikasikan bahwa  $H_0$  diterima karena taksiran varians antar kelompok kurang dari varians dalam kelompok sampel. Namun, jika rasio  $F$  semakin lebih dari 1,0, maka  $H_0$  ditolak dan kita dapat semakin yakin bahwa kategori-kategori variabel bebas memiliki pengaruh terhadap variabel terikat atau dengan kata lain, rata-ran

variabel terikat memiliki perbedaan yang signifikan antara kelompok yang satu dan kelompok lainnya pada variabel bebas.

Berikut akan diuraikan langkah-langkah perhitungan manual ANAVA satu jalur. Meskipun ketika bekerja dengan data penelitian yang nyata ke depannya, perhitungan ANAVA tidak akan menggunakan persamaan-persamaan berikut, perhitungan ANAVA secara manual akan memberikan pemahaman yang lebih tentang pembilang dan penyebut pada rasio  $F$  dari pada hanya mengklik menu pada program komputer dan membaca hasilnya.

#### 1. Hipotesis Uji

$H_0$ : tidak terdapat perbedaan yang signifikan secara statistik pada rata-rata variabel terikat antara suatu kategori dan kategori lainnya pada variabel bebas

$H_a$ : terdapat perbedaan yang signifikan secara statistik pada rata-rata variabel terikat antara suatu kategori dan kategori lainnya pada variabel bebas

#### 2. Hipotesis Statistik

$H_0$ :  $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \dots = \mu_{i-1} = \mu_i$

$H_a$ : minimal ada dua kombinasi rata-rata yang tidak sama

#### 3. Menghitung Jumlah Kuadrat Total ( $JK_T$ )

Jumlah kuadrat total ( $JK_T$ ) dapat dihitung dengan persamaan berikut.

$$JK_T = \sum X^2 - N_T \bar{X}_T^2$$

Dengan  $X$  = nilai yang diamati pada semua kategori/kelompok,  $N_T$  = jumlah total pengamatan pada suatu percobaan, dan  $\bar{X}_T$  = rata-rata total untuk semua pengamatan.

#### 4. Menghitung Jumlah Kuadrat antar Kelompok ( $JK_A$ )

Jumlah kuadrat antar kelompok ( $JK_A$ ) dapat dihitung dengan persamaan berikut (berlaku untuk  $n$  sama).

$$JK_A = n \sum_{i=1}^k (\bar{X}_i - \bar{X}_T)^2 \text{ atau } JK_A = n \sum_{i=1}^k \bar{X}_i^2 - N_T \bar{X}_T^2$$

Dengan  $n$  = jumlah pengamatan setiap kategori,  $k$  = jumlah kategori, dan  $\bar{X}_i$  = rata-rata kelompok  $i$ . Jika  $n$  tidak sama, maka berlaku persamaan

$$JK_A = \sum_{i=1}^k N_i (\bar{X}_i - \bar{X}_T)^2$$

Dengan  $N_i$  = banyaknya pengamatan pada kategori  $i$ .

5. Menghitung Jumlah Kuadrat dalam Kelompok ( $JK_D$ )

Jumlah kuadrat dalam kelompok ( $JK_D$ ) dapat dihitung dengan persamaan berikut.

$$JK_T = JK_A + JK_D$$

6. Menghitung Rataan Kuadrat antar Kelompok ( $RK_A$ )

Rataan kuadrat antar kelompok ( $RK_A$ ) merupakan taksiran untuk varians antar kelompok dan dapat diperoleh dengan persamaan berikut.

$$RK_A = \frac{JK_A}{db_A}$$

Dengan  $db_A$  = derajat kebebasan antar kelompok, yang mana  $db_A = k - 1$ .

7. Menghitung Rataan Kuadrat dalam Kelompok ( $RK_D$ )

Rataan kuadrat dalam kelompok ( $RK_D$ ) merupakan taksiran untuk varians dalam kelompok dan dapat diperoleh dengan persamaan berikut.

$$RK_D = \frac{JK_D}{db_D}$$

Dengan  $db_D$  = derajat kebebasan dalam kelompok, yang mana  $db_D = N_T - k$ .

8. Menghitung Rasio  $F$  ( $F_{hitung}$ )

Nilai rasio  $F$  dapat dihitung dengan persamaan berikut.

$$F_{hitung} = \frac{RK_A}{RK_D}$$

9. Menentukan Nilai  $F_{tabel}$

Nilai  $F_{tabel}$  dapat ditentukan dengan cara mencari nilai yang bersesuaian pada tabel  $F$ , dengan  $db$  pembilang =  $db_A$  dan  $db$  penyebut =  $db_D$ .

10. Membuat Keputusan dan Menarik Simpulan Hasil ANAVA

Jika  $F_{hitung} > F_{tabel}$  pada suatu taraf signifikansi tertentu (misalnya 5%), maka tolak  $H_0$  dan menyimpulkan bahwa terdapat



perbedaan rata-rata variabel terikat yang signifikan secara statistik antara suatu kategori dan kategori lainnya pada variabel bebas.

#### 11. Menghitung Ukuran Pengaruh Jika Hasil Signifikan

Jika hasil ANAVA menyatakan terdapat perbedaan yang signifikan secara statistik, maka perlu dihitung ukuran pengaruh untuk mengetahui seberapa besar perbedaan rata-rata yang terjadi. Ukuran pengaruh yang dipakai, yaitu *eta-squared* ( $\eta^2$ ), yang dihitung dengan persamaan berikut.

$$\eta^2 = \frac{db_A(F_{hitung} - 1)}{db_A F_{hitung} + db_D}$$

#### 12. Membuat Tabel Ringkasan ANAVA Satu Jalur

Tabel ringkasan tes ANAVA satu jalur memuat informasi, seperti taksiran varians antar kelompok, varians dalam kelompok, rata-rata kuadrat antar kelompok dan dalam kelompok, derajat kebebasan tiap rata-rata kuadrat,  $F_{hitung}$ ,  $F_{tabel}$ , dan nilai  $p$ .

### Contoh Masalah

Diberikan suatu kumpulan data hipotetis tentang percobaan tikus dalam labirin seperti yang disajikan pada tabel 12.1. Dalam percobaan tersebut, peneliti tertarik untuk mengamati banyaknya kesalahan yang dilakukan oleh sampel tikus dalam labirin berdasarkan jenis imbalan yang diberikan diakhir. Kelompok 1 mendapat imbalan berupa larutan akuades ( $H_2O$ ) 100%; kelompok 2 mendapat imbalan berupa larutan yang mengandung 50%  $H_2O$  dan 50% gula; dan kelompok 3 mendapat imbalan berupa larutan gula 100%. Hitunglah ANAVA satu jalur untuk data tersebut.

**Tabel 12.1.** Hasil Percobaan Tikus dalam Labirin.

No.	Kelompok 1	Kelompok 2	Kelompok 3	$\Sigma$	$\bar{X}$
1.	1	8	6	15	5
2.	3	7	5	15	5
3.	2	5	3	10	3,33
4.	1	4	3	8	2,67
5.	3	6	3	12	4
$\Sigma$	10	30	20	60	
$\bar{X}$	2	6	4		4

Penyelesaian:

1. Hipotesis Uji

$H_0$  : tidak ada perbedaan yang signifikan secara statistik pada banyaknya kesalahan yang dilakukan oleh tikus pada ketiga kelompok berdasarkan jenis imbalan yang diberikan

$H_a$  : ada perbedaan yang signifikan secara statistik pada banyaknya kesalahan yang dilakukan oleh tikus pada ketiga kelompok berdasarkan jenis imbalan yang diberikan

2. Hipotesis Statistik

$$H_0: \mu_1 = \mu_2 = \mu_3$$

$H_a$  : minimal ada dua kombinasi rataan yang tidak sama

3. Menghitung Jumlah Kuadrat Total ( $JK_T$ )

$$\begin{aligned} JK_T &= \sum X^2 - N_T \bar{X}_T^2 \\ &= (1^2 + 3^2 + \dots + 7^2 + \dots + 3^2 + 3^2) - 15 \cdot 4^2 \\ JK_T &= 302 - 240 = 62 \end{aligned}$$

4. Menghitung Jumlah Kuadrat antar Kelompok ( $JK_A$ )

$$\begin{aligned} JK_A &= n \sum_{i=1}^k (\bar{X}_i - \bar{X}_T)^2 \\ &= 5[(2-4)^2 + (6-4)^2 + (4-4)^2] \\ JK_A &= 5(4 + 4 + 0) = 40 \end{aligned}$$

5. Menghitung Jumlah Kuadrat dalam Kelompok ( $JK_D$ )

$$\begin{aligned} JK_T &= JK_A + JK_D \\ JK_D &= JK_T - JK_A = 62 - 40 = 22 \end{aligned}$$

6. Menghitung Rataan Kuadrat antar Kelompok ( $RK_A$ )

$$RK_A = \frac{JK_A}{db_A} = \frac{40}{3-1} = \frac{40}{2} = 20$$

7. Menghitung Rataan Kuadrat dalam Kelompok ( $RK_D$ )

$$RK_D = \frac{JK_D}{db_D} = \frac{22}{15-3} = \frac{22}{12} = 1,83$$

8. Menghitung Rasio  $F$  ( $F_{hitung}$ )

$$F_{hitung} = \frac{RK_A}{RK_D} = \frac{20}{1,83} = 10,9$$

9. Menentukan Nilai  $F_{tabel}$

Nilai  $F_{tabel}$  untuk  $db_A = 2$  dan  $db_D = 12$  dengan taraf signifikansi 5% adalah 3,88.

10. Membuat Keputusan dan Menarik Simpulan Hasil ANAVA

Karena  $F_{hitung} = 10,9 > 3,88 = F_{tabel}$ , maka tolak  $H_0$  dan menyimpulkan bahwa ada perbedaan yang signifikan secara

statistik pada banyaknya kesalahan yang dilakukan oleh tikus pada ketiga kelompok berdasarkan jenis imbalan yang diberikan.

11. Menghitung Ukuran Pengaruh Jika Hasil Signifikan

$$\eta^2 = \frac{db_A(F_{hitung} - 1)}{db_A F_{hitung} + db_D} = \frac{2(10,9 - 1)}{2(10,9) + 12} = \frac{2(9,9)}{21,8 + 12} = \frac{19,8}{33,8} = 0,586$$

(jenis imbalan yang diberikan berpengaruh cukup besar sekitar 58,6% terhadap banyaknya kesalahan yang dilakukan oleh tikus dalam labirin pada percobaan tersebut)

12. Membuat Tabel Ringkasan ANAVA Satu Jalur

**Tabel 12.2.** Ringkasan ANAVA Satu Jalur pada Percobaan Tikus dalam Labirin.

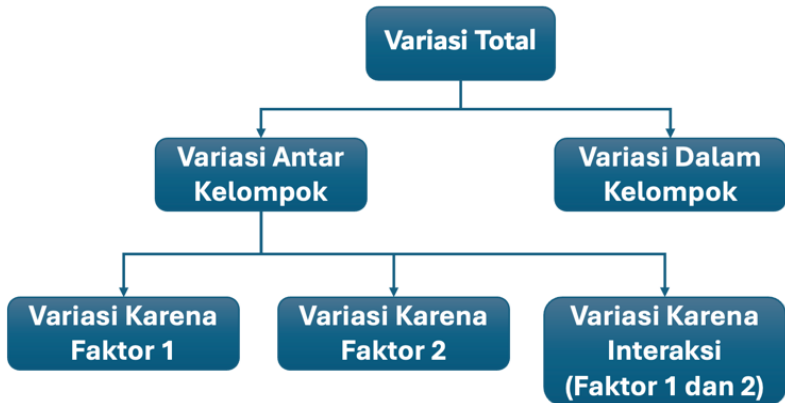
Sumber Varians	<i>JK</i>	<i>db</i>	<i>RK</i>	<i>F<sub>hitung</sub></i>	<i>p</i>
Antar Kelompok	40	2	20	10,9	<0,05
Dalam Kelompok	22	12	1,83		
Total	62	24			

## 12.3 Analisis Varians Dua Jalur (ANAVA Dua Jalur)

ANAVA satu jalur yang telah dijelaskan di bagian 12.2 digunakan untuk menyelidiki hubungan *satu* variabel bebas (faktor) dengan satu variabel terikat, dimana variabel bebas tersebut memiliki dua atau lebih kategori (atau kelompok). Sementara itu, teknik yang digunakan untuk menganalisis hubungan *dua* atau *lebih* variabel bebas dengan satu variabel terikat, dimana masing-masing variabel bebas tersebut memiliki dua atau lebih kategori (atau kelompok), disebut **desain faktorial**. Salah satu desain faktorial yang akan dibahas adalah analisis varians dua jalur (ANAVA dua jalur). Disebut *dua jalur* karena teknik ini menganalisis hubungan antara *dua* variabel bebas dengan satu variabel terikat.

Logika yang dipakai pada ANAVA dua jalur sama seperti pada ANAVA satu jalur, yaitu jumlah kuadrat total dibagi menjadi jumlah kuadrat *antar* kelompok dan jumlah kuadrat *dalam* kelompok. Namun, pada ANAVA dua jalur, jumlah kuadrat antar kelompok dibagi lagi menjadi variasi dikarenakan pengaruh faktor 1, variasi dikarenakan pengaruh faktor 2, dan variasi dikarenakan pengaruh kedua faktor

tersebut secara bersamaan (disebut sebagai *interaksi*) yang diilustrasikan pada gambar 12.1.



**Gambar 12.1.** Pembagian Variasi pada Desain Faktorial Dua Jalur.

Berikut akan diuraikan langkah-langkah perhitungan manual ANAVA dua jalur.

1. Hipotesis Uji

- 1) tidak terdapat perbedaan yang signifikan secara statistik pada rata-rata variabel terikat antara suatu kategori dan kategori lainnya pada faktor 1;
- 2) tidak terdapat perbedaan yang signifikan secara statistik pada rata-rata variabel terikat antara suatu kategori dan kategori lainnya pada faktor 2;
- 3) hubungan antara variabel terikat dan faktor 1 tidak berbeda berdasarkan kategori pada faktor 2 (atau tidak terdapat perbedaan yang signifikan secara statistik pada rata-rata variabel terikat berdasarkan kombinasi khusus dari kategori faktor 1 dan 2).

$H_0$ :

2. Menghitung Jumlah Kuadrat Total ( $JK_T$ )

Jumlah kuadrat total ( $JK_T$ ) dapat dihitung dengan persamaan berikut.

$$JK_T = \sum X^2 - N_T \bar{X}_T^2$$

Dengan  $X$  = nilai yang diamati pada semua kategori/kelompok,  $N_T$  = jumlah total pengamatan pada suatu percobaan, dan  $\bar{X}_T$  = rata-rata total untuk semua pengamatan.

3. Menghitung Jumlah Kuadrat antar Kelompok ( $JK_A$ )

Jumlah kuadrat antar kelompok ( $JK_A$ ) dapat dihitung dengan persamaan berikut (berlaku untuk  $n$  sama).

$$JK_A = n \sum_{i=1}^k (\bar{X}_i - \bar{X}_T)^2 \text{ atau } JK_A = n \sum_{i=1}^k \bar{X}_i^2 - N_T \bar{X}_T^2$$

Dengan  $n$  = jumlah pengamatan setiap kategori,  $k$  = jumlah kategori, dan  $\bar{X}_i$  = rata-rata kelompok  $i$ . Jika  $n$  tidak sama, maka berlaku persamaan

$$JK_A = \sum_{i=1}^k N_i (\bar{X}_i - \bar{X}_T)^2$$

Dengan  $N_i$  = banyaknya pengamatan pada kategori  $i$ .

4. Menghitung Jumlah Kuadrat dalam Kelompok ( $JK_D$ )

Jumlah kuadrat dalam kelompok ( $JK_D$ ) dapat dihitung dengan persamaan berikut.

$$JK_T = JK_A + JK_D$$

5. Menghitung Jumlah Kuadrat untuk Faktor 1 ( $JK_1$ )

$JK_1$  dapat dihitung dengan persamaan berikut.

$$JK_1 = rn \sum_{i=1}^c \bar{X}_i^2 - N_T \bar{X}_T^2$$

Dengan  $c$  = banyaknya kategori pada faktor 1, dan  $r$  = banyaknya kategori pada faktor 2.

6. Menghitung Jumlah Kuadrat untuk Faktor 2 ( $JK_2$ )

$JK_2$  dapat dihitung dengan persamaan berikut.

$$JK_2 = cn \sum_{i=1}^r \bar{X}_i^2 - N_T \bar{X}_T^2$$

7. Menghitung Jumlah Kuadrat untuk Interaksi ( $JK_{1 \times 2}$ )

$JK_{1 \times 2}$  dapat dihitung dengan persamaan berikut.

$$JK_{1 \times 2} = JK_A - JK_1 - JK_2$$

8. Menghitung Rataan Kuadrat dalam Kelompok ( $RK_D$ )

Rataan kuadrat dalam kelompok ( $RK_D$ ) merupakan taksiran untuk varians dalam kelompok dan dapat diperoleh dengan persamaan berikut.

$$RK_D = \frac{JK_D}{db_D}$$

Dengan  $db_D$  = derajat kebebasan dalam kelompok, yang mana  $db_D = N_T - rc$ .

9. Menghitung Rataan Kuadrat untuk Faktor 1 ( $RK_1$ )

$RK_1$  dapat diperoleh dengan persamaan berikut.

$$RK_1 = \frac{JK_1}{db_1}$$

Dengan  $db_1$  = derajat kebebasan untuk faktor 1, yang mana  $db_1 = c - 1$ .

10. Menghitung Rataan Kuadrat untuk Faktor 2 ( $RK_2$ )

$RK_2$  dapat diperoleh dengan persamaan berikut.

$$RK_2 = \frac{JK_2}{db_2}$$

Dengan  $db_2$  = derajat kebebasan untuk faktor 2, yang mana  $db_2 = r - 1$ .

11. Menghitung Rataan Kuadrat untuk Interaksi ( $RK_{1 \times 2}$ )

$RK_{1 \times 2}$  dapat diperoleh dengan persamaan berikut.

$$RK_{1 \times 2} = \frac{JK_{1 \times 2}}{db_{1 \times 2}}$$

Dengan  $db_{1 \times 2}$  = derajat kebebasan untuk interaksi faktor 1 dan 2, yang mana  $db_{1 \times 2} = (c - 1)(r - 1)$ .

12. Menghitung Rasio  $F$  ( $F_{hitung}$ )

Nilai rasio  $F$  untuk faktor 1 dapat dihitung dengan persamaan berikut.

$$F_1 = \frac{RK_1}{RK_D}$$

Nilai rasio  $F$  untuk faktor 2 dapat dihitung dengan persamaan berikut.

$$F_2 = \frac{RK_2}{RK_D}$$

Nilai rasio  $F$  untuk interaksi faktor 1 dan 2 dapat dihitung dengan persamaan berikut.

$$F_{1 \times 2} = \frac{RK_{1 \times 2}}{RK_d}$$

13. Menentukan Nilai  $F_{tabel}$

Nilai  $F_{tabel}$  dapat ditentukan dengan cara mencari nilai yang bersesuaian pada tabel  $F$ , dengan  $db$  pembilang dan  $db$  penyebut.

14. Membuat Keputusan dan Menarik Simpulan Hasil ANAVA

Jika  $F_{hitung} > F_{tabel}$  pada suatu taraf signifikansi tertentu (misalnya 5%) untuk masing-masing pengaruh faktor 1, 2, dan interaksinya, maka tolak  $H_0$ .

15. Membuat Tabel Ringkasan ANAVA Satu Jalur

Tabel ringkasan tes ANAVA satu jalur memuat informasi, seperti taksiran varians antar kelompok, varians dalam kelompok, varians untuk faktor 1, 2, dan interaksi, rata-rata kuadrat antar kelompok, dalam kelompok, rata-rata kuadrat untuk faktor 1, 2, dan interaksi, derajat kebebasan tiap rata-rata kuadrat, dan  $F_{hitung}$ .

### Contoh Masalah

Diberikan kumpulan data hipotetis nilai kuis statistika 18 mahasiswa berdasarkan gender dan tingkatan fobia terhadap matematika seperti yang disajikan pada tabel 12.3. Hitunglah ANAVA dua jalur untuk data tersebut.

**Tabel 12.3.** Nilai Kuis Statistika Berdasarkan Tingkat Fobia dan Gender.

Faktor	Fobia Rendah	Fobia Sedang	Fobia Tinggi	$\Sigma$	$\bar{X}$
Laki-laki	5	8	5	18	5,89
	4	5	6	15	
	4	7	9	20	
	$\bar{X} = 4,33$	$\bar{X} = 6,67$	$\bar{X} = 6,67$		
Perempuan	3	7	4	14	5,22
	4	5	9	18	
	2	6	7	15	
	$\bar{X} = 3$	$\bar{X} = 6$	$\bar{X} = 6,67$		

Faktor	Fobia Rendah	Fobia Sedang	Fobia Tinggi	$\Sigma$	$\bar{X}$
$\Sigma$	22	38	40	100	
$\bar{X}$	3,67	6,33	6,67		5,56

Penyelesaian:

Hipotesis uji:

- 1) tidak terdapat perbedaan yang signifikan secara statistik pada rata-rata nilai kuis statistika antara kategori mahasiswa yang memiliki fobia rendah, sedang, dan tinggi;
- $H_0$ : 2) tidak terdapat perbedaan yang signifikan secara statistik pada rata-rata nilai kuis statistika antara mahasiswa dan mahasiswi;
- 3) hubungan antara nilai kuis matematika dan tingkatan fobia matematika tidak berbeda berdasarkan gender.

Menentukan jumlah kuadrat:

$$\begin{aligned}
 JK_T &= \sum X^2 - N_T \bar{X}_T^2 \\
 &= (5^2 + 4^2 + \dots + 9^2 + 7^2) - 18(5,56)^2 \\
 &= 622 - 18(30,91) = 622 - 556,44 = 65,56
 \end{aligned}$$

$$\begin{aligned}
 JK_A &= n \sum_{i=1}^k \bar{X}_i^2 - N_T \bar{X}_T^2 \\
 &= 3(4,33^2 + 6,67^2 + \dots + 6^2 + 6,67^2) - 18(5,56)^2 \\
 &= 3(197,22) - 556,44 = 591,66 - 556,44 = 35,22
 \end{aligned}$$

$$JK_D = JK_T - JK_A = 65,56 - 35,22 = 30,34$$

$$\begin{aligned}
 JK_1 &= rn \sum_{i=1}^c \bar{X}_i^2 - N_T \bar{X}_T^2 \\
 &= 2 \cdot 3(3,67^2 + 6,33^2 + 6,67^2) - 18(5,56)^2 \\
 &= 6(98,03) - 556,44 = 588,18 - 556,44 = 31,74
 \end{aligned}$$

$$\begin{aligned}
 JK_2 &= cn \sum_{i=1}^r \bar{X}_i^2 - N_T \bar{X}_T^2 \\
 &= 3 \cdot 3(5,89^2 + 5,22^2) - 18(5,56)^2 \\
 &= 9(61,94) - 556,44 = 557,46 - 556,44 = 1,02
 \end{aligned}$$

$$JK_{1 \times 2} = JK_A - JK_1 - JK_2 = 35,22 - 31,74 - 1,02 = 2,46$$



Menentukan rata-rata kuadrat:

$$RK_D = \frac{JK_D}{db_D} = \frac{JK_D}{N_T - rc} = \frac{30,34}{18 - 6} = 2,53$$

$$RK_1 = \frac{JK_1}{db_1} = \frac{JK_1}{c - 1} = \frac{31,74}{3 - 1} = 15,87$$

$$RK_2 = \frac{JK_2}{db_2} = \frac{JK_2}{r - 1} = \frac{1,02}{2 - 1} = 1,02$$

$$RK_{1 \times 2} = \frac{JK_{1 \times 2}}{db_{1 \times 2}} = \frac{JK_{1 \times 2}}{(c - 1)(r - 1)} = \frac{2,46}{(3 - 1)(2 - 1)} = \frac{2,46}{2} = 1,23$$

Menentukan rasio  $F$ :

$$F_1 = \frac{RK_1}{RK_D} = \frac{15,87}{2,53} = 6,27$$

$$F_2 = \frac{RK_2}{RK_D} = \frac{1,02}{2,53} = 0,4$$

$$F_{1 \times 2} = \frac{RK_{1 \times 2}}{RK_D} = \frac{1,23}{2,53} = 0,49$$

Menentukan nilai  $F_{tabel}$ :

**Tabel 12.4.** Penentuan Nilai  $F_{tabel}$ .

No.	Sumber Variasi	Derajat Kebebasan	$F_{tabel}$
1.	Fobia Matematika	$db_1 = 2; db_D = 12$	3,88
2.	Gender	$db_2 = 1; db_D = 12$	4,75
3.	Interaksi (fobia x gender)	$db_{1 \times 2} = 2; db_D = 12$	3,88

Membuat keputusan dan menarik simpulan hasil ANAVA dua jalur:

Karena  $F_1 = 6,27 > 3,88 = F_{tabel}$  untuk faktor fobia, maka tolak  $H_0$  dan menyimpulkan bahwa ada perbedaan yang signifikan secara statistik pada rata-rata nilai kuis statistika mahasiswa berdasarkan tingkatan fobia. Karena  $F_2 = 0,4 < 4,75 = F_{tabel}$  untuk faktor gender, maka terima  $H_0$  dan menyimpulkan bahwa tidak ada perbedaan gender pada skor kuis statistika. Karena  $F_{1 \times 2} = 0,49 < 3,88 = F_{tabel}$  untuk interaksi fobia dan gender, maka terima  $H_0$  dan menyimpulkan bahwa hubungan antara nilai kuis statistika dan tingkatan fobia matematika tidak berbeda berdasarkan gender.

**Tabel 12.5.** Ringkasan ANAVA Dua Jalur.

Sumber variasi	$JK$	$db$	$RK$	$F_{hitung}$
Fobia	31,74	2	15,87	6,27
Gender	1,02	1	1,02	0,4
Interaksi (fobia x gender)	2,46	2	1,23	0,49
Dalam kategori (error)	30,34	12	2,53	

## DAFTAR PUSTAKA

- Agresti, A. 2018. Statistical Methods for the Social Sciences. Fifth Edition. Boston Pearson.
- Field, A. 2000. Discovering Statistics using SPSS for Windows. London: SAGE Publications Ltd.
- Gunawan, M. A. 2013. Statistik untuk Penelitian Pendidikan. Yogyakarta: Parama Publishing.
- Welkowitz, J., Cohen, B. H. and Ewen, R. B. 2006. Introductory Statistics for the Behavioural Sciences. Sixth Edition. Hoboken: John Wiley & Sons, Inc.



## **BIODATA PENULIS**



### **Prima Kurniati Hamzah**

Dosen tetap di UIN Imam Bonjol, Padang

Penulis lahir di Padang tanggal 18 September 1989. Penulis adalah dosen tetap di UIN Imam Bonjol, Padang. Menyelesaikan pendidikan S1 pada prodi Ilmu Kesehatan Masyarakat dengan peminatan epidemiologi di Fakultas Kedokteran, Universitas Andalas pada tahun 2011. Melanjutkan S2 pada prodi Ilmu Kesehatan Masyarakat peminatan epidemiologi lapangan, Fakultas Kedokteran Universitas Gadjah Mada pada tahun 2013. Penulis menyukai analisis dan manajemen data dan melakukan riset kolaborasi di bidang kesehatan. Memiliki publikasi di Jurnal nasional dan Internasional. Sekarang aktif terlibat dalam penguatan implementasi Kawasan Tanpa Rokok serta merancang kebijakan Kawasan Tanpa Rokok di beberapa Kota/ Kabupaten di Provinsi Sumatera Barat. Mendapatkan hibah project Tobacco Control dari Vital Strategies, Universitas John Hopkins, Amerika Serikat pada tahun 2023-2025. Penulis dapat dihubungi melalui e-mail: [primakurniati@uinib.ac.id](mailto:primakurniati@uinib.ac.id)

## BIODATA PENULIS



**Dr. MAhkam Alwi, S.Psi., M.Si**  
Dosen Program Studi Psikologi  
Fakultas Psikologi UNM

Penulis lahir di Padang tanggal 18 Mei 1970. Penulis adalah dosen tetap pada Program Studi Psikologi Universitas Negeri Makassar. Menyelesaikan pendidikan S1 pada Jurusan Psikologi Universitas Proklamasi 45 Yogyakarta, S2 pada Jurusan Psikologi UGM dan S3 Jurusan Psikologi minat Pendidikan dan Perkembangan di Universitas Airlangga

Penulis dapat dihubungi melalui e-mail: [m.ahkama@unm.ac.id](mailto:m.ahkama@unm.ac.id)

## **BIODATA PENULIS**



**Raden Lutfi Zakaria, S.T., M.M.**

**Dosen Program Studi Teknik Mesin - Fakultas Teknologi Industri  
Universitas Kebangsaan Republik Indonesia**

Penulis adalah dosen tetap pada Program Studi Teknik Mesin di Universitas Kebangsaan Republik Indonesia (UKRI) di Bandung. Di samping menjalankan tugasnya sebagai pendidik dan akademisi, Raden Lutfi Zakaria juga memiliki ketertarikan besar pada dunia otomotif, yang menjadi salah satu hobi yang ia tekuni di waktu luangnya. Dengan kombinasi keahlian akademik dan minat pribadinya, beliau memberikan kontribusi yang berharga dalam lingkungan akademis serta komunitas otomotif. Penulis dapat dihubungi melalui e-mail: [lut.zakaria2003@gmail.com](mailto:lut.zakaria2003@gmail.com)

## BIODATA PENULIS



**Susana Labuem, S.Pd., M.Pd**

Dosen Program Studi Pendidikan Matematika  
PSDKU ARU Fakultas Keguruan & Ilmu Pendidikan  
Universitas Pattimura

Penulis lahir di desa Longgar, Kabupaten Kepulauan Aru, Provinsi Maluku pada tanggal 7 Oktober 1990 dan merupakan anak kedua dari tiga bersaudara. Penulis adalah anak dari pasangan bapak Julianus Labuem dan ibu Abigael Siarukin. Penulis menempuh pendidikan di SD Negeri 5 Inpres Dobo mulai tahun 1996 dan lulus pada tahun 2002. Penulis kemudian melanjutkan pendidikan di SMP Negeri 1 Dobo mulai tahun 2002 dan lulus pada tahun 2005. Selanjutnya, penulis melanjutkan pendidikan di SMA Negeri 1 Dobo pada tahun 2005 dan lulus pada tahun 2008. Penulis kemudian melanjutkan pendidikan S1 di Universitas Pattimura dan lulus pada tahun 2013. Setelah lulus pendidikan S1, penulis mengajar di SMP Negeri 1 Dobo dan SMK PGRI Dobo dari tahun 2013 sampai 2017. Penulis kemudian melanjutkan pendidikan S2 pada tahun 2017 di Pascasarjana jurusan Matematika Universitas Negeri Malang dan kemudian lulus pada tahun 2019. Setelah lulus pendidikan S2, penulis mengajar di Program Studi Pendidikan Matematika PSDKU Universitas Pattimura di kabupaten Kepulauan Aru. Pada tahun 2021 penulis kemudian melanjutkan pendidikan S3 di Pascasarjana Program Studi Pendidikan Matematika Universitas Negeri Malang. Penulis dapat dihubungi melalui e-mail: [susana.labuem@lecturer.unpatti.ac.id](mailto:susana.labuem@lecturer.unpatti.ac.id)

## BIODATA PENULIS



**Ahmad Ridfah, S.Psi., M.Psi., Psikolog**

Dosen Program Studi Psikologi

Fakultas Psikologi Universitas Negeri Makassar

Penulis lahir di Ujung Pandang tanggal 10 Agustus 1982. Penulis adalah dosen PNS pada Program Studi Psikologi (S1) Fakultas Psikologi Universitas Negeri Makassar. Penulis menyelesaikan pendidikan sarjana psikologi pada Jurusan Psikologi Universitas Negeri Makassar pada tahun 2006. Kemudian pada tahun 2012, penulis menyelesaikan pendidikan pada Magister Psikologi Profesi Mayoring Klinis Dewasa Fakultas Psikologi Universitas Padjadjaran. Penulis merupakan pengampu mata kuliah Statistika Dasar, Statistika Lanjut, Pengantar Psikometrika, Psikometrika Lanjut, Pengantar Intervensi Psikologi, Psikologi Klinis, Pengukuran Kepribadian Non Proyektif, Pengukuran Intelligensi dan Bakat, Metode Observasi, dan Psikologi Forensik. Penulis dapat dihubungi melalui e-mail: [ahmad.ridfah@unm.ac.id](mailto:ahmad.ridfah@unm.ac.id).



## BIODATA PENULIS



**Vera Rosalina Bulu, M.Pd.**

Dosen Program Studi Pendidikan Guru Sekolah Dasar, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Nusa Cendana

Penulis lahir di Soe tanggal 4 Oktober 1990. Penulis adalah dosen tetap pada Program Studi Pendidikan Guru Sekolah Dasar, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Nusa Cendana. Menyelesaikan pendidikan S1 pada Jurusan Pendidikan Matematika, Universitas Nusa Cendana pada tahun 2011 dan melanjutkan S2 pada Jurusan Pendidikan Matematika, Universitas Sebelas Maret pada tahun 2015. Penulis menekuni bidang Menulis sejak tahun 2015. Tulisan yang dihasilkan berupa artikel yang diterbitkan pada jurnal internasional bereputasi, jurnal nasional terakreditasi dan buku ber ISBN Matematika Dasar PGSD.

Penulis dapat dihubungi melalui e-mail: [Veraros0451@gmail.com](mailto:Veraros0451@gmail.com).

## **BIODATA PENULIS**



### **Farida**

**Dosen Program Studi Teknik Industri  
Fakultas Teknik, Universitas Mercu Buana**

Penulis adalah dosen pada Program Studi Teknik Industri Fakultas Teknik Universitas Mercu Buana. Menyelesaikan pendidikan S1 pada Manajemen Agribisnis IPB, S2 pada Manajemen Agribisnis dan melanjutkan S3 pada Doktor Manajemen UMB. Penulis mengajar mata kuliah Teori Probabilitas, Statistik Industri, Manajemen Kualitas. Penulis dapat dihubungi melalui e-mail : [farida@mercubuana.ac.id](mailto:farida@mercubuana.ac.id)

## BIODATA PENULIS



**Dr. Ir. Herry Agung Prabowo, MSc. IPM. CLSSGB**  
Dosen Program Studi Magister Teknik Industri  
Fakultas Teknik Universitas Mercu Buana Jakarta

Dr. Ir. Herry Agung Prabowo, MSc. PhD. IPM. CLSSYB, telah menempuh Pendidikan S1 Teknik Mesin di Universitas Indonesia Depok (1987) kemudian mendapat beasiswa dari Pemerintah RI melalui Kemenristek Dikti pada tahun 1998 untuk menempuh Pendidikan S2 di University of Birmingham, England-UK dan meraih Master of Science (MSc) di bidang Manufacturing Engineering and Management tahun 1999. Pada tahun 2016 mendapat beasiswa dari Universitas Mercu Buana (UMB) Jakarta untuk melanjutkan program S3 di International Islamic University Malaysia, selesai pada tahun 2019 dengan gelar Doctor of Philosophy di bidang Manufacturing Engineering. Meraih Insinyur Profesional Madya (IPM) sejak tahun 2018. Dan meraih sertifikasi untuk Lean Six Sigma Green Belt (CLSSGB) pada tahun 2022. Saat ini menjabat sebagai Kepala ASMARU (Agile and Sustainable Manufacturing Research Center), Ketua Kelompok Bidang Ilmu Sistem Manufaktur dan Ketua Kelompok Riset & Pengabdian Masyarakat Prodi S1 dan S2 Teknik Industri UMB. Sehari-hari mengajar di program studi S1 dan S2 Teknik Industri Universitas Mercu Buana Jakarta untuk mata kuliah Multy Variate Statistics, Lean Manufacturing System Design, Total Productive Maintenance, Green Manufacturing dll. Berbagai penelitian terkait bidang Maintenance Management, TPM, Quality Management dan Lean and Green

Manufacturing secara rutin terus dilakukan. Puluhan artikel telah ditulis dan dipublikasikan dalam skala Nasional maupun International.

Link Google scholar:

<https://scholar.google.com/citations?user=wEx5KokAAAAJ&hl=id>

Penulis dapat dihubungi melalui e-mail: [herryagung@gmail.com](mailto:herryagung@gmail.com)

## BIODATA PENULIS



**Dr. Ir. Agung Wahyudi Biantoro, S.T., M.T.**  
Dosen Program Studi Magister Teknik Sipil  
Fakultas Teknik, Universitas Mercu Buana, Jakarta

Penulis lahir di Jakarta, 29 Oktober 1969. Penulis adalah dosen tetap pada Program Studi Program Studi Magister Teknik Sipil, Fakultas Teknik, Universitas Mercu Buana, Jakarta. Menyelesaikan pendidikan S1 pada Jurusan Teknik Mesin Univ. Yuppentek, S2 pada Jurusan Teknik Mesin Univ. Pancasila dan Teknik Sipil Univ Tamajagakarsa, selanjutnya S3 Teknik Sipil di Univ. Islam Sultan Agung, Semarang. Penulis menekuni bidang Penulis dan Konsultan Konstruksi Bangunan, Bangunan Gedung Hijau, K3 Konstruksi dan Energi Hijau.

Penulis dapat dihubungi melalui e-mail: [agungbntr30@gmail.com](mailto:agungbntr30@gmail.com)

## **BIODATA PENULIS**



**Yani Quarta Mondiana, M.Si**

Dosen Program Studi Kehutanan  
Fakultas Kehutanan Institut Pertanian Malang

Penulis lahir di Malang pada tanggal 11 April 1988. Penulis adalah dosen tetap pada Program Studi Kehutanan fakultas Kehutanan Institut Pertanian Malang. Menyelesaikan pendidikan S1 pada Program Studi statistika Universitas Brawijaya dan melanjutkan S2 pada program studi statistika terapan Institut Pertanian Bogor.

## BIODATA PENULIS



**Dr. Sawarni Hasibuan, M.T., IPU.**

Dosen Program Studi Magister Teknik Industri  
Fakultas Teknik, Universitas Mercu Buana, Jakarta

Sawarni Hasibuan, saat ini adalah Ass. Professor di Departemen Teknik Industri Universitas Mercu Buana, Jakarta. Menyelesaikan pendidikan Sarjana Teknologi Industri Pertanian di Institut Pertanian Bogor (1989), Magister Teknik Industri di Institut Teknologi Bandung (1995), dan memperoleh Doktor Teknologi Industri Pertanian di Institut Pertanian Bogor (2012). Sejak tahun 2017 sebagai Chief in Editor Jurnal Terakreditasi Sinta 3, *Operations Excellence: Journal of Applied Industrial Engineering*. Penulis telah banyak melakukan penelitian di bidang manajemen industri, *risk management*, *green manufacturing*, *renewable energy*, *circular economy*, dan *sustainable supply chain*. Penulis telah beberapa kali memperoleh hibah penelitian dari Kemendikbud yang membahas energi terbarukan seperti biodiesel, bioethanol, biomassa, dan terbaru adalah biojet. Penulis juga telah mempublikasikan hasil-hasil risetnya di berbagai jurnal nasional dan internasional, dengan H-indeks Scopus 8 pada Tahun 2024.

Penulis dapat dihubungi melalui e-mail: [sawarni02@gmail.com](mailto:sawarni02@gmail.com)

## BIODATA PENULIS



### **Daniel Williams Fointuna, S.Pd., M.A.**

**Dosen Program Studi Pendidikan Guru Sekolah Dasar  
Fakultas Keguruan dan Ilmu Pendidikan Universitas Nusa Cendana**

Penulis lahir di Ende tanggal 22 Juli 1997. Penulis adalah dosen tetap pada Program Studi Pendidikan Guru Sekolah Dasar Fakultas Keguruan dan Ilmu Pendidikan Universitas Nusa Cendana. Penulis menyelesaikan pendidikan S-1 pada Jurusan Pendidikan Matematika di Universitas Katolik Widya Mandira Kupang dan melanjutkan S-2 pada Jurusan Pendidikan STEM di Universitas Texas-Austin dengan dukungan penuh dari pemerintah Indonesia lewat LPDP. Penulis menekuni bidang pendidikan matematika dan berfokus pada penelitian kuantitatif. Beberapa karya penulis telah terbit di jurnal internasional maupun nasional. Di waktu luang, penulis suka berwisata ke tempat baru dan menonton drama. Penulis dapat dihubungi melalui e-mail: [daniel\\_fointuna@staf.undana.ac.id](mailto:daniel_fointuna@staf.undana.ac.id).