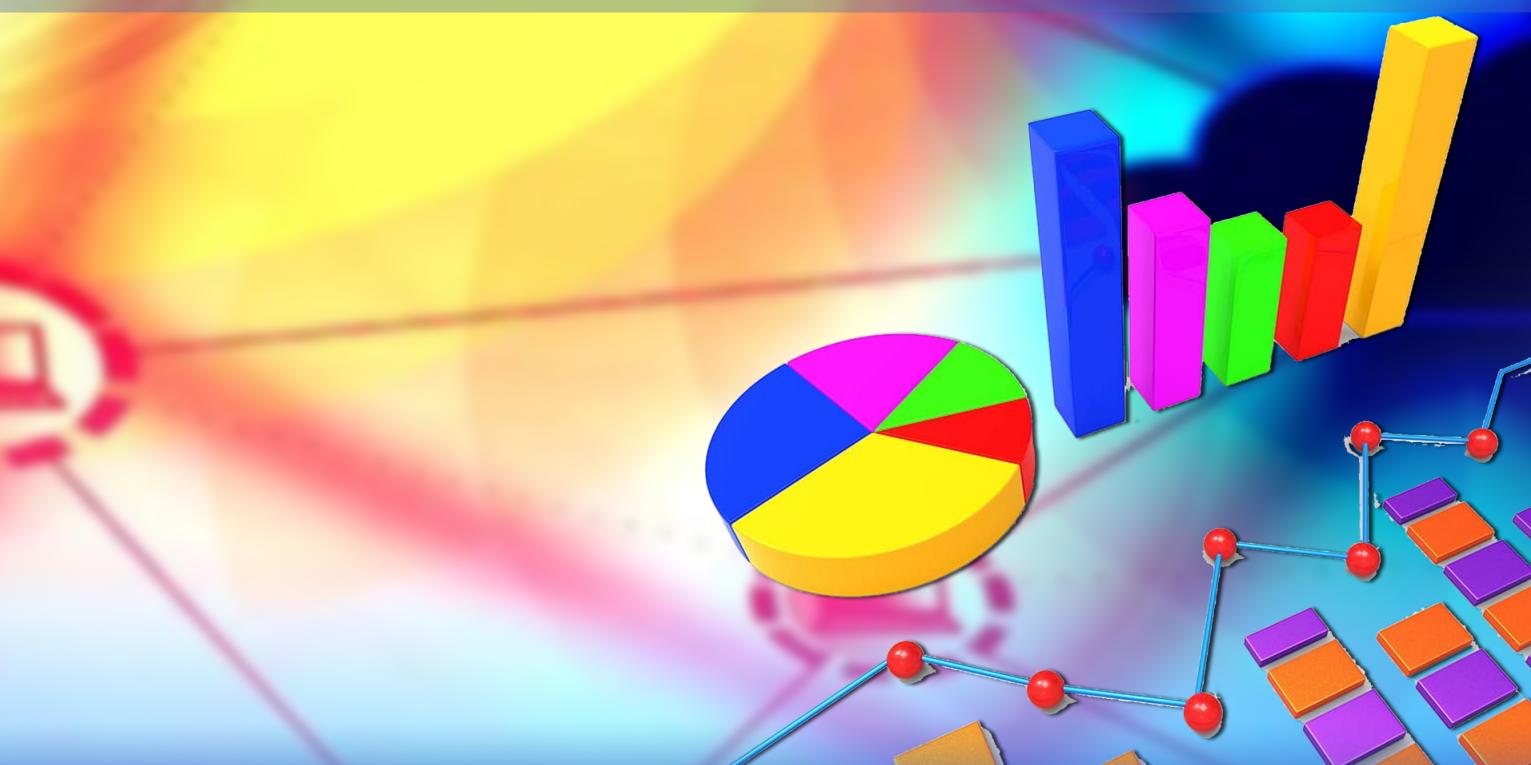




# Statistika

## Bidang Teknologi Informasi



Akbar Iskandar • Muttaqin • Sarini Vita Dewi • Jamaludin • Irawati HM • Cahyo Prianto  
Rosmita Sari Siregar • Muhammad Noor Hasan Siregar • Dina Chamidah  
Marzuki Sinambela • Albinur Limbong • Yusra Fadhillah • Janner Simarmata



# Statistik

## Bidang Teknologi Informasi



## UU 28 tahun 2014 tentang Hak Cipta

### Fungsi dan sifat hak cipta Pasal 4

Hak Cipta sebagaimana dimaksud dalam Pasal 3 huruf a merupakan hak eksklusif yang terdiri atas hak moral dan hak ekonomi.

### Pembatasan Perfilman dan Pasal 26

Ketentuan sebagaimana dimaksud dalam Pasal 23, Pasal 24, dan Pasal 25 tidak berlaku terhadap:

- a. penggunaan kutipan singkat Ciptaan dan/atau produk Hak Terkait untuk pelaporan peristiwa aktual yang ditujukan hanya untuk keperluan penyediaan informasi aktual;
- b. Penggandaan Ciptaan dan/atau produk Hak Terkait hanya untuk kepentingan penelitian ilmu pengetahuan;
- c. Penggandaan Ciptaan dan/atau produk Hak Terkait hanya untuk keperluan pengajaran, kecuali pertunjukan dan Fonogram yang telah dilakukan Pengumuman sebagai bahan ajar; dan
- d. penggunaan untuk kepentingan pendidikan dan pengembangan ilmu pengetahuan yang memungkinkan suatu Ciptaan dan/atau produk Hak Terkait dapat digunakan tanpa izin Pelaku Pertunjukan, Produser Fonogram, atau Lembaga Penyiaran.

### Sanksi Pelanggaran Pasal 113

1. Setiap Orang yang dengan tanpa hak dan/atau tanpa izin Pencipta atau pemegang Hak Cipta melakukan pelanggaran hak ekonomi Pencipta sebagaimana dimaksud dalam Pasal 9 ayat(1) huruf c, huruf d, huruf f, dan/atau huruf h untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 3 (tiga) tahun dan/atau pidana denda paling banyak Rp500.000.000,00 (lima ratus juta rupiah).
2. Setiap Orang yang dengan tanpa hak dan/atau tanpa izin Pencipta atau pemegang Hak Cipta melakukan pelanggaran hak ekonomi Pencipta sebagaimana dimaksud dalam Pasal 9 ayat(1) huruf a, huruf b, huruf e, dan/atau huruf g untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 4 (empat) tahun dan/atau pidana denda paling banyak Rp1.000.000.000,00 (satu miliar rupiah).

# **Statistika Bidang Teknologi Informasi**

Akbar Iskandar, Muttaqin, Sarini Vita Dewi, Jamaludin, Irawati HM Cahyo Prianto, Rosmita Sari Siregar, Muhammad Noor Hasan Siregar

Dina Chamidah, Marzuki Sinambela, Albinur Limpong

Yusra Fadhillah, Janner Simarmata



Penerbit Yayasan Kita Menulis

# **Statistika Bidang Teknologi Informasi**

Copyright © Yayasan Kita Menulis, 2021

Penulis:

Akbar Iskandar, Muttaqin, Sarini Vita Dewi, Jamaludin, Irawati HM  
Cahyo Prianto, Rosmita Sari Siregar, Muhammad Noor Hasan Siregar  
Dina Chamidah, Marzuki Sinambela, Albinur Limpong  
Yusra Fadhillah, Janner Simarmata

Editor: Abdul Karim

Desain Sampul: Devy Dian Pratama, S.Kom.

Penerbit

Yayasan Kita Menulis

Web: [kitamenulis.id](http://kitamenulis.id)

e-mail: [press@kitamenulis.id](mailto:press@kitamenulis.id)

WA: 0821-6453-7176

Anggota IKAPI: 044/SUT/2021

Akbar Iskandar, dkk.

Statistika Bidang Teknologi Informasi

Yayasan Kita Menulis, 2021

xvi; 192 hlm; 16 x 23 cm

ISBN: 978-623-342-047-1 (print)

E-ISBN: 978-623-342-048-8 (online)

Cetakan 1, April 2021

I. Statistika Bidang Teknologi Informasi

II. Yayasan Kita Menulis

## **Katalog Dalam Terbitan**

Hak cipta dilindungi undang-undang

Dilarang memperbanyak maupun mengedarkan buku tanpa

Izin tertulis dari penerbit maupun penulis

# Kata Pengantar

Puji syukur kami haturkan kepada sang pencipta alam semesta karena berkat rahmat dan hidayah-Nya kepada kita semua, sehingga penulisan buku yang berjudul “Statistika Bidang Teknologi Informasi” dapat diselesaikan dengan baik dan berjalan sesuai rencana. Penyusunan buku ini banyak mendapat masukan dari berbagai pihak terutama rekan-rekan kolaborasi dalam penulisan buku ini dengan tujuan memberikan wawasan tentang penerapan statistik dalam bidang informatika/teknologi informasi. Maka pada kesempatan ini kami ingin mengucapkan terima kasih yang setinggi-tingginya kepada semua pihak atas kerelaan dan keikhlasannya dalam proses penyusunan buku ini hingga selesai dan sesuai kaedah-kaedah penulisan karya ilmiah. Penulis juga menyampaikan terima kasih kepada orang tua dan keluarga yang sangat kami cintai, yang selalu memberikan motivasi dan setiap saat berdoa atas setiap langkah kebenaran yang kami lakukan seperti dalam penyusunan buku ini agar kelak bermanfaat untuk kami maupun bagi orang lain.

Akhirnya penulis menyadari bahwa dalam penyusunan buku ini masih banyak terdapat kekurangan, walaupun penulis sudah berusaha secara maksimal dalam mengerjakannya agar menjadi buku yang berkualitas. Saran dan kritik yang sifatnya sangat membangun untuk kesempurnaan dari berbagai pihak sangat kami harapkan jika menemukan kekurangan dalam penyusunan buku ini.

Terima kasih  
Penulis



# Daftar Isi

Kata Pengantar .....	v
Daftar Isi .....	vii
Daftar Gambar .....	xiii
Daftar Tabel.....	xv

## Bab 1 Konsep Dasar Statistika

1.1 Pendahuluan.....	1
1.2 Fungsi dan Tujuan Statistika.....	3
1.3 Syarat Data yang baik dan Langkah Analisisnya.....	4
1.4 Analisis Statistika Deskriptif dan Inferensial .....	5
1.4.1 Tendensi Sentral Data Tunggal.....	5
1.4.2 Mean .....	6
1.4.3 Median .....	7
1.4.4 Modus .....	7
1.4.5 Kuartil, Desil dan Persentil.....	8
1.5 Hubungan Informatika dan Statistika .....	13
1.5.1 Statistik dalam Kecerdasan Buatan (AI) .....	14
1.5.2 Statistik dalam Visualisasi Data.....	14
1.5.3 Statistik dalam Jaringan Saraf .....	14

## Bab 2 Jenis Data dan Skala Pengukuran

2.1 Pendahuluan.....	15
2.2 Pentingnya Memahami Jenis-Jenis Data .....	16
2.3 Skala Pengukuran Data .....	17
2.3.1 Skala Nominal .....	17
2.3.2 Skala Ordinal .....	19
2.3.3 Skala Interval .....	21
2.3.4 Skala Rasio .....	23
2.4 Perubahan Skala .....	25

**Bab 3 Pemeriksaan Data**

3.1 Pendahuluan .....	27
3.2 Data .....	28
3.2.1 Data kualitatif .....	28
3.2.2 Data Kuantitatif .....	28
3.3 Pemeriksaan Data .....	29
3.3.1 Normalitas Data .....	29
3.3.2 Homogen .....	30
3.3.3 Linieritas Data .....	31
3.4 Uji Normalitas .....	32
3.4.1 Uji Grafik .....	32
2.4.2 Metode Chi-Square .....	33
3.4.3 Metode Lilliefors .....	34
3.4.4 Metode Kolmogorov-Smirnov .....	34
3.4.5 Metode Shapiro Wilk .....	35
3.5 Model Pemeriksaan Data .....	36

**Bab 4 Teknik Penyajian Data**

4.1 Pendahuluan .....	39
4.2 Konsep Pengumpulan Data .....	40
4.3 Konsep Penyajian Data .....	41
4.4 Manfaat dan Tujuan Penyajian Data .....	42
4.5 Teknik Penyajian Data .....	43
4.5.1 Narasi atau Teks .....	43
4.5.2 Tabel .....	43
4.5.3 Grafik .....	45
4.6 Penyajian Data Spasial .....	47

**Bab 5 Populasi dan Sampel**

5.1 Pendahuluan .....	49
5.2 Populasi .....	50
5.2.1 Pengertian Populasi .....	50
5.2.2 Jenis- Jenis Populasi .....	52
5.3 Sampel .....	56
5.3.1 Pengertian Sampel .....	56
5.3.2 Teknik Pengambilan Sampel .....	57
5.4 Menentukan Ukuran Sampel .....	61

**Bab 6 Probabilitas**

6.1 Pendahuluan .....	65
6.2 Pendekatan Menentukan Probabilitas .....	66
6.3 Pendekatan Perumusan Probabilitas .....	68
6.3.1 Pendekatan Klasik.....	68
6.3.2 Pendekatan Frekuensi Relatif.....	70
6.3.3 Pendekatan Subjektif .....	72
6.4 Probabilitas Kejadian Majemuk .....	73
6.4.1 Kejadian Saling Lepas .....	74
6.4.2 Kejadian Saling Bebas .....	75
6.5 Probabilitas Bersyarat.....	78

**Bab 7 Data Kelompok dan Analisisnya**

7.1 Pendahuluan .....	81
7.2 Jenis-Jenis Analisis Data .....	82
7.2.1 Data Bermuatan Kualitatif .....	83
7.2.2 Data Bermuatan Kuantitatif.....	83
7.3 Teknik Analisis Data Kelompok .....	83
7.3.1 Distribusi Frekuensi Data Kelompok .....	83
7.3.2 Beberapa Istilah Pada Tabel Distribusi Frekuensi .....	84
7.3.3 Prosedur Umum Membuat Tabel Distribusi Frekuensi.....	85
7.3.4 Macam - Macam Tabel Distribusi Frekuensi.....	87
7.3.5 Pengukuran Nilai Sentral Data Kelompok .....	89
7.3.6 Mean Data Kelompok.....	89
7.3.7 Modus Data Kelompok.....	89
7.3.8 Median Data Kelompok.....	90
7.3.9 Kuartil Data Kelompok.....	91

**Bab 8 Ukuran Pemusatan**

8.1 Pendahuluan .....	93
8.2 Mean (Rata-Rata) .....	93
8.2.1 Mean dari data tunggal .....	94
8.2.2 Mean Aritmatika dari Data Berdaftar Distribusi Frekuensi .....	95
8.2.1 Mean data gabungan .....	99
8.3 Median.....	100
8.3.1 Median dari Data Tunggal.....	100
8.3.2 Median dari Data Berdistribusi Frekuensi.....	101
8.4 Modus.....	102
8.4.1 Modus dari Data Tunggal.....	102

8.4.2 Modus dari Data Berdistribusi Frekuensi .....	103
8.5 Hubungan antara Mean, Modus, dan Median .....	104
8.6 Implementasi Ukuran Pemusatan Menggunakan Python .....	105
8.6.1 Mean.....	105
8.6.2 Median .....	105
8.6.3 Modus .....	106

### Bab 9 Ukuran Penyebaran

9.1 Pendahuluan.....	109
9.2 Konsep Ukuran Penyebaran .....	111
9.3 Tipe Ukuran Penyebaran .....	112
9.4 Fungsi Ukuran Penyebaran.....	113
9.5 Pengukuran Dispersi .....	113
9.6 Jenis Ukuran Variabilitas/Dispersi .....	115
9.7 Range.....	115
9.8 Range Interkuartil .....	118
9.9 Deviasi Kuartil .....	118
9.9.1 Sifat Deviasi Kuartil .....	120
9.9.2 Batasan Penyimpangan Kuartil.....	120
9.10 Deviasi Rata-Rata.....	120
9.11 Standar Deviasi.....	121

### Bab 10 Normalitas dan Homogenitas

10.1 Pendahuluan.....	123
10.2 Normalitas.....	126
10.3 Homogenitas .....	131

### Bab 11 Analisis Korelasi

11.1 Pendahuluan.....	135
11.2 Menentukan Koefisien Korelasi.....	136
11.2.1 Diagram Pencar .....	136
11.2.2 Koefisien Korelasi Pearson .....	138
11.2.3 Koefisien korelasi data berkelompok .....	140
11.2.4 Koefisien Korelasi Peringkat (Rank) Spearman.....	142
11.3 Koefisien determinasi.....	145
11.4 Uji statistik uji Korelasi.....	146

**Bab 12 Analisis Regresi**

12.1 Pendahuluan.....	149
12.2 Jenis Hubungan Antar Variabel .....	150
12.3 Analisis Regresi Sederhana .....	151
12.3.1 Pengertian Regresi Linier Sederhana .....	151
12.3.2 Model Analisis Regresi Linier Sederhana.....	152
12.3.3 Menghitung Regresi Linier dengan Aplikasi.....	153
12.4 Analisis Regresi Linier Berganda .....	156
12.4.1 Pengertian Regresi Linier Berganda.....	156
12.4.2 Model Regresi Linier Berganda.....	157
12.4.3 Menghitung Regresi Linier dengan Aplikasi.....	159

**Bab 13 Aplikasi Statistika**

13.1 Penerapan Statistik .....	163
13.2 Aplikasi Statistik.....	165
13.3 Jenis-Jenis Aplikasi Statistik.....	165

Daftar Pustaka .....	173
Biodata Penulis .....	185

# Daftar Gambar

Gambar 1.1: Histogram data mean, median dan modus .....	5
Gambar 1.2: Frekuensi perkembangan mahasiswa STMIK AKBA .....	11
Gambar 3.1: Linieritas Data (a) Grafik menunjukkan nilai linier. (b) Grafik menunjukkan nilai meningkat secara signifikan .....	32
Gambar 3.2: Grafik Data normal.....	32
Gambar 3.3: Grafik Distribusi Data normal .....	33
Gambar 3.4: Gambar kurva data chi-square .....	33
Gambar 4.1: Contoh diagram grafik (a) .....	46
Gambar 4.2: Contoh diagram grafik (b).....	46
Gambar 4.3: Contoh diagram grafik (c).....	47
Gambar 4.4: Peta sebaran bandara di Indonesia 2008-2009 .....	48
Gambar 8.1: Pola distribusi data (variabel acak kontinu) dengan posisi Mean, Median dan Modus .....	104
Gambar 9.1: Pengukuran Dispersi.....	114
Gambar 10.1: Visual plot grafik dalam uji normalitas .....	126
Gambar 10.2: Shapiro-Wilk Uji Normal .....	128
Gambar 10.3: Tes Kolmogorov-Smirnov .....	129
Gambar 10.4: kolmogrov smirnov normality test .....	129
Gambar 10.5: regression analysis .....	133
Gambar 11.1: Lima kemungkinan korelasi linier untuk data X dan Y .....	137
Gambar 12.1: Hubungan positif dan negatif Antar Variabel dalam Regresi ..	150
Gambar 12.2: Diagram tidak ada hubungan antara Variabel dalam Regresi ..	151
Gambar 12.3: Variabel View .....	154
Gambar 12.4: Data View .....	154
Gambar 12.5: Menu Analyze-Regression-Linear .....	156
Gambar 12.6: Kotak Dialog Linear Regressions.....	156
Gambar 12.7: Output Analisis Regresi .....	157
Gambar 12.8: Variabel View .....	160
Gambar 12.9: Data View .....	161
Gambar 12.10: Menu Analyze-Regression-Linear .....	161
Gambar 12.11: Kotak Dialog Linear Regressions .....	162
Gambar 12.12: Output Analisis Regresi Linier Berganda.....	162

Gambar 13.1: SPSS .....	166
Gambar 13.2: Logo Minitab .....	166
Gambar 13.3: Logo Microsoft Access .....	167
Gambar 13.4: Logo LISREL .....	168
Gambar 13.5: Logo SAS.....	169
Gambar 13.6: Logo Eviews .....	169
Gambar 13.7: Logo STATA.....	170
Gambar 13.8: Logo AMOS .....	170
Gambar 13.9: R-Software .....	171
Gambar 13.10 Logo SOFA.....	172

# Daftar Tabel

Tabel 1.1: Contoh hasil ujian mata kuliah statistika.....	6
Tabel 1.2: Contoh letak modus.....	8
Tabel 1.3: Contoh analisis varians.....	11
Tabel 1.4: Contoh analisis standar deviasi.....	12
Tabel 2.1: Berbagai Skala Pengukuran Variabel dengan Ciri yang Dipunyai.	26
Tabel 3.1: Format Data liliefors.....	34
Tabel 3.2: Kolmogorov-Smirnov .....	35
Tabel 4.1: Jumlah penduduk putus sekolah untuk SD/MI di Desa X Tahun 2007 – 2009 .....	44
Tabel 7.1: Tabel Penilaian.....	82
Tabel 7.2: Tinggi 100 siswa SMA XYZ.....	84
Tabel 7.3: Penilaian siswa .....	86
Tabel 7.4: Frekuensi Penilaian.....	86
Tabel 7.5: Distribusi Frekuensi Relatif .....	87
Tabel 7.6: Frekuensi Kumulatif.....	88
Tabel 7.7: Perhitungan Frekuensi Kumulatif.....	88
Tabel 8.1: Frekuensi dan nilai akhir mahasiswa.....	95
Tabel 8.2: Data jumlah transaksi per hari selama 60 hari .....	96
Tabel 8.3: Tabulasi distribusi frekuensi .....	97
Tabel 8.4: Distribusi frekuensi dan nilai tengah .....	98
Tabel 8.5: Rata rata transaksi pada masing masing kelas yang diperoleh....	98
Tabel 8.6: Distribusi nilai 40 mahasiswa dalam suatu mata kuliah: .....	101
Tabel 9.1: Tipe Ukuran Penyebaran .....	112
Tabel 11.1: Arti magnitudo koefisien korelasi .....	140
Tabel 11.2: Hubungan besar koefisien korelasi dan determinasi (Holcomb, 1998).....	145
Tabel 12.1: Data Biaya dan Jumlah Produksi.....	153
Tabel 12.2: Daftar Nilai Statistika Mahasiswa .....	159



# **Bab 1**

## **Konsep Dasar Statistika**

### **1.1 Pendahuluan**

Statistika adalah sebuah cabang matematika yang berhubungan dengan ilmu melakukan pengumpulan data, analisis, deskripsi, penyajian, interpretasi dan penarikan kesimpulan dari data kuantitatif untuk memahami data yang dikumpulkan (Wild, Utts and Horton, 2018), (Lindgren, 2017). Sedangkan statistik adalah hasil analisis data, interpretasi dan penyajiannya. Dengan kata lain, beberapa komputasi telah dilakukan yang memberikan pemahaman tentang arti data meskipun tidak harus selalu disajikan dalam bentuk tabel, bagan, atau grafik. Statistik sering kali dilaporkan oleh lembaga pemerintah misalnya statistik pengangguran, pertumbuhan penduduk, pertumbuhan ekonomi atau statistik melek teknologi (Jarmin, 2019).

Statistika salah satu disiplin ilmu yang paling penting untuk menyediakan alat dan metode untuk menemukan struktur dan memberikan wawasan yang lebih dalam tentang data dan disiplin yang paling penting untuk menganalisis dan mengukur ketidakpastian. Maka dari itu, metode statistik diterapkan di semua bidang yang melibatkan pengambilan keputusan untuk membuat kesimpulan yang akurat dari kumpulan data, sehingga menghasilkan beberapa informasi yang berarti tentang data itu atau membuat keputusan dalam menghadapi ketidakpastian berdasarkan metodologi statistik (Sethna, 2021).

Data merupakan fakta dan gambar yang dikumpulkan, dianalisis, dan diringkas untuk presentasi dan interpretasi. Data dapat diklasifikasikan sebagai kuantitatif atau kualitatif. Data kuantitatif mengukur seberapa banyak atau seberapa banyak sesuatu, dan data kualitatif memberikan label, atau nama, untuk kategori barang serupa. Misalnya karakteristik seperti usia, jenis kelamin, status perkawinan, dan pendapatan tahunan.

Statistik digunakan di hampir semua disiplin ilmu seperti ilmu fisik dan sosial, serta dalam bisnis, humaniora, pemerintahan, manufaktur dan informatika. Dalam praktiknya, statistik adalah gagasan yang dapat kita pelajari tentang sifat-sifat kumpulan besar objek atau peristiwa (suatu populasi) dengan mempelajari karakteristik sejumlah kecil objek atau peristiwa serupa (sampel). Karena dalam banyak kasus pengumpulan data komprehensif tentang seluruh populasi terlalu mahal, sulit, atau sama sekali tidak mungkin, statistik dimulai dengan sampel yang dapat diamati dengan mudah atau terjangkau (Singpurwala, 2018).

Analisis data adalah proses pemeriksaan, pembersihan, transformasi dan pemodelan data dengan tujuan menemukan informasi yang berguna, menginformasikan kesimpulan dan mendukung pengambilan keputusan. Analisis data memiliki banyak aspek dan pendekatan, mencakup beragam teknik dengan berbagai nama dan digunakan dalam domain bisnis, sains, dan ilmu sosial yang berbeda. Dalam dunia bisnis saat ini analisis data berperan dalam membuat keputusan yang lebih ilmiah dan membantu bisnis beroperasi lebih efektif (Jarmin, 2019).

Penggunaan komputer modern telah mempercepat penghitungan statistik skala besar dan juga memungkinkan metode baru yang tidak praktis untuk dilakukan secara manual. Statistika terus menjadi bidang kajian penelitian yang aktif misalnya tentang masalah bagaimana menganalisis data besar (Big data). Saat ini kebutuhan untuk mengolah data dengan jumlah besar tersedia di banyak bidang terapan menjadi informasi yang berguna dan mendorong perkembangan teoritis maupun praktis.

## 1.2 Fungsi dan Tujuan Statistika

Statistika merupakan disiplin yang menyangkut pengumpulan, organisasi, analisis, interpretasi, dan penyajian data. Maka fungsi statistika secara umum yaitu:

1. Statistika membantu dalam memberikan pemahaman yang lebih baik dan deskripsi yang akurat tentang fenomena alam.
2. Statistika membantu dalam perencanaan penyelidikan statistik yang tepat dan efisien dalam bidang studi apa pun.
3. Statistika membantu dalam mengumpulkan data kuantitatif yang sesuai.
4. Statistika membantu dalam menyajikan data kompleks dalam bentuk tabel, diagram dan grafik yang sesuai untuk pemahaman data yang mudah dan jelas.
5. Statistika membantu dalam memahami sifat dan pola variabilitas suatu fenomena melalui observasi kuantitatif.
6. Statistika membantu dalam menarik kesimpulan yang valid, bersama dengan ukuran keandalannya tentang parameter populasi dari data sampel.
7. Statistik membantu dalam merumuskan dan menguji hipotesis atau teori baru.
8. Statistik dapat membantu kita dalam melakukan analisis korelasi sekaligus sekaligus membandingkan antar kelompok data.
9. Statistika dapat digunakan untuk memprediksi tindakan fenomena di masa depan.
10. Statistika memungkinkan seseorang untuk memperbesar cakrawala atau meningkatkan pengetahuan.

**Tujuan analisis statistika sebagai berikut:**

1. Menentukan jenis dan jumlah data yang perlu dikumpulkan.
2. Mengatur dan meringkas data.
3. Menganalisis data dan menarik kesimpulan darinya.
4. Menilai kekuatan kesimpulan dan mengevaluasi ketidakpastiannya.

5. Mendeskripsikan kumpulan data populasi sesuai tujuan analisis

## 1.3 Syarat Data yang baik dan Langkah Analisisnya

Kualitas data mengacu pada keadaan informasi kualitatif atau kuantitatif. Ada banyak definisi tentang kualitas data, tetapi data umumnya dianggap berkualitas jika "sesuai untuk tujuan penggunaan dalam operasi, pengambilan keputusan dan perencanaan".

Maka syarat-syarat data yang baik yaitu:

1. Harus obyektif (sesuai dengan keadaan sebenarnya)
2. Harus bisa mewakili seluruh data (representatif)
3. Harus memiliki tingkat ketelitian yang tinggi (standar error harus kecil)
4. Harus tepat waktu (up to date)
5. Harus relevan

### **Langkah-langkah analisis data statistika:**

Tujuan statistik adalah menarik kesimpulan dari data. Setiap analisis data melibatkan langkah-langkah berikut:

1. Persiapan (persiapkan semua data yang sudah terkumpul, cek kelengkapan instrument yang akan digunakan dalam mengumpulkan data).
2. Merumuskan masalah penelitian.
3. Mendefinisikan populasi dan sampel.
4. Melakukan analisis data deskriptif.
5. Menggunakan metode statistik yang tepat untuk menyelesaikan masalah penelitian.
6. Melaporkan hasil.

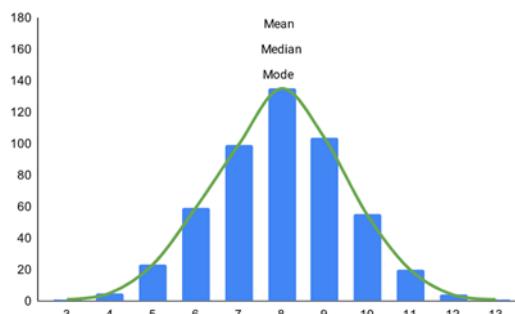
## 1.4 Analisis Statistika Deskriptif dan Inferensial

Statistika memainkan peran utama dalam bidang penelitian serta membantu kita dalam pengumpulan, analisis dan penyajian data empiris. Terdapat dua jenis metode statistik yang digunakan dalam menganalisis data yaitu statistik deskriptif dan statistik inferensial. Statistik deskriptif menggambarkan karakteristik penting dari data dengan menggunakan ukuran tendensi sentral seperti mean, median, modus dan ukuran dispersi seperti rentang, frekuensi, standar deviasi, varians, dll. Ukuran penyebaran seringkali ditampilkan secara visual dalam tabel, diagram lingkaran, diagram batang dan histogram untuk membantu dalam memahami tren dalam data (Rayat, 2018).

Singkatnya, statistik deskriptif membantu mendeskripsikan dan memahami fitur dari kumpulan data tertentu dengan memberikan ringkasan singkat tentang sampel dan ukuran data. Statistik deskriptif hanyalah representasi dari data yang tersedia dan tidak berdasarkan pada teori probabilitas apa pun. Jenis statistik deskriptif yang paling dikenal adalah ukuran pemusatan: mean, median dan modus yang digunakan di hampir semua tingkat matematika dan statistik.

### 1.4.1 Tendensi Sentral Data Tunggal

Ukuran tendensi sentral adalah nilai tunggal yang mencoba mendeskripsikan sekumpulan data dengan mengidentifikasi posisi sentral dalam kumpulan data tersebut. Ini bertujuan untuk memberikan deskripsi yang akurat dari seluruh data.



Gambar 1.1: Histogram data mean, median dan modus

### 1.4.2 Mean

Mean atau rerata dihitung dengan cara menjumlahkan semua angka dalam kumpulan data, kemudian dibagi dengan jumlah seluruh angka dalam kumpulan data tersebut (disebut sebagai n).

Rumus Mean

$$\underline{x} = \frac{x_1 + x_2 + \dots + x_n}{n} \quad \text{atau} \quad \underline{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Keterangan:

$\sum x_i$  = Jumlah seluruh harga x

n = Banyak data

Contoh:

Jumlah penerimaan mahasiswa baru STMIK AKBA dalam satu tahun (TA. 2020/2021) adalah 1000 orang dengan sebaran data setiap bulan yaitu (50, 30, 40, 50, 60, 70, 50, 60, 75, 74, 60, 50, 46, 55, 35, 45, 50, 50, 50).

Maka rerata dari data tersebut adalah 52,63 (1000/19).

Hasil ujian statistika pada mahasiswa Universitas Teknologi Akba Makassar (UTAMA) pada kelas teknik informatika yaitu:

**Tabel 1.1:** Contoh hasil ujian mata kuliah statistika

No	Nama	Skor hasil tes
1.	Awaluddin	85
2.	Akbar Iskandar	70
3.	Rahmat	80
4.	Anwar	87
5.	Sappa	50
6.	Lippu	90
7.	Saripa	85
8.	Palili	75

Maka cara untuk menghitung mean atau rerata dapat dilakukan dengan cara menjumlahkan seluruh data dibagi dengan jumlah data:  $85+70+80+87+50+90+85+75/8$  (622/8) maka mean atau rerata hasil perhitungan tersebut adalah 77.75.

### 1.4.3 Median

Median adalah titik yang membagi seluruh data menjadi dua bagian yang sama besar atau suatu nilai yang membagi 50% frekuensi bagian atas dan 50% frekuensi bagian bawah. Maka untuk menentukan nilai median dalam urutan angka, terlebih dahulu angka-angka tersebut diurutkan atau diatur dalam urutan nilai dari angka terendah ke tertinggi atau tertinggi ke terendah. Jika jumlah datanya ganjil, maka letak dan nilai mediannya adalah angka yang berada di tengah. Namun jika jumlah data genap, maka harus menentukan pasangan tengah terlebih dahulu, kemudian menjumlahkannya dan membaginya dengan dua.

Contoh:

Hasil pengamatan diperoleh kumpulan data {50, 30, 40, 50, 60, 70, 50, 60, 75, 74, 60, 50, 46, 55, 35, 45, 50, 50, 50}. Dari data tersebut selanjutnya diurutkan menjadi {30, 35, 40, 45, 46, 50, 50, 50, 50, 50, 50, 55, 60, 60, 60, 70, 74, 75}. Maka median dari kelompok data tersebut yaitu angka yang berada di tengah {30, 35, 40, 45, 46, 50, 50, 50, 50, 50, 50, 55, 60, 60, 60, 70, 74, 75} dalam contoh ini adalah 50 karena ada sembilan angka di kedua sisinya.

Untuk mencari nilai median dengan jumlah angka genap, terlebih dahulu seseorang harus menentukan pasangan nilai tengah, menjumlahkannya, dan membaginya dengan dua. Sekali lagi, susunlah angka-angkanya dari yang terendah hingga tertinggi.

Contoh:

Data berat badan 6 mahasiswa STMIK AKBA yaitu {70 kg, 69 kg, 45 kg, 80 kg, 56 kg, 50 kg}. Kemudian diurutkan menjadi {45 kg, 50 kg, 56 kg, 69 kg, 70 kg, 80 kg}. Maka mediannya adalah rata-rata dari dua angka di tengah yaitu {45 kg, 50 kg, 56 kg, 69 kg, 70 kg, 80 kg} data ke 3 dan ke 4  $(56+69) \div 2 = 62,5$  kg.

### 1.4.4 Modus

Modus adalah titik data yang paling sering muncul dalam kumpulan data atau mempunyai frekuensi terbesar dari sekelompok data. Modus ini berguna jika ada banyak nilai berulang dalam kumpulan data. Sebuah data dapat memiliki satu atau lebih dari satu mode. Jika hanya ada satu nomor yang muncul beberapa kali dalam kelompok data, berarti datanya memiliki satu mode disebut Uni-modal. Jika terdapat dua angka yang sering muncul dalam

kelompok data dan jumlah kemunculan maksimum sama, maka data data tersebut memiliki dua mode disebut Bi-modal. Lebih dari dua modus disebut Multi-modal sedangkan jika tidak ada modus disebut dengan Non-modal.

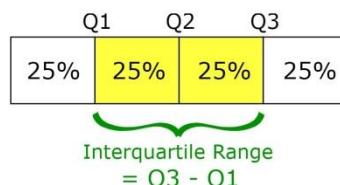
**Tabel 1.2:** Contoh letak modus

<b>X<sub>i</sub></b>	<b>f<sub>i</sub></b>	<b>Letak modus</b>
12	4	Berdasarkan data
14	2	disamping maka letak
28	2	modus berada pada
34	1	frekuensi terbesar yaitu
		frekuensi 4 dan Nilai
		modusnya adalah <b>12</b> .

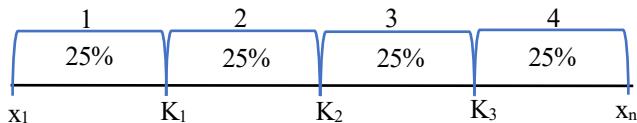
#### 1.4.5 Kuartil, Desil dan Persentil

Saat kita membagi data menjadi persentase kita biasanya berbicara tentang kuartil, desil dan persentil. Kuartil membagi kelompok data menjadi empat (4) bagian bagian yang sama. Desil dan persentil biasanya diterapkan pada kumpulan data besar dan proses analisis desil mirip dengan cara analisis persentil. Desil membagi kumpulan data menjadi sepuluh (10) bagian yang sama sedangkan persentil membagi data menjadi seratus (100) bagian (Kaur et al., 2018).

Penjelasan pertama kita akan membahas tentang kuartil, di mana kuartil digunakan untuk meringkas sekelompok angka menjadi 4 bagian yang sama banyak sesudah disusun menurut urutan nilainya, maka bilangan pembaginya disebut kuartil yaitu Q1, Q2, Q3 dan untuk memahami kuartil, penting untuk memahami median sebagai ukuran tendensi sentral (Selvamuthu and Das, 2018). Kuartil bekerja seperti median membagi data menjadi dua bagian sehingga 50% pengukuran berada di bawah median dan 50% berada di atasnya. Sedangkan kuartil memecah data menjadi beberapa kuartal sehingga 25% pengukuran lebih kecil dari kuartil bawah (Q1), 50 % kurang dari rata-rata, dan 75% kurang dari kuartil atas. Ilustrasi dapat dilihat seperti berikut:



atau



Tahapan menghitung kuartil, desil dan persentil

1. Susun data menurut urutan nilainya
2. Tentukan letak kuartil, desil dan persentil
3. Tentukan nilai kuartil, desil dan persentil

Rumus kuartil:  $K_i = \frac{i(n+1)}{4}$

Keterangan:

$n$  = jumlah data

$i = 1, 2$  dan  $3$

$K_i$  = kuartil ke- $i$

Contoh:

Tentukan nilai kuartil 1 dan kuartil 3 dari data berikut!

Data: 75, 82, 66, 57, 64, 56, 92, 94, 86, 52, 60, 70

Jawaban:

Kuartil 1

Urutan: 52, 56, 57, 60, 64, 66, 70, 75, 82, 86, 92, 94

$$K_i = \frac{1(12 + 1)}{4}$$

Letak  $K_i = 3,25$

Jadi nilai  $K_1 = \text{data ke tiga} + 0,25$  (data keempat – data ke tiga).

Maka nilai  $K_1 = 57 + 0,25$  ( $60 - 57$ ) = 57,75

Kuartil 2

$$K_i = \frac{2(12 + 1)}{4}$$

Letak  $K_i = 6,5$

Jadi nilai K2 = data ke enam + 0,5 (data ketujuh – data ke enam).

Maka nilai K2 = 66+0,5 (70-66)= 68

Setelah kuartil selanjutnya kita akan bahas tentang desil dan persentil, dan desil membagi himpunan data (n) menjadi sepuluh (10) bagian yang sama, sehingga terdapat 9 nilai desil yaitu D1, D2, D3, ..., D9. Desil pertama adalah D1. Sedangkan persentil membagi data menjadi seratus bagian seperti yang dijelaskan sebelumnya, sehingga diperoleh 99 nilai persentil yaitu P1, P2.....P99. Tahapan analisis desil dan persentil sama seperti proses analisis kuartil yaitu:

Rumus Desil:

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

Keterangan:

n = jumlah data

i = 1, 2 ..... dan 9

Di = Desil ke-i

Rumus Persentil:

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

Keterangan:

n = jumlah data

i = 1, 2 ..... dan 100

Pi = Persentil ke-i

Ukuran variasi (dispersi) data tunggal

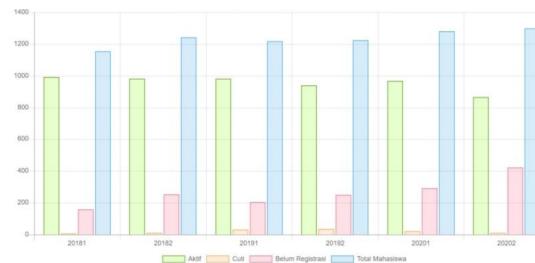
Dispersi dalam statistik adalah cara untuk mendeskripsikan seberapa tersebar suatu kumpulan data. Penyebaran kumpulan data dapat dijelaskan dengan berbagai statistik deskriptif termasuk rentang, frekuensi, standar deviasi dan varians.

Jangkauan (rentang): Selisih antara nilai terbesar dan terkecil

$$r = X_n - X_1$$

$$r = \text{Nilai maksimum} - \text{Nilai minimum}$$

Distribusi Frekuensi: Distribusi frekuensi biasanya disajikan dalam bentuk tabel atau grafik.

**Gambar 1.2:** Frekuensi perkembangan mahasiswa STMIK AKBA

Varians: Tingkat varian berkorelasi dengan ukuran rentang angka secara keseluruhan, artinya varian lebih besar bila ada rentang angka yang lebih luas dalam kelompok data dan varian lebih kecil jika ada rentang angka yang lebih sempit. Kemudian, salah satu teknik statistik yang digunakan untuk menjelaskan homogenitas kelompok data adalah dengan varians. Di mana varians menggabungkan semua nilai dalam kumpulan data untuk menghasilkan ukuran penyebaran. Varians untuk sampel (dilambangkan dengan  $S^2$ ) dan standar deviasi (akar kuadrat dari varians, dilambangkan dengan  $S$ ), sedangkan simbol varians untuk populasi  $\sigma^2$ .

Rumus varians

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

$$\sigma^2 = \frac{\sum (x - \bar{x})^2}{f-1} \text{ untuk populasi}$$

**Tabel 1.3:** Contoh analisis varians

$X$ (kg)	$(X - \bar{x})$	$(X - \bar{x})^2$
48	9	81
52	5	25
56	1	1
62	5	25
67	10	100
$\Sigma X = 285$		$\Sigma (X - \bar{x})^2 = 232$
$\bar{x} = \frac{\sum X}{n} = \frac{285}{5} = 57$	$s^2 = \frac{\sum (x - \bar{x})^2}{n-1} = \frac{232}{5-1} = 58$	

Standar deviasi: Ukuran penyebaran data dalam statistik atau seberapa banyak data anda tersebar di sekitar mean.

**Tabel 1.4:** Contoh analisis standar deviasi

X (kg)	(X - $\bar{x}$ )	(X - $\bar{x}$ ) <sup>2</sup>
48	9	81
52	5	25
56	1	1
62	5	25
67	10	100
$\Sigma X = 285$		$(\Sigma (X - \bar{x}))^2 = 232$
$\bar{x} = \frac{\sum X}{n} = \frac{285}{5} = 57$		$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}} = \sqrt{\frac{232}{5-1}} = \sqrt{58} = 7,6$

Setelah analisis statistik secara deskriptif di atas, selanjutnya statistik inferensial yang terbagi atas parametric dan nonparametric. Statistic parametric mensyaratkan data harus berdistribusi normal sedangkan nonparametric data tidak harus berdistribusi normal (Gibbons and Chakraborti, 2020), (Amrhein, Trafimow and Greenland, 2018). Tujuan analisis statistic infrensial adalah untuk membantu kita dalam melakukan analisis keputusan, prediksi atau generalisasi tentang suatu populasi, berdasarkan sampel.

Selanjutnya, statistik inferensial digunakan untuk membuat generalisasi tentang kelompok besar, seperti memperkirakan permintaan rata-rata untuk suatu produk dengan mensurvei sampel kebiasaan membeli konsumen atau untuk mencoba memprediksi peristiwa masa depan.

Maka statistik inferensial mencakup:

1. Probabilitas atau teori kemungkinan
2. Distribusi teoritis
3. Sampling dan distribusi sampling
4. Pendugaan populasi atau teori populasi
5. Uji hipotesis

## 1.5 Hubungan Informatika dan Statistika

Statistik dalam ilmu komputer atau informatika digunakan untuk beberapa hal, termasuk penggalian data, kompresi data, dan pengenalan suara secara ilmiah. Area lain di mana statistik digunakan dalam ilmu komputer memainkan peran penting untuk memberikan kontribusi yang signifikan bidang rekayasa perangkat lunak termasuk analisis gambar, kecerdasan buatan, visualisasi data, pengenalan pola, data mining, bioinformatika, dll. Teknik statistik tidak hanya membantu membuat model ilmiah tetapi juga mengukur keandalan, produktivitas dan ketidakpastian. Seiring perkembangan teknologi komputer saat ini, kontribusi statistika dan ilmu komputer semakin penting karena statistik dan ilmu komputer sama-sama fokus tentang data (Oussous et al., 2018). Sehingga menjadi media yang digunakan oleh banyak orang untuk menyelesaikan permasalahan, seperti permasalahan bidang manajemen, penelitian, bisnis, pemasaran, kualitas operasi, peramalan dan analisis risiko.

Dua puluh tahun yang lalu sebagian besar data dikumpulkan secara manual dan set data dilakukan dalam bentuk sederhana. Tapi saat ini telah terjadi banyak perubahan sifat data dan aplikasi komputer dapat dimanfaatkan untuk memperoleh informasi yang maksimal dengan mengurangi sedikit waktu dan biaya dalam pengumpulan data. Maka dengan hadirnya teknologi, sejumlah data besar secara otomatis direkam dengan komputer dan dikelola dengan sistem manajemen basis data untuk menyimpan dan mengambil informasi dengan cepat.

Teknologi saat ini telah memungkinkan ekstraksi otomatis informasi prediktif yang tersembunyi dari basis data. Dengan data mining, seseorang menerapkan berbagai metode statistik, analisis data dan mesin cerdas untuk mengeksplorasi dan menganalisis kumpulan data yang besar untuk mengekstrak informasi baru serta menguntungkan pemilik data (Ratner, 2017). Dengan menggunakan data mining, organisasi dapat menemukan wawasan yang dapat ditindaklanjuti dari data mereka yang ada. Misalnya, dengan menganalisis postingan media sosial, perusahaan makanan ringan mungkin terkejut mengetahui bahwa pasar terbesar mereka adalah anak kecil. Algoritma telah menjadi kebutuhan dalam banyak aspek pemrograman komputer dan penggalian data (Weihs and Ickstadt, 2018).

### **1.5.1 Statistik dalam Kecerdasan Buatan (AI)**

Kecerdasan buatan (AI) secara intrinsik digerakkan oleh data dan ini membutuhkan penerapan konsep statistik melalui kolaborasi manusia dan mesin selama pembuatan data, pengembangan algoritma, dan evaluasi hasil. Statistik umumnya dianggap sebagai prasyarat untuk bidang pembelajaran kecerdasan buatan karena AI dan statistik adalah dua bidang studi yang memiliki hubungan erat. Analisis data AI mampu melampaui analisis diagnostik sederhana dari data kuantitatif dan menangani data kualitatif untuk analisis diagnostik, prediktif, dan preskriptif (Brynjolfsson, Rock and Syverson, 2018). Sistem kecerdasan bisnis yang didukung AI memungkinkan kita untuk mengetahui mengapa sesuatu berhasil atau mungkin tidak berhasil.

### **1.5.2 Statistik dalam Visualisasi Data**

Visualisasi data terkait dengan semua kebutuhan bisnis dan merupakan salah satu aspek terpenting dari fungsi bisnis. Dengan meningkatnya analisis data, sebagian besar data sekarang dikumpulkan dan diproses oleh komputer. Jika anda ingin menyajikan data ini, maka anda memerlukan beberapa visualisasi, dan jenis alat ini membutuhkan pengetahuan statistic (McCormick and Salcedo, 2017). Dari industri seperti perawatan kesehatan hingga mobil, visualisasi data menjadi semakin penting. Jadi, di masa depan programmer mungkin harus melakukan pekerjaan ganda karena dia berfungsi pengembang dan berinteraksi dengan manajer data serta membuat keputusan (Friedman, 1998).

### **1.5.3 Statistik dalam Jaringan Saraf**

Aplikasi statistika dalam jaringan saraf tiruan bekerja mirip dengan jaringan saraf otak manusia. Sebuah "neuron" dalam jaringan saraf adalah fungsi matematika yang mengumpulkan dan mengklasifikasikan informasi menurut arsitektur tertentu. Jaringan memiliki kemiripan yang kuat dengan metode statistik seperti penyesuaian kurva dan analisis regresi. Jaringan saraf tiruan digunakan dalam berbagai aplikasi di layanan keuangan, mulai dari peramalan dan riset pemasaran hingga deteksi penipuan dan penilaian risiko. Penggunaan statistik dalam pemrograman meningkatkan pemahaman tentang sistem yang mendasari yang mengarah ke hasil yang lebih bermakna. Statistik dalam rekayasa perangkat lunak mengarah pada penentuan kualitas dan kinerja optimal yang lebih konklusif (Miller, 2017).

## **Bab 2**

# **Jenis Data dan Skala Pengukuran**

### **2.1 Pendahuluan**

Data adalah kumpulan informasi atau fakta yang belum diolah, biasanya berbentuk simbol, angka, kata-kata, citra atau lain sebagainya yang didapat dari hasil pengamatan atau sumber-sumber tertentu. Di era modern saat ini, data diibaratkan seperti tumpukan harta karun yang harganya semahal emas. Pentingnya sebuah data memicu banyak perusahaan mulai mencari tenaga ahli di bidang data sehingga lowongan profesional di bidang data semakin meningkat. Fenomena ini menyebabkan banyak orang mulai tertarik dan belajar data, salah satunya dengan mulai mengikuti kursus belajar data (Kurniawati, 2020).

Dalam bidang penelitian penelitian, kita telah mengenal macam-macam teknik pengumpulan data yang bisa kita gunakan, baik teknik pengumpulan data kuantitatif maupun kualitatif. Selain teknik pengumpulan data yang baik, jenis data dan skala pengukuran yang tepat untuk mengolah data tersebut juga diperlukan dalam sebuah penelitian supaya hasilnya objektif serta memiliki kredibilitas yang tinggi. Lalu, apakah jenis-jenis dari skala pengukuran yang

biasa digunakan untuk mendapatkan nilai statistik dari suatu data?(Sharon, 2020).

Skala pengukuran data dalam penelitian pada dasarnya dimaksudkan untuk mengklasifikasikan Variabel yang akan diukur agar tidak terjadi kesalahan dalam menentukan teknik analisis data dan tahapan penelitian selanjutnya. Skala pengukuran data merupakan seperangkat aturan yang diperlukan untuk ‘mengkuantitatifkan’ data dari pengukuran suatu variabel. Dalam melakukan analisis statistik, perbedaan jenis data sangat berpengaruh terhadap pemilihan model atau alat uji statistik. Tidak sembarang jenis data dapat digunakan oleh alat uji tertentu. Untuk itu skala pengukuran data (variabel) sangat menentukan dalam uji statistik (Rahman, 2020).

## 2.2 Pentingnya Memahami Jenis-Jenis Data

Di dunia ini, jenis data sangat beragam mulai dari data terstruktur hingga tidak terstruktur, data berbentuk angka, simbol, huruf, kata, gambar, bahkan ada data berbentuk audio. Pemahaman mengenai jenis data akan membantu proses pemilihan metode yang tepat untuk penelitian. Misalnya, data kontinu dan data kategorik dianalisis menggunakan metode yang berbeda. Pemahaman mengenai jenis data juga akan membantu kita untuk mengoptimalkan proses perhitungan. Selain itu, kita juga lebih mudah untuk menuliskan coding pada software analisis data sehingga informasi yang didapat lebih optimal dan mudah dipahami. Tidak semua data dapat divisualisasikan, misalnya data gambar tidak bisa divisualisasikan menjadi diagram batang, Oleh karena itu, pemahaman jenis-jenis data merupakan ilmu dasar yang harus kita kuasai sebelum menganalisis data.

Pengukuran merupakan proses untuk menilai atau membedakan objek yang akan diukur. Setiap operasi matematika atau metode analisis data memiliki syarat khusus pada skala pengukuran datanya. Ketidaksesuaian antara metode dan skala pengukuran juga akan berpengaruh pada kesimpulan. Hal ini dapat menyebabkan adanya kesimpulan yang bias atau kurang tepat (Kurniawati, 2020).

## 2.3 Skala Pengukuran Data

Dalam ilmu statistik, ada empat jenis pengukuran data, yaitu nominal, ordinal, interval, dan rasio. Skala-skala tersebut diperlukan terutama dalam pengukuran untuk menilai dan membedakan objek yang diteliti.

### 2.3.1 Skala Nominal

Apa itu skala nominal? Skala nominal adalah skala yang hanya bisa membedakan benda, peristiwa, ataupun objek yang diteliti antara satu dengan yang lainnya berdasarkan nama atau predikatnya. Skala nominal ini umumnya digunakan untuk mengklasifikasi sebuah objek, individu, atau kelompok secara kategorik atau kualitatif. Umumnya, pengukuran dengan skala ini dilakukan dengan pemberian angka atau simbol pada objek tersebut. Pemberian angka dan simbol ini tidak memiliki maksud kuantitatif serta hanya menunjukkan adanya atribut atau karakteristik pada objek tersebut. Hal ini berarti skala ini hanya diberikan sebagai label saja serta tidak mengandung pengertian tingkatan tertentu (Sharon, 2020).

Menurut Kurniawati (2020) Skala nominal merupakan skala yang paling rendah dibandingkan skala pengukuran yang lain. Skala ini hanya dapat digunakan sebagai label atau predikatnya saja. Pemberian angka atau simbol pada skala ini hanya berupa label dan tidak dapat merepresentasikan nilai kuantitatifnya. Sedangkan menurut Rahman (2020) Skala nominal adalah skala yang hanya mendasarkan pada pengelompokan atau pengkategorian peristiwa atau fakta dan apabila menggunakan notasi angka hal itu sama sekali tidak menunjukkan perbedaan kuantitatif melainkan hanya menunjukkan perbedaan kualitatif.

Contoh skala nominal adalah label laki-laki dan perempuan pada jenis kelamin seseorang. Skala ini tidak dapat digunakan untuk operasi matematika seperti penjumlahan, pengurangan, dan lain sebagainya namun, dapat divisualkan menjadi diagram batang, diagram lingkaran, dan lain sebagainya (Kurniawati, 2020). Banyak variabel dalam penelitian sosial menggunakan skala nominal seperti agama, jenis kelamin, tempat lahir, asal sekolah, dsb.

Untuk itu skala nominal mempunyai sifat (Rahman, 2020):

1. Kategori data bersifat mutually exclusive (saling memisah),
2. Kategori data tidak mempunyai aturan yang logis (bisa sembarang).

Skala nominal merupakan skala yang paling sederhana disusun menurut jenis (kategorinya) atau fungsi bilangan hanya sebagai simbol untuk membedakan sebuah karakteristik lainnya.

Skala nominal merupakan skala yang paling lemah/rendah di antara skala pengukuran yang ada. Skala nominal hanya bisa membedakan benda atau peristiwa yang satu dengan yang lainnya berdasarkan nama (predikat). Skala pengukuran nominal digunakan untuk mengklasifikasi objek, individu atau kelompok dalam bentuk kategori. Pemberian angka atau simbol pada skala nominal tidak memiliki maksud kuantitatif hanya menunjukkan ada atau tidaknya atribut atau karakteristik pada objek yang diukur. Misalnya, jenis kelamin diberi kode 1 untuk laki-laki dan kode 2 untuk perempuan. Angka ini hanya berfungsi sebagai label. Kategori tanpa memiliki nilai intrinsik dan tidak memiliki arti apa pun. Kita tidak bisa mengatakan perempuan dua kali dari laki-laki. Kita bisa saja mengkode laki-laki menjadi 2 dan perempuan dengan kode 1, atau bilangan apapun asal kodennya berbeda antara laki-laki dan perempuan. Misalnya lagi untuk agama, kita bisa mengkode 1 = Islam, 2 = Kristen, 3 = Hindu, 4 = Budha, dst. Kita bisa menukar angka-angka tersebut, selama satu karakteristik memiliki angka yang berbeda dengan karakteristik lainnya. Karena tidak memiliki nilai intrinsik, maka angka-angka (kode-kode) yang kita berikan tersebut tidak memiliki sifat sebagaimana bilangan pada umumnya.

Oleh karenanya, pada variabel dengan skala nominal tidak dapat diterapkan operasi matematika standar (aritmatik) seperti pengurangan, penjumlahan, perkalian, dan lainnya. Peralatan statistik yang sesuai dengan skala nominal adalah proposisi seperti modus, distribusi frekuensi, Chi Square dan beberapa peralatan statistik non-parametrik lainnya.

Ciri-ciri Skala nominal (Rahman, 2020):

1. Hasil penghitungan tidak dijumpai bilangan pecahan,
2. Angka yang tertera hanya label saja,
3. Tidak mempunyai urutan (ranking),
4. Tidak mempunyai ukuran baru,

5. Tidak mempunyai nol mutlak,
6. Tes statistik yang digunakan adalah statistik non parametrik.

Contoh Skala nominal:

1. Jenis kulit: Hitam Kuning Putih
2. Suku Daerah: Jawa Madura Bugis
3. Agama yang dianut: Islam Kristen Hindu
4. Partai pemenang pemilu: Golkar Demokrat PKB
5. Jenis kelamin: Laki Perempuan
6. Jenis Pekerjaan: PNS Swasta Tani dll
7. Status Perkawinan: Kawin Tidak Kawin

Contoh Skala nominal yang Tidak Sebenarnya:

1. Kelulusan: Lulus Tidak Lulus
2. Ijazah yang dipunyai: SD SMP SMA S1 S2 S3
3. Tahun Produksi Kendaraan: 2004 2005 2006 2007
4. Aktivitas: Bekerja Menganggur

### 2.3.2 Skala Ordinal

Skala ordinal merupakan skala yang lebih tinggi daripada skala nominal. Skala ini juga disebut skala peringkat. Skala ordinal biasa digunakan untuk mengukur data non-numerik seperti kepuasan pelanggan, kesenangan, kekecewaan, dan lain sebagainya. Arti dari ordinal adalah order atau urutan. Jadi, skala ini memberi label sesuai urutan. Contohnya kepuasan masyarakat dengan layanan transportasi umum dengan pilihan tidak puas diberi nilai satu, biasa saja diberi nilai dua, dan puas diberi nilai tiga. Urutan nilai satu, dua, dan tiga tidak boleh ditukar karena nilai tersebut merepresentasikan labelnya. Karakteristik dari skala ordinal adalah walaupun kita sudah mengetahui batas yang jelas, tetapi kita tidak mengetahui jarak antar batasnya. Misalnya, walaupun nilai satu menunjukkan variabel tidak puas, dua menunjukkan variabel biasa saja, dan tiga menunjukkan variabel puas, kita belum bisa mengetahui berapa jarak antara puas dan tidak puas sehingga kita tidak bisa mengatakan bahwa rasa puas tiga kali lebih tinggi dari rasa tidak puas. Ukuran pusat yang dapat dihitung pada skala ini adalah nilai median dan modus. Selain

itu, pada skala ini kita juga dapat menghitung rasio dan persentil. Sama seperti skala nominal, skala ordinal ini juga dapat divisualisasikan menggunakan diagram batang, diagram lingkaran, dan lain-lain (Kurniawati, 2020).

Skala ordinal dalam pengukuran di mana skala yang dipergunakan disusun berdasarkan atas jenjang dalam atribut tertentu sehingga penyusunannya disusun secara terurut dari yang rendah sampai yang tinggi menurut suatu ciri tertentu, namun antara urutan (ranking) yang satu dengan yang lainnya tidak mempunyai jarak yang sama.

Skala ordinal banyak dipergunakan dalam penelitian sosial dan pendidikan terutama berkaitan dengan pengukuran kepentingan, persepsi, motivasi serta sikap, apabila mengukur sikap responden terhadap suatu kebijakan pendidikan, responden dapat diurutkan dari mulai Sangat Setuju (1), Setuju (2), Tidak Berpendapat (3), Kurang Setuju (4), dan Tidak Setuju (5), maka angka-angka tersebut hanya sekedar menunjukkan urutan responden, bukan nilai untuk variabel tersebut.

Adapun ciri dari skala ordinal adalah (Rahman, 2020):

1. Kategori data bersifat saling memisah,
2. Kategori data mempunyai aturan yang logis,
3. Kategori data ditentukan skalanya berdasarkan jumlah karakteristik khusus yang dimilikinya.

Dapat juga dikatakan bahwa skala ordinal merupakan skala yang didasarkan pada ranking diurutkan dari jenjang yang lebih tinggi sampai jenjang yang lebih rendah atau sebaliknya. Skala ordinal ini lebih tinggi daripada skala nominal, dan sering juga disebut dengan skala peringkat. Hal ini karena dalam skala ordinal, lambing-lambang bilangan hasil pengukuran selain menunjukkan perbedaan juga menunjukkan urutan atau tingkatan objek yang diukur menurut karakteristik tertentu.

Misalnya tingkat kepuasan seseorang terhadap produk. Bisa kita beri angka dengan 5 = sangat puas, 4 = puas, 3 = kurang puas, 2 = tidak puas, dan 1 = sangat tidak puas. Atau misalnya dalam suatu lomba, pemenangnya diberi peringkat 1, 2, 3, dst. Dalam skala ordinal, tidak seperti skala nominal, ketika kita ingin mengganti angka-angkanya, harus dilakukan secara berurut dari besar ke kecil atau dari kecil ke besar. Jadi, tidak boleh kita buat 1 = sangat puas, 2 = tidak puas, 3 = puas, dst. Yang boleh adalah 1 = sangat puas, 2 = puas, 3 = kurang puas, dst.

Selain itu, yang perlu diperhatikan dari karakteristik skala ordinal adalah meskipun nilainya sudah memiliki batas yang jelas tetapi belum memiliki jarak (selisih). Kita tidak tahu berapa jarak kepuasan dari tidak puas ke kurang puas. Dengan kata lain juga, walaupun sangat puas kita beri angka 5 dan sangat tidak puas kita beri angka 1, kita tidak bisa mengatakan bahwa kepuasan yang sangat puas lima kali lebih tinggi dibandingkan yang sangat tidak puas. Sebagaimana halnya pada skala nominal, pada skala ordinal kita juga tidak dapat menerapkan operasi matematika standar (aritmatik) seperti pengurangan, penjumlahan, perkalian, dan lainnya. Peralatan statistik yang sesuai dengan skala ordinal juga adalah peralatan statistik yang berbasiskan (berdasarkan) jumlah dan proposisi seperti modus, distribusi frekuensi, Chi Square dan beberapa peralatan statistik non-parametrik lainnya.

Contoh Skala ordinal:

1. Stadium penyakit (berat, sedang, ringan)
2. Tingkat Pendidikan ( SD, SLTP, SLTA, PT )
3. Umur ( bayi, anak, dewasa, tua)
4. Status keadaan ekonomi ( bawah, menengah, atas )
5. Skala likert ( 1= sangat setuju 2= setuju 3= ragu-ragu 4= tidak setuju 5= sangat tidak setuju).

### 2.3.3 Skala Interval

Skala interval adalah skala numerik di mana kita bisa mengetahui urutan dan perbedaan jarak antar datanya. Karakteristik skala interval sama dengan skala nominal dan ordinal hanya saja ditambah dengan interval yang tetap. Namun skala ini belum memiliki nilai nol mutlak. Artinya, jarak yang dapat dihitung tidak berarti kelipatan. Contohnya, di sebuah kelas, siswa A memperoleh nilai 50, siswa B memperoleh nilai 80, dan siswa C memperoleh nilai 100. Dalam kasus ini kita dapat mengatakan bahwa selisih nilai siswa C dan A adalah 50, tetapi kita tidak bisa mengatakan bahwa siswa C dua kali lebih pintar dari siswa A. Hal ini karena skala interval sudah berupa angka-angka dan dapat dihitung menggunakan operasi matematika tetapi belum bisa digunakan untuk menghitung rasio atau koefisien variasi pada statistika (Kurniawati, 2020).

Skalah interval adalah skala pengukuran di mana jarak satu tingkat dengan tingkat lainnya sama, oleh karena itu skala interval dapat juga disebut skala

unit yang sama (equal unit scale). Skala interval merupakan skala yang dihasilkan dari proses pengukuran, di mana dalam pengukuran tersebut diasumsikan terdapat satuan (unit) pengukuran yang sama. Contoh yang sangat dikenal adalah temperatur.

Adapun ciri-ciri skala interval adalah (Rahman, 2020):

1. Kategori data bersifat saling memisah,
2. Kategori data bersifat logis,
3. Kategori data ditentukan skalanya berdasarkan jumlah karakteristik khusus yang dimilikinya,
4. Perbedaan karakteristik yang sama tergambar dalam perbedaan yang sama dalam jumlah yang dikenakan pada kategori,
5. Angka nol hanya menggambarkan suatu titik dalam skala (tidak punya nilai nol absolut).

Dengan demikian skala interval merupakan skala yang menunjukkan jarak antara satu data dengan data yang lain dan mempunyai bobot yang sama. Analisis statistik yang digunakan adalah mempunyai karakteristik uji statistik parametrik. Skala interval mempunyai karakteristik seperti yang dimiliki oleh skala nominal dan ordinal dengan ditambah karakteristik lain, yaitu berupa adanya interval yang tetap. Dengan demikian, skala interval sudah memiliki nilai intrinsik, sudah memiliki jarak, tetapi jarak tersebut belum merupakan kelipatan. Pengertian “jarak belum merupakan kelipatan” ini kadang-kadang diartikan bahwa skala interval tidak memiliki nilai nol mutlak. Misal pada pengukuran suhu. Kalau ada tiga daerah dengan suhu daerah A = 10°C, daerah B = 15°C dan daerah C = 20°C. Kita bisa mengatakan bahwa selisih suhu daerah B 5°C lebih panas dibandingkan daerah A, dan selisih suhu daerah C dengan daerah B adalah 5°C (ini menunjukkan pengukuran interval sudah memiliki jarak tetap). Tetapi, kita bisa mengatakan bahwa suhu daerah C dua kali lebih panas dibandingkan daerah A (artinya tidak bisa jadi kelipatan). Kenapa? karena dalam derajat Celcius tidak memiliki nol absolut. (Titik nolnya pada 0°C Bukan berarti Tidak ada Suhu sama sekali).

#### **Contoh Skala Interval:**

Frekuensi nafas, tekanan darah, nadi, Hb, pengetahuan dan derajat Celcius. Misalnya 0° C, adalah titik beku dan bukan bererti tidak ada suhunya ini menunjukkan bahwa nol punya nilai (nol bukan absolut). Karena tidak ada titik

nol sebenarnya pada skala, maka tidak dapat dikatakan bahwa 4 derajat Celcius menggambarkan  $4 \times$  sepanas nol derajat Celcius (Rahman, 2020).

### 2.3.4 Skala Rasio

Skala rasio adalah skala tertinggi dibanding jenis skala pengukuran yang lain. Skala rasio sudah mencakup seluruh karakteristik pada skala nominal, ordinal, dan interval, Skala ini juga sudah bisa diinterpretasikan pada semua operasi statistik bahkan dapat menghitung rasio atau perbandingan karena skala ini sudah memiliki nilai nol mutlak. Arti dari nilai nol mutlak adalah nilai dasar yang sudah tidak dapat diubah meskipun menggunakan skala lain. Contoh skala rasio adalah berat dan tinggi badan seseorang. Misalnya, berat si A 20 kg sedangkan berat si B 40 kg, kita dapat mengatakan bahwa berat si B adalah dua kali berat si A (Kurniawati, 2020).

Skala rasio merupakan skala pengukuran yang mempunyai nilai Nol mutlak dan mempunyai jarak yang sama. Skala interval yang benar-benar memiliki nilai nol mutlak disebut skala rasio, dengan demikian skala rasio menunjukkan jenis pengukuran yang sangat jelas dan akurat (precise). Jika kita memiliki skala rasio, kita dapat menyatakan tidak hanya jarak yang sama antara satu nilai dengan nilai lainnya dalam skala, tapi juga tentang jumlah proporsional karakteristik yang dimiliki dua objek atau lebih, dan contoh untuk skala ini adalah uang.

Adapun ciri-ciri dari skala rasio adalah:

1. Kategori data bersifat saling memisah,
2. Kategori data mempunyai aturan yang logis,
3. Kategori data ditentukan skalanya berdasarkan jumlah karakteristik khusus yang dimilikinya,
4. Perbedaan karakteristik yang sama tergambar dalam perbedaan yang sama dalam jumlah yang dikenakan pada kategori,
5. Angka nol menggambarkan suatu titik dalam skala yang menunjukkan ketidadaan karakteristik (punya nilai nol absolut).

Tes yang digunakan adalah tes statistik parametrik. Skala rasio adalah skala data dengan kualitas paling tinggi. Pada skala rasio, terdapat semua karakteristik skala nominal, ordinal, dan skala interval ditambah dengan sifat adanya nilai nol yang bersifat adanya nilai nol bersifat mutlak. Nilai nol mutlak

ini artinya adalah nilai dasar yang tidak bisa diubah meskipun menggunakan skala yang lain. Oleh karenanya, pada skala ratio, pengukuran sudah mempunyai nilai perbandingan/rasio. Pengukuran-pengukuran dalam skala rasio yang sering digunakan adalah pengukuran tinggi dan berat. Misalnya berat benda A adalah 30 kg, sedangkan benda B adalah 60 kg, maka dapat dikatakan bahwa benda B lebih berat dua kali dibandingkan benda A

Contoh Skala Rasio:

1. Umur manusia dalam ... tahun
2. Ukuran timbangan dalam ... kg
3. Berat badan dalam ... kg
4. Tinggi pohon dalam... cm
5. Tinggi badan manusia dalam ... cm
6. Jarak dalam ... meter

Dari uraian di atas jelas bahwa Skala Ratio, Interval, Ordinal dan Nominal berturut-turut memiliki nilai kuantitatif dari yang Paling Rinci ke yang Kurang Rinci. Skala Ratio mempunyai sifat-sifat yang dimiliki Skala Interval, Ordinal dan Nominal. Skala Interval memiliki ciri-ciri yang dimiliki Skala Ordinal dan Nominal, sedangkan Skala Ordinal memiliki sifat yang dimiliki Skala Nominal. Adanya perbedaan tingkat pengukuran memungkinkan terjadinya Transformasi Skala Ratio dan Interval menjadi Ordinal atau Nominal. Transformasi ini dikenal sebagai Data Reduction atau Data Collapsing. Hal ini dimaksudkan agar dapat menerapkan metode statistik tertentu, terutama yang menghendaki skala data dalam bentuk Ordinal atau Nominal.

Sebaliknya, Skala Ordinal dan Nominal tidak dapat diubah menjadi Interval atau Ratio. Skala Nominal yang diberi label 0, 1 atau 2 dikenal sebagai Dummy Variable (Variabel Rekayasa). Misalnya: Pemberian label 1 untuk laki – laki dan 2 untuk perempuan tidak mempunyai arti kuantitatif (tidak mempunyai nilai / hanya kode). Dengan demikian, perempuan tidak dapat dikatakan 1 lebih banyak dari laki – laki. Pemberian label tersebut dimaksudkan untuk mengubah kategori huruf (Alfabet) menjadi kategori Angka (Numerik), sehingga memudahkan analisis data (Rahman, 2020).

## 2.4 Perubahan Skala

Peneliti dalam menganalisis data biasanya merubah skala pengukuran, misalnya dari skala rasio diubah menjadi nominal. Sebagai contoh variabel umur (diukur dalam tahun) skalanya rasio, diubah menjadi umur berisiko dan tidak berisiko ini menjadi skala nominal. Seharusnya peneliti mempertahankan skala pengukuran tetap rasio, karena ini termasuk menurunkan skala dari rasio turun ke nominal. Konsekwensi dari penurunan skala tentunya berdampak kepada hasil analisis statistik menjadi tidak sahih. Sebagai ilustrasi adalah variabel umur diukur dalam tahun, kita berubah menjadi umur bersiko jika  $> 35$  tahun, tidak berisiko jika  $\leq 35$  tahun. Dari sini dapat diartikan bahwa jika seseorang berumur 36 tahun termasuk berisiko, begitu pula jika seseorang berumur 50 tahun juga termasuk berisiko. Jadi seseorang yang berumur 36 tahun disamakan risikonya dengan 50 tahun, padahal secara fisik orang yang berumur 36 tahun dan 50 tahun pasti berbeda. Kesalahan ini tentunya berdampak pada hasil analisis statisitk, yang seharusnya  $H_0$  ditolak menjadi  $H_0$  diterima begitu pula sebaliknya. Harapan sebagai peneliti seharusnya jangan menurunkan skala, kalau perlu menaikkan skala karena sudah dijelaskan di atas bahwa semakin tertinggi skala yang digunakan, maka semakin tinggi pula ketelitiannya (Rahman, 2020).

Contoh Perubahan Skala:

1. Penurunan skala: Berat badan lahir (grm) skala rasio diturunkan menjadi berat badan normal dan tidak normal skala nominal.
2. Menaikkan skala: Tingkat pendidikan (SD, SMP, SMA, PT) skala ordinal dinaikkan menjadi lama menempuh pendidikan (thn) skala rasio.

Untuk memperjelas perbedaan sifat yang dimiliki oleh empat skala pengukuran dapat dilihat pada Tabel 2.1

**Tabel 2.1:** Berbagai Skala Pengukuran Variabel dengan Ciri yang Dipunyai  
(Rahman, 2020)

Skala Pengukuran Variabel	C	i	r	i
	Membedakan	Tingkatan	Besar beda	Kelipatan/ada titik nol
Nominal	Ya	Tdk	Tdk	Tdk
Ordinal	Ya	Ya	Tdk	Tdk
Interval	Ya	Ya	Ya	Tdk
Rasio	Ya	Ya	Ya	Ya

# **Bab 3**

## **Pemeriksaan Data**

### **3.1 Pendahuluan**

Data merupakan bagian penting dari informasi, data adalah fakta mentah yang belum diolah sehingga belum memiliki makna. Data bisa berupa angka, teks, gambar bahkan video. Agar suatu data menjadi sebuah informasi maka diperlukan perlakuan khusus atau yang disebut sebagai pengolahan data. Data yang telah diolah diperlukan suatu metode untuk penyajiannya, atau yang disebut dengan penyajian data.

Penyajian data adalah proses menampilkan hasil olahan data sehingga dapat dimengerti dan tersusun. Penyajian data dapat ditampilkan dalam bentuk diagram maupun tabel sesuai dengan kebutuhannya. Data yang telah disajikan dapat dijadikan acuan dalam proses analisa dan pengambilan kesimpulan atau keputusan secara baik dan benar. Dalam proses penyajian data, data dibagi menjadi dua yaitu data tunggal dan data kelompok. Untuk dapat menyajikan data dengan baik dan benar maka diperlukan pemeriksaan data terlebih dahulu.

Proses pemeriksaan data merupakan salah satu ilmu statistika yang dapat dijadikan dasar untuk mengambil kesimpulan kondisi dari sampel data yang diambil. Dalam proses ini diperlukan hipotesa awal yang dapat diuji agar bisa dilakukan validitas jika syarat terpenuhi.

## 3.2 Data

Berdasarkan jenisnya data dibagi menjadi dua macam yaitu data kualitatif dan data kuantitatif. Berikut penjelasan singkat tentang data kualitatif dan data kuantitatif.

### 3.2.1 Data kualitatif

Data kualitatif adalah data yang direpresentasikan dalam bentuk kata, kalimat, dan gambar, data yang dimiliki bisa berbentuk naratif atau deskriptif. Data kualitatif sering digunakan dalam penelitian yang bersifat non statistik. Data kualitatif bisa didapatkan dari wawancara dengan narasumber terkait dengan data yang ingin kita dapatkan, data tersebut diolah menjadi informasi sehingga dapat dijadikan pedoman untuk proses pengambilan data

Contoh data kualitatif:

1. Kalimat penjelasan dari narasumber.
2. Wawancara
3. Observasi langsung kelapangan
4. Fenomena yang terjadi seperti bencana alam dan lainnya

Studi kasus: Untuk mengetahui permasalahan pada suatu objek penelitian kita harus melakukan wawancara dengan pihak terkait guna mengetahui jawaban untuk kita jadikan sebagai data penelitian.

### 3.2.2 Data Kuantitatif

Data kuantitatif adalah data yang bersifat statistik atau numerik yang menunjukkan hasil pengukuran dari data yang telah diambil, data tersebut berbentuk angka atau nilainya dapat dijumlahkan. Data kuantitatif diperoleh dengan cara membagikan angket atau survei yang dilakukan langsung ke subjek penelitian. Metode ini dilakukan untuk mengetahui kondisi lapangan dan direpresentasikan ke dalam bentuk angka atau numerik.

Contoh data kuantitatif:

1. Data berupa angka
2. Nilai rata-rata
3. Data bisa dijumlahkan

Studi kasus: untuk mengetahui tingkat kepuasan pelanggan maka data yang cocok digunakan adalah jenis data kuantitatif. Karena pengambilan data yang digunakan adalah dengan membagikan angket atau survey.

## 3.3 Pemeriksaan Data

Pemeriksaan data adalah proses validasi apakah data yang telah diambil dapat dijadikan sebagai data atau tidak. Fungsi pemeriksaan data ini adalah untuk memastikan tidak adanya kesalahan dan kekurangan dari data yang telah diambil. Pemeriksaan data akan berpengaruh pada keakuratan hasil dari pengolahan data, setelah pencocokan data dilakukan pengamatan terhadap respon variabel. Pada tahapan selanjutnya akan dilakukan pemeriksaan dan menentukan apakah model yang dimiliki sudah memenuhi kriteria kelayakan atau tidak .

Pemeriksaan data dilakukan berdasarkan hipotesa awal dari hasil pengamatan data yang telah dikumpulkan, pengujian hipotesa awal dinyatakan valid bila telah memenuhi syarat-syarat normalitas, homogen dan linier terhadap kebutuhan data. Model data yang digunakan akan dicocokan dengan data observasi sesuai dengan variabel respon yang dimiliki, dengan demikian langkah selanjutnya dapat dinyatakan apakah model tersebut dapat digunakan untuk data yang dimiliki atau tidak. Ada beberapa cara yang dapat dilakukan untuk model pemeriksaan kelayakan salah satunya dengan menguji normalitas data dengan syarat semua komponen data linier dengan model data yang digunakan.

### 3.3.1 Normalitas Data

Uji Normalitas adalah melakukan pengujian data dengan tujuan untuk menilai data yang dimiliki pada sebuah kelompok data atau variabel, apakah data tersebut berjalan normal, sesuai atau tidak. Manfaat pengujian Normalitas adalah untuk menjadikan data yang telah dikumpulkan memiliki nilai normal sesuai dengan populasi dan sampel data yang diambil. Metode umum yang sering digunakan dalam pengujian normalitas data yaitu berdasarkan pengalaman empiris beberapa pakar statistik, data yang banyaknya lebih dari 30 angka ( $n > 30$ ), maka sudah dapat diasumsikan berdistribusi normal. Biasa dikatakan sebagai sampel besar.

Untuk memastikan apakah data berdistribusi normal atau tidak, maka perlu dilakukan uji normalitas. Hal ini disebabkan karena data yang dimiliki walaupun jumlahnya lebih dari 30 data tetapi tidak dapat dipastikan nilainya, demikian sebaliknya data yang banyaknya kurang dari 30 belum juga belum tentu tidak berdistribusi normal, untuk itu perlu suatu pembuktian uji statistik normalitas yang dapat digunakan diantaranya Chi-Square, Kolmogorov Smirnov, Lilliefors, Shapiro Wilk, Jarque Bera.

### 3.3.2 Homogen

walaupun memiliki persamaan sama-sama digunakan sebagai syarat dalam uji parameter namun Uji homogen berbeda dengan uji normalitas. Perbedaannya terletak pada perlakuan datanya, jika uji normalitas diperlukan untuk semua uji parameter, maka uji homogenitas tidak selalu digunakan untuk semua parameter. Uji homogenitas digunakan hanya pada uji parameter yang menguji perbedaan antara kedua kelompok atau beberapa kelompok yang berbeda subjeknya atau sumber datanya. Oleh karena itu, uji homogenitas diperlukan sebagai asumsi dari uji independent T test dan uji Anova. Sedangkan pada uji regresi data linear, homogenitas tidak diperlukan sebagai syarat dikarenakan uji regresi linear tidak menguji perbedaan pada beberapa kelompok data (Anwar, 2017).

Risiko kesalahan asumsi atau homogenitas tidak terpenuhi berbeda-beda tergantung pada hipotesis utama dari tiap penelitian. Misalkan pada uji Anova, jika asumsi homogenitas tidak terpenuhi, maka peneliti dapat menggunakan koreksi oleh uji brown forsythe atau welch's F. Sedangkan jika asumsi homogenitas tidak terpenuhi pada uji independen t test, peneliti dapat menggunakan uji independen t test unequal variance atau menggunakan uji indepeden welch's test (Anwar, 2017).

#### Contoh Uji Homogenitas

Uji homogenitas memiliki beberapa metode perhitungan yang sering digunakan dalam pemeriksaan data, yang populer adalah: Uji Levene test, Fisher F dan Bartlett Test.

Secara harfiah Uji homogenitas dapat diartikan sebagai metode uji yang menilai apakah data yang dimiliki memiliki perbedaan varian atau jenis dari dua kelompok atau lebih.

### **Persamaan Uji Normalitas dan Homogenitas**

Uji normalitas dan uji homogen sama-sama digunakan sebagai syarat dalam pengujian parameter pemeriksaan data. Dengan pengujian ini maka akan diketahui nilai dari tiap hipotesis dan asumsi awal data.

### **Perbedaan Uji Normalitas dan Homogenitas**

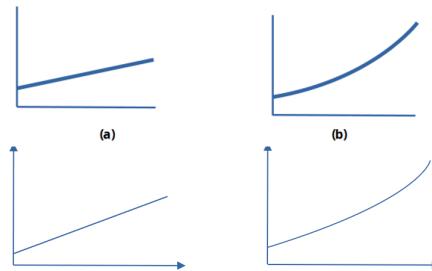
Perbedaannya terletak pada perlakuan datanya, jika uji normalitas diperlukan untuk semua uji parameter, maka uji homogenitas tidak selalu digunakan untuk semua parameter. Uji normalitas digunakan sebagai asumsi atau syarat dari setiap uji parametris, sedangkan uji homogenitas hanya diperlukan untuk menguji parameter yang memiliki nilai atau menilai perbedaan antara dua kelompok atau lebih.

### **3.3.3 Linieritas Data**

Dalam beberapa analisis statistik parametrik, seperti korelasi pearson dan analisis regresi, salah satu asumsi yang mendasari analisis adalah hubungan antar variabel membentuk pola (model) yang linear. Model linear artinya pola hubungan kedua variabel independen dan dependen akan membentuk satu garis lurus. Beberapa pendapat muncul terkait perlu tidaknya kita menguji asumsi linearitas ini terlebih dahulu sebelum melakukan uji hipotesis. Pendapat ini wajar saja, karena memang uji hipotesis, dengan korelasi pearson misalnya, mendasarkan hubungannya harus linear, jadi ketika hasil korelasi signifikan, sudah dipastikan asumsi linearitas juga terpenuhi. Jadi menguji asumsi linearitas di awal adalah sesuatu yang mubazir.

Namun pada kenyataannya, tidak semua variabel Psikologi di dunia ini hubungannya linear. Ada yang membentuk kurve. Seperti hubungan antara uang dan kebahagiaan, pada tingkatan rendah, uang memang akan dapat meningkatkan kebahagiaan. Tapi pada satu titik tertentu, peningkatan uang tidak terlalu berpengaruh lagi terhadap peningkatan kebahagiaan. Tetapi tetap kita harus akui, ada hubungan antara uang dan kebahagiaan, meskipun pola hubungannya tidak linear sempurna. Di sini saya akan menyajikan contoh penelitian yang dapat menjelaskan perlunya kita melihat linearitas hubungan dan apa pengaruhnya terhadap penyimpulan statistik kita.

Hubungan antara variable dependent dan independen yang membentuk pola liner dan kurve dapat dilihat pada Gambar 3.1.



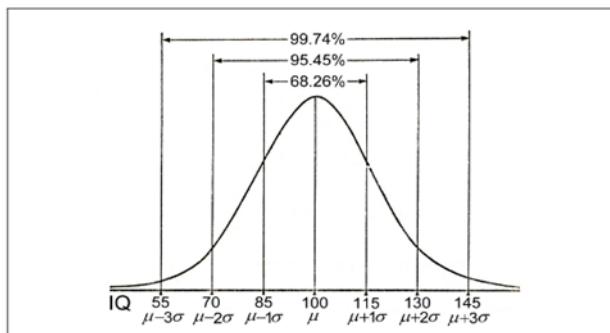
**Gambar 3.1:** Linieritas Data (a) Grafik menunjukkan nilai linier. (b) Grafik menunjukkan nilai meningkat secara signifikan

## 3.4 Uji Normalitas

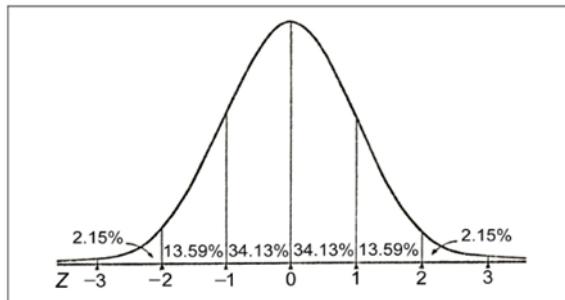
Seperti yang dijelaskan pada point 3.3.1, uji normalitas memiliki beberapa cara atau metode yang dapat digunakan untuk menormalisasi data yaitu:

### 3.4.1 Uji Grafik

Uji grafik merupakan metode pengamatan distribusi data pada sumber data, yaitu titik diagonal pada grafik. Data dapat dinyatakan bernilai normal bila penyebaran titik data berada di area garis dan mengikuti diagonal grafik, maka nilai data tersebut adalah normal.



**Gambar 3.2:** Grafik Data normal



**Gambar 3.3:** Grafik Distribusi Data normal

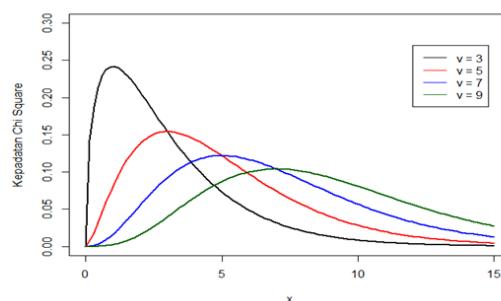
Gambar 3.2 dan 3.3 menunjukkan garis grafik untuk pengujian data normal dengan nilai garis continue.

#### 2.4.2 Metode Chi-Square

Chi-Square adalah metode yang digunakan untuk Uji Goodness of fit Distribusi data Normal dengan menggunakan pendekatan penjumlahan penyimpangan data berdasarkan hasil observasi dari tiap kelas.

Syarat dari Metode Chi Square adalah

1. Data tersusun berkelompok atau dikelompokkan dalam satu tabel dengan distribusi frekuensi yang sama
2. Digunakan untuk data yang memiliki jumlah data yang memiliki angka besar ( $n > 30$ )



**Gambar 3.4:** Gambar kurva data chi-square

### 3.4.3 Metode Lilliefors

Lilliefors merupakan metode yang menggunakan data yang belum diolah baik dalam tabel maupun distribusi grafik. Data dasar akan diubah menjadi nilai z agar dapat dihitung berapa jumlah data normal yang dimiliki sebagai nilai probabilitas keseluruhan.

Syarat metode Lilliefors

1. Data yang digunakan berskala interval atau rasio yang dikenal dengan angka penjumlahan
2. Data yang dimiliki berupa data tunggal atau belum dikelompokkan dalam satu tabel dengan distribusi frekuensi yang sama
3. Dapat digunakan untuk jumlah data n lebih besar atau n lebih kecil.

**Tabel 3.1:** Format Data lilliefors

DF	Probabilitas					Lilliefors
	0.01	0.05	0.1	0.2	0.3	
1	1.03100	0.88600	0.80500	0.76800	0.73600	
2	0.72903	0.62650	0.56922	0.54306	0.52043	
3	0.59525	0.51153	0.46477	0.44341	0.42493	
4	0.51550	0.44300	0.40250	0.38400	0.36800	
5	0.46108	0.39623	0.36001	0.34346	0.32915	
6	0.42090	0.36171	0.32864	0.31353	0.30047	
7	0.38968	0.33488	0.30426	0.29028	0.27818	
8	0.36451	0.31325	0.28461	0.27153	0.26022	
9	0.34367	0.29533	0.26833	0.25600	0.24533	
10	0.32603	0.28018	0.25456	0.24286	0.23274	Lilliefors Tabel

### 3.4.4 Metode Kolmogorov-Smirnov

Metode Kolmogorov-Smirnov tidak berbeda jauh dengan metode Lilliefors. Langkah-langkah Penyelesaian dan penggunaan rumusnya sama, yang membedakannya adalah pada hasil signifikansi. Signifikansi Metode Kolmogorov-Smirnov menggunakan tabel pembanding Kolmogorov-Smirnov, sedangkan Metode Lilliefors menggunakan tabel pembanding metode Lilliefors.

Syarat yang dimiliki oleh metode ini adalah

1. Data yang digunakan berskala interval atau rasio yang dikenal dengan angka penjumlahan
2. Data yang dimiliki berupa data tunggal atau belum dikelompokkan dalam satu tabel dengan distribusi frekuensi yang sama

3. Dapat digunakan untuk jumlah data n lebih besar atau n lebih kecil.

**Tabel 3.2:** Kolmogorov-Smirnov

No	X <sub>i</sub>	Z = $\frac{X_i - \bar{X}}{SD}$	F <sub>T</sub>	F <sub>S</sub>	F <sub>T</sub> - F <sub>S</sub>
1					
2					
3					
dst					

Keterangan:

X<sub>i</sub> = Angka pada data

Z = Transformasi dari angka ke notasi pada distribusi normal

F<sub>T</sub> = Probabilitas komulatif normal

F<sub>S</sub> = Probabilitas komulatif empiris.

### 3.4.5 Metode Shapiro Wilk

metode ini menggunakan data awal yang belum diolah baik dalam tabel maupun distribusi grafik. Data diurutkan dan akan dibagi menjadi dua kelompok untuk dikonversi dalam Shapiro Wilk. Dapat juga dilanjutkan transformasi dalam nilai Z untuk dapat dihitung luasan kurva normal. Metode Metode Shapiro Wilk menguji hipotesis yang berasal dari distribusi normal tidak bergantung pada nilai rata-rata dan variansi.

Syarat yang dimiliki oleh metode ini adalah

1. Data yang digunakan berskala interval atau rasio yang dikenal dengan angka penjumlahan
2. Data yang dimiliki berupa data tunggal atau belum dikelompokkan dalam satu tabel dengan distribusi frekuensi yang sama
3. Menggunakan data dari sampel acak.

Langkah-langkah pengujian Metode Shapiro Wilk

1. Merumuskan hipotesis yang akan di uji

H<sub>0</sub>: data berdistribusi Normal

H<sub>1</sub>: data tidak berdistribusi Normal

2. Menentukan tingkat signifikansi ( $\alpha$ )
3. Menentukan statistik uji Statistik

Uji Shapiro-Wilk didefinisikan dengan

$$w = \frac{b^2}{s^2} = \frac{\left( \sum_{i=1}^n a_i x_i \right)^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \dots \dots \dots \text{(Putri Rika Dwiana, 2020)}$$

di mana, jika  $n$  genap, maka  $n = 2k$

### 3.5 Model Pemeriksaan Data

Menurut penelitian yang dilakukan oleh Nusar Hajarisman, model pemeriksaan data akan berpengaruh dalam regresi Gamma. Model regresi gamma merupakan model regresi yang variabel responnya didasarkan pada distribusi gamma, yang juga merupakan anggota dari eksponensial, sehingga terminologi dari model linear terampat dapat diterapkan dalam pemodelan statistik.

Misalkan diamati suatu variabel respon  $y_i$  untuk  $n$  buah pengamatan. Asumsi dasar yang diperlukan dalam model gamma ini adalah

$$\text{var}(y_i) = \sigma^2 [E(y_i)]^2 \text{ untuk } i = 1, \dots, n \text{ (sumber pemeriksaan data)}$$

yaitu, koefisien variasi pengamatannya merupakan suatu konstanta dan koefisien variasi umum dinyatakan dengan  $\sigma^2$ . Apabila nilai yang mungkin dari variabel respon berupa bilangan nyata positif dan apabila respon tersebut berasal dari distribusi gamma, maka akan diperoleh bentuk khusus di mana  $\sigma^2 = 1/v$  dan  $v$  merupakan parameter bentuk (shape parameter). (sumber pemeriksaan data)

Langkah-langkah pemodelan statistika merupakan proses pencocokan data dengan variabel respon, dengan mencocokkan hipotesis yang berhubungan dengan pemeriksaan kelayakan data suatu model. Setelah model dicocokan dengan data hasil observasi dengan variabel respon, maka tahapan selanjutnya adalah melakukan pemeriksaan apakah model yang digunakan layak atau tidak. Ada beberapa cara yang dapat dijadikan sebagai model hipotesa, dengan syarat komponen yang dimiliki sistematis dan linear dengan model data yang digunakan (Nusar Hajarisman, 2010).

Contoh,

Dalam sebuah pemeriksaan data tidak menyertakan variable prediksi untuk dijadikan hipotesis dalam model dan nilai-nilai yang diambil tidak memiliki variable yang jelas dan perlu ditransformasikan lagi. Transformasi dari peluang respon yang didapat menjadi kurang tepat. Data akan berisi observasi dengan nilai data yang kecil, hal ini dapat menyebabkan model data menjadi tidak bagus. Data observasi yang berpengaruh atau dikenal dengan influence data, memiliki pengaruh dan berperan besar dalam hasil kesimpulan dari analisis.

Teknik yang dipakai untuk menentukan kelayakan model hipotesis secara kolektif disebut sebagai Diagnostics. Teknik ini dilakukan dengan uji statistik formal, namun sering juga digunakan untuk evaluasi non-formal dengan tabel dari nilai statistik tertentu melalui gambaran grafik atau plot dari nilai-nilai yang dimiliki.



# **Bab 4**

## **Teknik Penyajian Data**

### **4.1 Pendahuluan**

Kegiatan pengumpulan data di lapangan akan menghasilkan data berupa angka-angka yang menunjukkan bahwa data tersebut belum diolah dengan menggunakan teknik statistik tertentu. Jadi data tersebut masih berwujud sebagaimana data itu diperoleh yang biasanya berupa skor dan relative banyak tidak beraturan. Dalam pembuatan laporan penelitian, data termasuk yang harus dilaporkan. Agar dapat memberikan gambaran yang bermakna, data-data itu haruslah disajikan ke dalam tampilan yang sistematis dan untuk keperluan penganalisisan. Penyajian data ini bertujuan memudahkan pengolahan data dan pembaca memahami data.

Pada dasarnya banyak cara untuk menyajikan data, sehingga ia dapat dipahami dan digunakan secara tepat oleh pengolah data. Beberapa jenis penyajian data di antaranya adalah berupa teks, tabel dan grafik. Tabel menyajikan data ke dalam bentuk baris atau kolom sedemikian rupa sehingga memberikan informasi lebih kepada peneliti. Sedangkan grafik menyajikan data dari tabel tersebut menjadi bentuk visual yang lebih informatif lagi. Penyajian data ini tidak hanya sangat membantu peneliti untuk mengetahui gambaran data awal, namun digunakan juga pada analisis inti penelitian atau pelaporan.

## 4.2 Konsep Pengumpulan Data

Pengumpulan dan penyajian data merupakan proses yang tidak bisa dipisahkan dalam sebuah penelitian, karena proses pengumpulan dan penyajian data tersebut merupakan salah satu bagian yang sangat penting dalam pembuatan laporan atau penelitian. Laporan jenis apapun selalu diawali dari pengumpulan data kemudian menyajikan data tersebut yang terkandung dalam laporan tersebut. Sebelum memulai penyajian data, kita pahami dahulu bagaimana cara-cara pengumpulan data.

Menurut Sugiyono (2016), menyatakan bahwa ada beberapa metode dalam melakukan pengumpulan data, antara lain :

1. Observasi (pengamatan yang melibatkan semua indera yaitu penglihatan, pendengaran, penciuman, pembau, perasa)
2. Wawancara, terbagi menjadi :
  - a. Wawancara tidak terstruktur
    - Merupakan langkah persiapan wawancara terstruktur
    - Pertanyaan yang diajukan merupakan upaya menggali isu awal
    - Sifat pertanyaan spontan
  - b. Wawancara terstruktur
    - Pertanyaan sudah disiapkan, karena sudah dirancang data atau informasi apa yang dibutuhkan
    - Adapun jenis wawancara adalah wawancara langsung (face to face) dan wawancara tidak langsung (misalnya dengan telepon atau internet/on-line)
3. Kuesioner (daftar pertanyaan)
  - a. Kuesioner adalah daftar pertanyaan tertulis yang ditujukan kepada responden. Jawaban responden atas semua pertanyaan dalam kuesioner kemudian dicatat atau direkam.
  - b. Kuesioner merupakan metode pengumpulan data yang efisien bila peneliti mengetahui secara pasti data atau informasi apa yang dibutuhkan dan bagaimana variabel yang menyatakan informasi yang dibutuhkan tersebut diukur.

#### 4. Pengukuran Fisik

Jenis data dapat berupa data fisik, yaitu berupa objek atau benda fisik, seperti tanah, bangunan, buku, kendaraan, dan lain-lain. Sementara itu, metode pengukuran fisik ini berupa memetakan obyek empirik ke obyek angka-angka dengan perubahan yang setara. Misalnya pada data tanah, kita dapat mengumpulkan data fisik tanah dari ukuran panjang (meter), lebar (meter), jenis tanah atau jenis batuannya.

#### 5. Percobaan Laboratorium

Pengumpulan data dengan metode ini dilakukan di laboratorium. Sebelumnya, dilakukan perencanaan percobaan (experimen design). Laboratorium yang dimaksud tidak hanya berupa ruangan untuk kegiatan biologi, kimia, fisika, kedokteran, atau ilmu rekayasa, namun dapat pula berupa setiap ruang yang memiliki pengendalian ketat untuk setiap faktornya.

Sebagai contoh adalah penelitian untuk mengetahui apakah pH memengaruhi pertumbuhan tanaman. Dalam hal ini pertumbuhan tanaman sangat dipengaruhi oleh banyak faktor, oleh karena itu perlu pengendalian yang ketat bagi tanaman sebagai sampel. Pengendalian berupa penggunaan laboratorium, di mana sampel dikontrol secara ketat di ruangan dan hanya pH yang diberikan pada tanaman

### 4.3 Konsep Penyajian Data

Penyajian data merupakan salah satu kegiatan dalam pelaporan hasil penelitian yang telah dilakukan agar dapat dipahami dan dianalisis sesuai dengan tujuan yang diinginkan. Data yang disajikan harus sederhana dan jelas agar mudah dibaca. Penyajian data juga dimaksudkan agar para pengamat dapat dengan mudah memahami apa yang disajikan untuk selanjutnya dilakukan penilaian atau perbandingan.

Menurut Syukra (2016), menyatakan bahwa prinsip dasar penyajian data adalah bagaimana data dapat komunikatif dan lengkap. Data yang disajikan harus menarik perhatian pihak lain untuk membaca dan tentunya mudah

dipahami. Penyajian data ini berfungsi untuk memberikan gambaran awal dari hasil pengumpulan data, informasi data lebih cepat dimengerti, dan memudahkan proses analisis data.

## 4.4 Manfaat dan Tujuan Penyajian Data

Menurut Supranto (2000), mengemukakan bahwa penyajian data bermanfaat sebagai informasi dan menyajikan data bisa dalam bentuk textular, tabular dan grafik tergantung dari jenis data dan skala pengukurannya. Untuk mendapatkan penyajian data yang benar, memerlukan pengolahan data terlebih dahulu. Data statistik tidak hanya cukup dikumpulkan dan diolah, tetapi perlu disajikan dalam bentuk dan mudah dibaca serta dimengerti oleh pengambil keputusan. Keuntungan penyajian data adalah data tersebut akan lebih cepat ditangkap dan dimengerti daripada disajikan dalam bentuk kata-kata.

Menurut Swarjana (2016), penyajian data tidak hanya sekedar penyajian angka, teks atau grafik, melainkan menyajikan data yang akan bermanfaat untuk menarik kesimpulan dengan cepat dan tepat, serta mempercepat pengambilan keputusan. Data statistik maupun penelitian dapat disajikan dengan berbagai cara, di antaranya dapat disajikan dalam bentuk tabel, grafik dan teks. Kelemahan data dalam bentuk tekstual adalah diperlukan waktu yang lebih banyak untuk menyusun kata dan kalimat menjadi satu atau lebih paragraf agar mampu menjelaskan angka atau data statistik.

Menurut Aunuddin (2005), menyatakan bahwa beberapa manfaat penyajian data, di antaranya:

1. Meringkas data, baik data kualitatif maupun kuantitatif.
2. Data kualitatif berupa distribusi frekuensi, frekuensi relatif, persen distribusi frekuensi, grafik batang, dan grafik lingkaran.
3. Data kuantitatif berupa distribusi frekuensi, relatif frekuensi dan persen distribusi frekuensi, diagram atau plot titik, distribusi kumulatif, dan ogive.
4. Dapat digunakan untuk melakukan eksplorasi data.
5. Membuat tabulasi silang dan diagram sebaran data

Tujuan penyajian data adalah sebagai berikut (Furqoni, 2021):

1. Memberikan gambaran yang sistematis tentang peristiwa-peristiwa yang merupakan hasil dari penelitian atau observasi.
2. Data lebih cepat ditangkap, dipahami, dan dimengerti.
3. Memudahkan observer dalam membuat analisis data.
4. Membuat proses pengambilan kesimpulan dan keputusan lebih tepat, cepat, dan akurat.

## 4.5 Teknik Penyajian Data

Pada dasarnya banyak cara untuk menyajikan data sehingga ia dapat dipahami dan digunakan secara tepat oleh pengolah data. Namun untuk menghasilkan gambaran data yang komunikatif, harus diingat untuk menyajikan sesuai kebutuhan. Penyajian data dilakukan untuk menganalisis masalah agar mudah dicari pemecahannya. Cara penyajian data dapat dilakukan dengan narasi atau teks, tabel dan grafik (Rasdihan, 2009).

### 4.5.1 Narasi atau Teks

Penyajian data dalam bentuk narasi atau secara teks adalah penyajian data hasil penelitian dalam bentuk kalimat. Misalnya, penyebaran suatu penyakit di daerah pantai lebih tinggi bila dibandingkan dengan penduduk pedesaan. Data yang disajikan merupakan gambaran umum tentang kesimpulan tentang hasil pengamatan. Penyajian dalam bentuk teks banyak digunakan dalam bidang sosial, ekonomi, psikologi dan lain-lain, dan berperan sebagai laporan hasil penelitian kualitatif (Furqoni, 2021).

### 4.5.2 Tabel

Tabel merupakan susunan data dalam bentuk baris dan kolom. Penyajian data dalam bentuk tabel berarti mengumpulkan data-data ke dalam kelompok yang sama pada suatu baris atau kolom, sehingga setiap kelompok memiliki frekuensi (jumlah) (Subchan, 2015). Tabel merupakan kumpulan angka-angka yang disusun menurut kategori-kategori (misalnya: jumlah pegawai menurut pendidikan dan masa kerja) sehingga memudahkan dalam pembuatan analisis data. Penyajian data dalam bentuk tabel bertujuan untuk memberikan

informasi dan gambaran mengenai jumlah secara terperinci sehingga memudahkan pengolah data dalam menganalisis data tersebut.

Penggunaan tabel sangatlah luas. Tabel muncul di media cetak, catatan tulisan tangan, perangkat lunak komputer, ornamen arsitektur, rambu lalu lintas dan banyak tempat lainnya. Banyak cara untuk menggambarkan tabel, tergantung pada fungsinya. Jenis tabel yang sering digunakan untuk keperluan analisis data antara lain tabel baris-kolom dan tabel distribusi frekuensi.(Sudibyo, 2020). Menurut Nuryadi, dkk (2017), penyajian tabel digunakan sebagai pilihan yang sering dipakai oleh peneliti atau penyaji informasi. Pengolahan data untuk keperluan analisis awal atau analisis lanjutan akan lebih baik apabila disajikan terlebih dahulu dalam tabel yang baik.

Berikut contoh sebuah tabel sebagai bahan untuk menjelaskan komponen tabel seperti pada tabel 4.1.

**Tabel 4.1:** Jumlah penduduk putus sekolah untuk SD/MI di Desa X Tahun 2007 – 2009

Tahun	Jumlah
2007	115
2008	121
2009	132
Jumlah	368

Keterangan tabel

1. Nomor tabel, diatas judul tabel terdapat nomor tabel yaitu 4.1. bila tabel yang disajikan lebih dari satu maka hendaknya diberi nomor agar mudah untuk mencari kembali bila dibutuhkan.
2. Judul tabel, di atas tabel dituliskan judul tabel. Judul tabel memuat informasi mengenai: data serta tempat dan waktu pengumpulannya.
3. Baris, tabel tersebut mempunyai baris 2007 – 115, 2008 – 121, 2009 – 132 dan jumlah – 368.
4. Kolom, tabel di atas mempunyai kolom tahun dan jumlah penduduk putus SD/MI.

5. Sel adalah data yang menjadi pertemuan baris dan kolom, yaitu 155, 121, 132 dan 368.
6. Sumber adalah asal dari mana data dikutip. Sumber merupakan pihak yang melakukan pengumpulan data. Jika tabel tidak memuat sumber berarti data dikumpulkan dan ditabulasikan sendiri oleh pembuat tabel.

### 4.5.3 Grafik

Grafik digunakan untuk menampilkan data numerik dan menggambarkan hubungan antara beberapa variabel. Tampilan grafik dapat menjelaskan pola data dan memperlihatkan pengaruh perubahan satu variabel terhadap variabel lainnya. Ada beberapa macam grafik yang sering digunakan yaitu grafik bar, pie, garis, histogram dan scatter (Sudibyo, 2020).

Penyajian data dalam bentuk grafik umumnya lebih menarik perhatian dan mengesankan. Penyajian data statistik secara grafis mempunyai berbagai fungsi. Grafik atau diagram seringkali digunakan dalam iklan dengan maksud agar konsumen memperoleh kesan yang mendalam terhadap ciri-ciri produk yang diiklankan. Kegiatan produksi lebih mudah dilihat dan dipelajari secara visual bila dinyatakan dalam angka-angka dan digambarkan secara grafis. Peta pengawasan kualitas merupakan alat yang penting dalam melakukan pengawasan produk maupun pengawasan proses produksi. Grafik penjualan suatu perusahaan memberi gambaran yang sederhana dan menarik mengenai perkembangan hasil penjualan yang telah dicapai oleh perusahaan yang bersangkutan. Pada hakekatnya grafik dan tabel seyogyanya digunakan secara bersama-sama. Grafik statistik lebih mudah dan menarik dibanding tabel statistik. Selain itu, grafik dapat melukiskan suatu peristiwa secara lebih mengesankan dan tidak membosankan. Namun demikian, penyajian secara grafis hanyalah bersifat aprosimatif. Angka-angka yang pasti dan rinci tentang suatu peristiwa dimuat dalam tabel. Oleh karena itu, analisis dan interpretasi data umumnya dilakukan terhadap data yang terdapat dalam tabel statistik (Nuryadi, 2017).

Menurut Nuryadi, dkk (2017), jenis-jenis grafik statistik dibedakan menjadi:

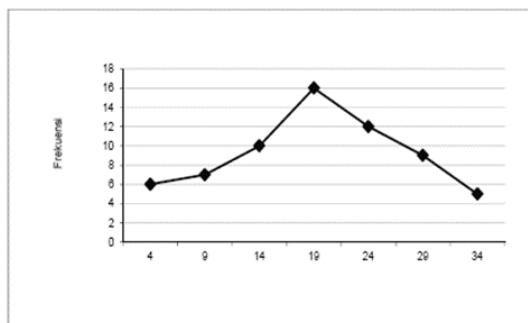
1. Diagram garis (line chart)

Diagram garis sering disebut juga peta garis (line chart) atau kurva (curve), merupakan bentuk penyajian yang paling banyak dipakai

dalam berbagai laporan perusahaan maupun penelitian ilmiah. Data statistik dapat diklasifikasikan atas ciri-ciri kronologis, geografis, kuantitatif maupun kualitatif. Salah satu bentuk data yang dapat diklasifikasi secara kronologis adalah data deret berkala (time series). Sebagian besar distribusi data dapat diklasifikasi secara kuantitatif dalam bentuk distribusi frekuensi. Hasil kedua cara klasifikasi tersebut dapat digambarkan secara visual dalam bentuk kurva. Sedangkan data yang diklasifikasikan berdasarkan geografis maupun kualitatif, jarang digambarkan dalam bentuk kurva. Data demikian dapat digambarkan dengan peta balok (bar chart) atau bentuk peta lainnya.

Grafik menunjukkan perilaku sepanjang waktu, digunakan untuk mengkomunikasikan bagaimana satu variabel memengaruhi atau tergantung pada variabel yang lain.

Contoh diagram garis bisa dilihat pada gambar 4.1.

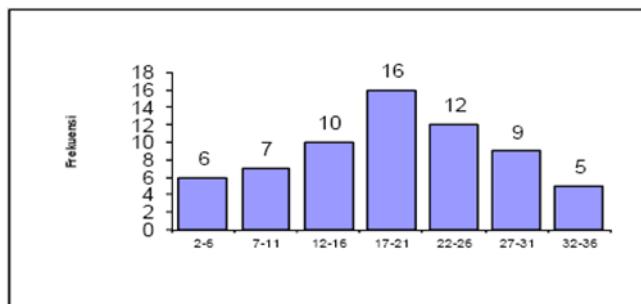


**Gambar 4.1:** Contoh diagram grafik (a) (Nuryadi, 2017)

## 2. Diagram batang (bar chart)

Grafik batang menunjukkan jumlah dari suatu kategori, digunakan untuk mengkomunikasikan variabel kuantitatif dan kualitatif variabel. Grafik batang menampilkan hasil pengukuran dari beberapa variabel. Diagram batang digunakan untuk memahami persoalan secara visual. Dalam diagram batang, lebar kelas diambil dari selang kelas distribusi frekuensi, sedangkan frekuensi masing-masing kelas ditunjukkan oleh tinggi batang.

Contoh diagram batang bisa dilihat pada gambar 4.2.

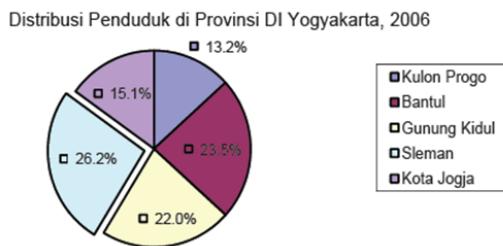


**Gambar 4.2:** Contoh diagram grafik (b) (Nuryadi, 2017)

### 3. Diagram lingkaran (pie chart)

Diagram lingkaran biasa digunakan untuk menyatakan perbandingan jika data terdiri atas beberapa kelompok. Diagram lingkaran digunakan untuk menggambarkan fraksi atau persentase, sebagian dari keseluruhan.

Contoh diagram lingkaran bisa dilihat pada gambar 4.3.



**Gambar 4.3:** Contoh diagram grafik ©(Nuryadi, 2017)

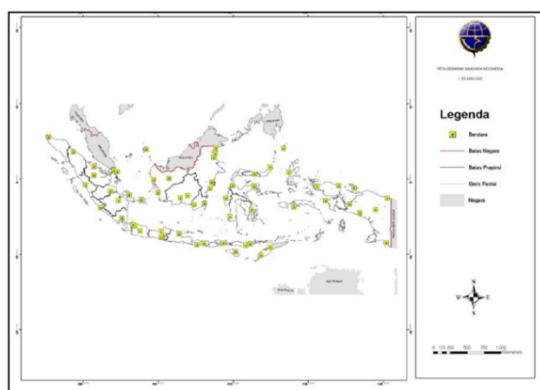
## 4.6 Penyajian Data Spasial

Menurut Bambang Widjanarko Otok dan Dewi Juliah Ratnaningsih (2016), spasial, secara bahasa berkenaan dengan ruang (space) atau tempat. Dengan demikian, secara sederhana data spasial dapat diartikan sebagai data yang

memiliki referensi keruangan (geografi). Setiap bagian dari data tersebut selain memberikan gambaran tentang suatu fenomena, juga selalu dapat memberikan informasi mengenai lokasi dan juga persebaran dari fenomena tersebut dalam suatu ruang (wilayah). Bentuk yang paling tepat dengan cara penyajian data spasial adalah peta.

Penyajian data dalam bentuk peta pada dasarnya dilakukan dengan mengikuti kaidah-kaidah kartografis yang pada intinya menekankan pada kejelasan informasi tanpa mengabaikan unsur estetika dari peta sebagai sebuah karya seni. Kartografis itu sendiri merupakan studi dan praktik membuat peta atau globe.

Contoh penyajian data spasial bisa dilihat pada gambar 4.4.



**Gambar 4.4:** Peta sebaran bandara di Indonesia 2008-2009

## **Bab 5**

# **Populasi dan Sampel**

### **5.1 Pendahuluan**

Konsep mengenai populasi dan sampel merupakan sebuah konsep yang sangat penting dipahami setiap peneliti. Memahami setiap perbedaan antara populasi dan sampel itu sangat mudah. Kita harus mengingat salah satu hukum dasar statistik yakni sampel merupakan suatu kelompok yang selalu lebih kecil dalam suatu populasi tertentu. Populasi dalam suatu penelitian tidaklah harus selalu menggunakan populasi manusia. Melainkan bisa juga berupa parameter data apapun yang memiliki suatu sifat yang sama. Dalam suatu riset dan statistik, setiap studi memiliki pertanyaan penting. Pengamatan dan percobaan sampel dan populasi menentukan suatu hasil penyelidikan ini. Hal tersebut dilakukan untuk mendapatkan wawasan yang menjelaskan mengenai fenomena dalam seluruh populasi.

Amirullah (2017) menjelaskan syarat utama dikatakan sampel itu baik, yaitu apabila sampel itu memiliki sifat yang representatif. Untuk memenuhi syarat tersebut diperlukan suatu teknik pengambilan sampel yang baik pula. Pengambilan sampel dalam penelitian dapat dilakukan dengan berbagai teknik. Adapun teknik pengambilan sampel yang secara umum dapat dikelompokkan menjadi dua teknik, yaitu teknik *nonprobability sampling* dan *probability sampling*.

Menurut Supardi (1993) penentuan populasi dan sampel penelitian menjadi sangat penting, karena hasil penelitian pada umumnya akan mengambil kesimpulan secara luas atau menggeneralisasi hasil penelitian. Ketepatan serta keakuratan dalam suatu penentuan populasi dan sampel penelitian memberikan bobot dan kualitas hasil penelitian. Penentuan populasi terhadap suatu penelitian akan memberikan kebenaran terhadap generalisasi kesimpulan hasil penelitian yang diperoleh nantinya. Oleh karena itu, mempersiapkan desain penelitian peneliti harus mampu benar-benar menentukan secara baik populasi dan sampel penelitian.

Menurut Susilana (2020) seorang peneliti harus dapat memahami konsep populasi dan sampel untuk dapat melaksanakan penelitian dengan baik. Populasi adalah keseluruhan objek maupun subjek penelitian, sedangkan sampel merupakan sebagian atau wakil populasi yang memiliki karakteristik representasi dari populasi. Untuk menentukan atau menentukan banyaknya sampel yang diperlukan, pemahaman yang baik dari peneliti mengenai sampling, baik dalam hal penentuan jumlah maupun dalam menentukan sampel yang mana yang diambil. Kesalahan dalam menentukan populasi akan mengakibatkan tidak tepatnya data yang dikumpulkan sehingga hasil penelitian pun tidak memiliki kualitas yang baik, tidak representatif, dan tidak memiliki daya generalisasi yang baik (Sugiyono, 2019).

Untuk lebih jelasnya, perbedaan populasi dan sampel adalah, jika populasi merupakan bagian besar dari sampel dan merupakan objek penelitian. Sementara objek yang akan dijadikan riset secara langsung berupa sampel.

## 5.2 Populasi

### 5.2.1 Pengertian Populasi

Dalam ilmu statistik, populasi merupakan suatu kumpulan data yang memiliki karakteristik yang juga sama dan menjadi suatu objek inferensi. Inferensi statistik didasarkan pada dua konsep dasar, yaitu data penduduk secara keseluruhan, baik nyata atau imajinasi dan sampel, sebagai bagian dari populasi yang digunakan untuk membuat kesimpulan (pendekatan atau penggambaran) dari populasi di mana ia berasal (Kurniawan, 2021).

Kata populasi pada umumnya merujuk pada sekumpulan individu yang memiliki karakteristik khas dan menjadi suatu perhatian dalam penelitian atau pengamatan.

Populasi adalah suatu kesatuan individu atau subjek pada wilayah dan waktu dengan kualitas tertentu yang akan diteliti atau diamati (Supardi, 1993). Populasi adalah keseluruhan dari kumpulan elemen yang memiliki sejumlah karakteristik umum, yang terdiri dari bidang-bidang untuk diteliti atau populasi adalah keseluruhan kelompok dari orang-orang, peristiwa atau barang-barang yang diminati oleh peneliti untuk diteliti (Amirullah, 2017). Dengan demikian, populasi merupakan seluruh kumpulan elemen yang dapat digunakan untuk membuat beberapa kesimpulan.

Menurut Sugiyono (2016), dalam bukunya mengemukakan mengenai Populasi adalah wilayah generalisasi yang terdiri atas objek atau subjek yang mempunyai kualitas dan karakteristik tertentu yang ditetapkan oleh penelitian untuk dipelajari yang kemudian ditarik kesimpulannya.

Uma & Bougie (2017) menjelaskan populasi mengacu pada seluruh kelompok masyarakat, peristiwa atau hal yang menarik untuk diteliti oleh peneliti. Dari pengertian di atas, dapat disimpulkan bahwa populasi bukan hanya sekedar jumlah yang ada pada objek atau subjek yang dipelajari, akan tetapi meliputi seluruh karakteristik atau sifat yang dimiliki oleh subjek atau objek tersebut. Hanlon & Large (2011) menjelaskan bahwa suatu populasi didefinisikan sebagai sekelompok individu dari spesies yang sama dalam wilayah tertentu. Anggota suatu populasi sering bergantung pada sumber daya yang sama serta sebuah populasi adalah semua individu atau unit yang menarik.

Populasi sebenarnya juga dapat dikatakan sebagai totalitas dari semua objek yang pada nantinya akan diteliti. Menurut Kurniawan (2021) sampel dianggap mewakili populasi. Sampel diambil dari populasi tidak dapat digunakan untuk mewakili populasi lain. Sensus dilakukan untuk mendapatkan karakteristik populasi nyata. Karakteristik yang dimiliki oleh parameter populasi disebut. Untuk beberapa karakteristik statistik dari sampel, nilai parameter adalah nilai harapan atau nilai yang diharapkan.

Populasi sebenarnya merupakan sebuah kelompok objek pengamatan atau riset yang pada nantinya dapat berbentuk manusia, barang hingga barang dan masih banyak lagi (Adytya, 2019). Biasanya sebagai bahan yang digunakan untuk riset, populasi ini mempunyai karakteristik tertentu, lengkap dan jelas.

Sehingga Perlu dipahami bahwa populasi dapat pula disebut sebagai keseluruhan dari semua objek yang diteliti.

### 5.2.2 Jenis- Jenis Populasi

Menurut Saddoen (2021) bahwa secara umum, jenis populasi dibagi menjadi tiga bagian, yakni populasi yang berdasarkan jumlahnya, populasi berdasarkan sifatnya dan juga populasi berdasarkan sifat lain.

Adapun jenis-jenis populasi ini akan dijelaskan secara lengkap di bawah ini:

#### 1. Populasi Berdasarkan Jumlahnya

Berdasarkan jumlahnya, populasi dibagi menjadi dua jenis yakni populasi terbatas dan populasi tak terbatas. Populasi terbatas adalah suatu sumber daya yang dapat dihitung karena terdapat batasan secara kuantitatif. Adapun contoh populasi ini adalah pada tahun 1985 hanya ada 3 wanita saja yang mengikuti program KB. Sedangkan populasi tak terbatas adalah suatu sumber daya yang tidak dapat dinyatakan bentuk dan jumlahnya karena jumlahnya tak terbatas. Contohnya: jumlah dari narapidana di Indonesia sangatlah banyak.

#### 2. Populasi Berdasarkan Sifatnya

Berdasarkan sifatnya, populasi dibagi menjadi dua yakni populasi homogen dan populasi heterogen. Populasi homogen adalah populasi yang unsurnya memiliki sifat yang sama, sehingga jika dilihat secara kuantitatif jumlahnya tidak dipermasalahkan. Sedangkan populasi heterogen adalah populasi yang dalam unsurnya terdapat sifat variasi sehingga ada batasan baik secara kuantitatif maupun secara kualitatif.

#### 3. Populasi Berdasarkan Perbedaan Lain

Jenis populasi ini juga dibagi menjadi dua, yakni populasi target dan populasi survey. Populasi target adalah populasi yang ditentukan sesuai dengan yang tertera dalam masalah penelitian. sedangkan populasi survey adalah populasi yang terliput di dalam penelitian yang sedang dilaksanakan.

Sedangkan Hayati (2020) menjelaskan bahwa secara umum terdapat dua jenis populasi dalam sebuah penelitian, yakni:

1. Populasi Target (Target Population)

Populasi target mengacu pada seluruh kelompok individu atau objek yang menarik minat peneliti untuk menggeneralisasi kesimpulan. Populasi target biasanya memiliki karakteristik yang berbeda-beda dan disebut juga dengan populasi teoritis.

2. Populasi yang Dapat Dicapai (Accessible Population)

Populasi yang dapat dicapai adalah populasi dalam penelitian yang kesimpulannya dapat diterapkan oleh peneliti. Populasi ini adalah bagian dari populasi target dan disebut juga populasi studi. Dari populasi yang dapat diakses itulah peneliti menarik sampel mereka.

Menurut Silalahi (2003) dalam (Niningprastiwi, 2017) Jenis populasi dapat dilihat dari beberapa sudut, yaitu:

1. Berdasarkan jumlah populasi

Berdasarkan jumlah populasi, populasi dibagi menjadi dua, yaitu:

- a. Populasi terbatas (finite)

- Sumber data jelas batasannya secara kuantitatif, sehingga relatif dapat dihitung jumlahnya.

• Memiliki ciri terbatas. Contoh: Murid di suatu sekolah, atau mahasiswa di sesuatu fakultas, dan karyawan serta guru/dosen yang ada di situ, jelas merupakan sesuatu yang bisa dan mudah dihitung. Bahkan di suatu kecamatan, kabupaten, provinsi, bahkan nasional pun masih bisa dan mudah dihitung (walau mungkin tidak tepat benar).

- b. Populasi tak terbatas (infinite)

Sumber daya yang tidak dapat ditentukan batasannya.

Contoh: Jumlah orang yang suka (sukarela) mengikuti pengajian di sesuatu pondok pesantren besar (lebih-lebih di sekitar banyak pesantren) mungkin menjadi tak terhingga, karena kehadirannya secara terus-menerus.

## 2. Berdasarkan sifat populasi

Berdasarkan sifatnya, populasi dibagi menjadi dua, yaitu:

### a. Populasi homogen

Sumber data yang unsurnya memiliki sifat yang sama. Contoh: Darah, misalnya, termasuk yang memiliki kesamaan sifat atau kondisi (berkaitan dengan golongan darah) di seluruh tubuh, pekerjaan yang dibatasi (misalnya: Ibu Rumah Tangga), maka subjek penelitian homogen untuk jenis pekerjaan.

### b. Populasi heterogen

- Sumber data yang unsurnya memiliki sifat yang bervariasi sehingga perlu ditempatkan batas-batasnya baik secara kualitatif dan kuantitatif.
- Ketidaksamaan itu dapat terjadi antara lain karena di antara anggota-anggotanya ada perbedaan dari aspek sebagai berikut:

## 3. Strata atau lapisan.

Misalnya: status ekonomi (perbedaan pemilikan harta benda): ada miliarder, jutawan, menengah, miskin, dan di bawah garis kemiskinan; tingkat pendidikan (tingkat pendidikan formal yang pernah ditempuh): ada yang berpendidikan Perguruan Tinggi, Sekolah Menengah Atas, Sekolah Menengah Pertama, dan Sekolah Dasar; lapisan kemasyarakatan atau sosial: ada kelompok elite, menengah, dan bawah. Tingkatan usia: pada bayi, anak-anak, remaja, orang dewasa, dan lansia; tingkatan kelas di sekolah: ada Kelas XII, XI, X SMA; Kelas IX, VIII, VII SMP; dan Kelas VI, V, IV, III, II, I SD.

## 4. Cluster [klaster] atau golongan, dan juga gugus atau kelompok.

Misalnya: golongan berdasarkan pemelukannya agama: ada yang beragama Islam, Katolik, Kristen, Hindu, Budha, dan Kong Hu Cu. Jenis kelamin: ada laki-laki dan perempuan. Pekerjaan: ada petani, PNS, pedagang, buruh bangunan, pegawai swasta, wirausaha dan sebagainya. Kelompok atau gugus: guru di satu sekolah, murid di satu kelas, sekolah di satu gugus sekolah. Ada orang yang

menyamakan cluster dengan strata, maksudnya sebutan strata sama dengan cluster (di dalamnya tercakup baik lapisan, maupun golongan).

5. Area (wilayah), geografis dan atau administratif (juga ada yang menyebutnya strata). Misalnya: geografis: ada desa, pinggiran kota, kota, dan metropolitan. Administratif: ada desa atau kelurahan, kecamatan, kabupaten dan provinsi.

Heterogenitas (keberagaman) tersebut perlu diperhatikan dalam pengambilan sampel manakala diduga atau diperkirakan akan membawa perbedaan terhadap hasil penelitian (sesuai objek yang diteliti). Misalnya, jika dianggap jenis kelamin tidak berkaitan dengan prestasi belajar, maka unsur jenis kelamin itu tidak perlu diperhatikan dalam pengambilan sampel penelitian yang akan meneliti tentang prestasi belajar. Maksudnya adalah tidak harus unsur jenis kelamin laki-laki terwakili, perempuan juga terwakili.

Berdasarkan perbedaan lain, dibagi menjadi:

1. Populasi target, yaitu jenis populasi yang telah ditentukan sesuai dengan masalah penelitian. Populasi target adalah populasi yang dengan alasan yang kuat (reasonable) memiliki kesamaan dengan populasi terukur (Sukmadinata, 2009). Populasi target adalah seluruh populasi yang ada dalam ini, jumlahnya tak terbatas karena tidak dibatasi tempat dan waktu. Populasi target adalah unit di mana suatu hasil penelitian akan diterapkan atau digeneralisir.

Contoh: Populasi anak usia 5 tahun.

2. Populasi terukur atau terjangkau (accessible population)  
Populasi terjangkau adalah populasi yang secara riil dijadikan dasar dalam penentuan sampel dan secara langsung menjadi lingkup sasaran keberlakuan kesimpulan (Sugiyono, 2009). Populasi terjangkau adalah populasi yang terukur karena dibatasi oleh tempat dan waktu.

Populasi terjangkau adalah merupakan bagian dari populasi target, di mana peneliti mampu menjangkaunya, karena dibatasi oleh karakteristik demografi

(letak wilayah), waktu untuk menjangkau seluruh anggota populasi, ketersediaan dana untuk melaksanakan penelitian pada seluruh anggota populasi, ketersediaan sumber daya manusia sebagai pelaksana penelitian (Dharma, 2011).

Contoh: Kemampuan bahasa anak usia 5 tahun di sebuah kabupaten X Tahun 2016. Karena tingkat kecerdasan, kematangan berbahasa, usia, lingkungan dan status sosial ekonomi, anak-anak di kabupaten X sama dengan di Kabupaten Y. Kesimpulannya adalah kemampuan berbahasa anak usia 5 tahun di kabupaten X berlaku untuk Kabupaten Y.

## 5.3 Sampel

### 5.3.1 Pengertian Sampel

Menurut Sugiyono (2019) sampel adalah bagian dari jumlah dan karakteristik yang dimiliki oleh populasi, sehingga sampel merupakan bagian dari populasi yang ada dan pengambilan sampel harus menggunakan cara tertentu yang didasarkan oleh pertimbangan serta syarat-syarat yang ada. Sedangkan Sujarweni (2015) menjelaskan sampel adalah bagian dari sejumlah karakteristik yang dimiliki oleh populasi yang digunakan untuk penelitian. Sampel juga diambil dari populasi yang benar-benar mewakili dan valid yaitu dapat mengukur sesuatu yang seharusnya diukur.

Sampel merupakan bagian kecil dari populasi itu sendiri yang diambil sebagai objek dalam sebuah pengamatan atau penelitian lantaran dianggap mampu mewakili populasi (Adytya, 2019). Sampel adalah pemilihan acak anggota suatu populasi. Ini adalah kelompok yang lebih kecil yang diambil dari populasi yang memiliki karakteristik dari seluruh populasi. Pengamatan dan kesimpulan yang dibuat terhadap data sampel dikaitkan dengan populasi secara keseluruhan (Momoh, 2021).

Sampel menurut Uma & Bougie (2017) adalah bagian dari populasi. Terdiri dari beberapa anggota yang dipilih dari populasi. Dengan kata lain, beberapa, tapi tidak semua elemen populasi dari sampel. Untuk menentukan besarnya sampel tersebut bisa dilakukan, baik secara statistik maupun berdasarkan estimasi penelitian, selain itu juga perlu diperhatikan bahwa sampel yang

dipilih harus representative artinya segala karakteristik populasi hendaknya tercermin dalam sampel yang dipilih.

Menurut Arikunto (2006) sampel adalah sebagian dari jumlah dan karakteristik yang dimiliki oleh populasi. Bila populasi besar, dan peneliti tidak mungkin mempelajari semua yang ada pada populasi, misalnya karena keterbatasan dana, tenaga dan waktu, maka peneliti dapat menggunakan sampel yang diambil dari populasi itu. Apa yang dipelajari dari sampel itu, kesimpulannya akan diberlakukan untuk populasi. Untuk itu sampel yang diambil dari populasi harus betul-betul representatif.

## 5.3.2 Teknik Pengambilan Sampel

### Teknik Sampling Secara Probabilitas

Probability sampling menurut Sugiyono (2019) adalah teknik sampling yang memberikan peluang yang sama bagi setiap unsur (anggota) populasi untuk dipilih menjadi anggota sampel. Menurut Retnawati (2017) probability sampling adalah teknik pengambilan sampel yang memberikan peluang yang sama bagi setiap unsur (anggota) populasi untuk dipilih menjadi anggota sampel. Teknik ini merupakan teknik yang memungkinkan peneliti atau evaluator untuk membuat generalisasi dari karakteristik sampel menjadi karakteristik populasi. Probability sampling dapat digolongkan menjadi *simple random sampling*, *systematic random sampling*, *stratified sampling*, dan *cluster sampling*, seperti yang dijelaskan berikut ini.

#### Simple Random Sampling

Penyampelan acak sederhana, dimaksudkan bahwa sebanyak ( $n$ ) sampel diambil dari populasi  $N$  dan tiap anggota populasi mempunyai peluang yang sama untuk terambil.

Terdapat 3 (tiga) cara untuk menentukan sampel dengan menggunakan teknik ini, yaitu:

1. Cara undian;
2. Cara tabel bilangan random;

Contoh: Diketahui  $N = 1000$ , akan dipilih  $n = 20$  dengan menggunakan teknik simple random sampling.

Solusi: Misalnya ke-1000 data tersebut adalah 001,002,003,...,999,000 dengan 000 adalah data ke-1000. Pertama-tama, tentukan aturan penggunaan tabel random, misal dimulai dari kolom pertama baris pertama sampai baris ke 20. Jadi didapatkan 104, 213,243, ..., 070. (Scheaffer, 1986:43). Dengan menggunakan komputer untuk mengacak, misalnya dengan bantuan SPSS.

### **Stratified Random Sampling**

Pada penyampelan jenis ini, anggota populasi dikelompokkan berdasarkan stratanya, misalnya tinggi, sedang, dan rendah. Kemudian dipilih sampel yang mewakili masing-masing strata.

Langkah-langkah dalam menentukan *Stratified Random sampling*:

1. Menentukan data pendukung tentang populasi yang diambil berikut strata-strata yang ada di dalamnya;
2. Mengklasifikasikan populasi ke dalam grup atau strata yang saling lepas;
3. Menentukan ukuran sample untuk tiap stratum;
4. Memilih secara acak setiap stratum dengan menggunakan simple random sampling.

Contoh: Sebuah evaluasi dilakukan untuk mengetahui pelaksanaan program pembelajaran kesehatan. Populasi yang diambil adalah seluruh sekolah menengah atas di 34 provinsi di Indonesia, misalnya 340 sekolah.

Solusi: Langkah pertama yang dilakukan yakni membagi sekolah di tiap provinsi berdasarkan hasil UN dengan strata, yaitu strata dengan nilai UN tinggi, nilai sedang, dan nilai rendah. Masing-masing provinsi 10 sekolah.

### **Sistematic Sampling**

Penyampelan dengan cara ini dilakukan dengan mengurutkan terlebih dahulu semua anggota, kemudian dipilih urutan tertentu untuk dijadikan anggota sampel.

### **Cluster Sampling**

Pada penyampelan jenis ini, populasi dibagi menjadi wilayah atau klaster. Jika terpilih klasternya, seluruh anggota dalam klaster tersebut yang menjadi sampel.

Langkah-langkah dalam pengambilan sampel dengan cluster sampling:

1. Menentukan cluster-cluster nya;
2. Menentukan banyak cluster yang akan dijadikan sampel;
3. Memilih secara acak cluster sebanyak cluster;
4. Semua anggota yang terdapat dalam klaster yang terpilih merupakan sampel studi atau penelitian atau evaluasi.

Contoh:

Sebuah evaluasi tentang tingkat kesehatan siswa SMA akan melibatkan seluruh SMA yang ada di Indonesia. Ada 34 provinsi, sehingga dapat dibagi menjadi 34 cluster. Misalnya akan diambil sebanyak 7 klaster, maka dipilih secara acak 7 propinsi dari 34 provinsi. Semua SMA yang berasal dari 7 provinsi tersebut merupakan sampel.

### **Teknik Sampling Secara Non Probabilitas**

Menurut Sugiyono (2019) nonprobability sampling adalah teknik yang tidak memberi peluang atau kesempatan yang sama bagi setiap unsur atau anggota populasi untuk dipilih menjadi sampel. Menurut (Retnawati, 2017) nonprobability sampling adalah teknik pengambilan sampel yang tidak memberi peluang atau kesempatan sama bagi setiap unsur atau anggota populasi yang dipilih menjadi sampel. Teknik pengambilan sampel ini di antaranya sampling incidental, sampling bertujuan, sampling bola salju (snowball sampling), dan sampling kuota. Non probability sampling ini tidak bisa digunakan untuk membuat generalisasi. menurut Retnawati (2017), teknik sampel ini meliputi:

#### **Sampling Insidental (Reliance Available Sampling)**

Teknik sampling ini mengandalkan pada keberadaan subjek untuk dijadikan sampel yaitu siapa saja yang secara kebetulan bertemu dengan peneliti dan dipandang cocok sebagai sumber data maka subjek tersebut dijadikan sampel. Sebagai contoh misalnya suatu penelitian dilakukan untuk mengevaluasi pemanfaatan media komputer pada proses pembelajaran. Sampel yang akan diambil yaitu guru yang secara kebetulan bertemu dengan peneliti dan dianggap cocok oleh peneliti untuk dijadikan sumber data. Pengambilan sampling semacam ini tidak dapat digunakan untuk membuat generalisasi sifat sampel menjadi sifat populasi.

### **Sampling Purposive (Purposive or Judgment Sampling)**

Sampling purposive adalah teknik penentuan sampel berdasarkan pertimbangan peneliti atau evaluator tentang sampel mana yang paling bermanfaat dan representative (Babbie, 2004). Terkadang sampel yang akan diambil ditentukan berdasarkan pengetahuan tentang suatu populasi, anggotanya dan tujuan dari penelitian. Jenis sampel ini sangat baik jika dimanfaatkan untuk studi awal untuk penelitian atau evaluasi, yang kemudian diikuti oleh penelitian lanjutan yang sampelnya diambil secara acak (random).

Contoh: Suatu evaluasi dilakukan untuk mengevaluasi pembiasaan pola hidup sehat yang digunakan di Sekolah Luar Biasa-B (tunarungu). Dalam hal ini, sekolah-sekolah yang dijadikan sampel yakni SLB-B, yang ditetapkan sesuai tujuan evaluasi.

### **Sampling Bola Salju (Snowball Sampling)**

Sampling snowball dapat dilakukan jika keberadaan dari suatu populasi sulit untuk ditemukan. Dengan kata lain, cara ini banyak dipakai ketika peneliti atau evaluator tidak banyak tahu tentang populasi penelitian atau evaluasinya. Pada sampling bola salju, peneliti mengumpulkan data dari beberapa sampel yang dapat ditemukan oleh peneliti sendiri, selanjutnya peneliti meminta individu yang telah dijadikan sampel tersebut untuk memberitahukan keberadaan anggota yang lainnya yang tidak dapat ditemukan oleh peneliti untuk dapat melengkapi data (Babbie, 2004). Pada penelitian kualitatif banyak menggunakan sampel purposive dan snowball.

Sebagai contoh misalnya evaluasi dilakukan untuk mengetahui efektivitas bidan desa yang diprogramkan di suatu daerah. Salah satu orang yang dapat dijadikan sumber data adalah salah satu sesepuh dari masyarakat tersebut, dan ditanyai perlunya bidan desa. Selanjutnya dari sesepuh yang dijadikan sampel tersebut diminta untuk memberikan informasi tentang keberadaan anggota masyarakat yang lain yang dapat dijadikan sumber data.

### **Sampling Quota**

Teknik sampling kuota adalah teknik menentukan sampel dari populasi yang mempunyai ciri-ciri tertentu sampai jumlah (kuota) yang diinginkan. Pada sampling kuota, dimulai dengan membuat tabel atau matriks yang berisi penjabaran karakteristik dari populasi yang ingin dicapai atau karakteristik populasi yang sesuai dengan tujuan dari penelitian untuk selanjutnya ditentukan sampel yang memenuhi ciri-ciri dari populasi tersebut.

Prosedur yang dalam sampling kuota:

1. Pertama, populasi dibagi-bagi menjadi strata yang relevan seperti usia, jenis kelamin, lokasi, dan sebagainya.
2. Proporsi tiap strata diperkirakan atau ditentukan berdasarkan data eksternal kemudian total sampel dibagi-bagi sesuai proporsi ke tiap strata (kuota).
3. Untuk memenuhi jumlah sampel untuk tiap strata, peneliti menggunakan expert judgement.

## 5.4 Menentukan Ukuran Sampel

Ukuran sampel (sample size) adalah banyaknya individu, subjek atau elemen dari populasi yang diambil sebagai sampel. Jika ukuran sampel yang diambil terlalu besar atau terlalu kecil maka akan menjadi masalah dalam penelitian itu. Oleh karena itu, ukuran sampel harus betul-betul diperhatikan oleh peneliti dalam melakukan penelitiannya (Amirullah, 2017).

Berikut ini disajikan pendapat beberapa ahli tentang ukuran sampel;

1. Gay & Diehl (1992) berpendapat bahwa sampel haruslah sebesar-besarnya. Pendapat ini mengasumsikan bahwa semakin banyak sampel yang diambil, maka akan semakin representatif, dan hasilnya dapat digeneralisir. Namun, ukuran sampel yang dapat diterima akan sangat bergantung pada jenis penelitiannya:
  - a. apabila penelitiannya bersifat deskriptif, maka sampel minimumnya adalah 10% dari populasi,
  - b. penelitian yang bersifat korelasional, sample minimumnya 30 subjek,
  - c. penelitian kausal atau perbandingan, sampelnya sebanyak 30 subjek pergroup, dan
  - d. penelitian eksperimental, sampel minimumnya adalah 15 subjek per group.

2. Roscoe (1975) memberikan panduan untuk menentukan ukuran sampel:
  - a. Pada setiap penelitian, ukuran sampel harus berkisar antara 30 dan 500.
  - b. Apabila faktor yang digunakan dalam penelitian itu banyak, maka ukuran sampel minimal 10 kali atau lebih dari jumlah faktor.
  - c. Jika sampel akan dipecah-pecah menjadi beberapa bagian, maka ukuran sampel minimum 30 untuk tiap bagian yang diperlukan.
3. Slovin (1960) menentukan ukuran sampel dari suatu populasi dengan rumus sebagai berikut:

$$\frac{N}{1 + Ne^2}$$

**Gambar 5.1:** Menentukan Ukuran Sampel

Di mana:

$n$  = jumlah sampel

$N$  = ukuran populasi

$e$  = batas kesalahan

4. Frankel & Norman (1993) menyarankan, besar sampel minimum untuk:
    - a. Penelitian deskriptif sebanyak 100.
    - b. Penelitian korelasional sebanyak 50.
    - c. Penelitian kausal atau perbandingan 30 pergroup.
    - d. Penelitian eksperimental sebanyak 30 atau 15.
  5. Malhotra (1993) besarnya jumlah sampel yang diambil dapat ditentukan dengan cara mengalikan jumlah variabel dengan 5, atau  $5 \times$  jumlah variabel. Jika variabel yang diamati berjumlah 20, maka sampel minimalnya adalah 200 yakni ( $5 \times 20$ ).
- Cara menghitung jumlah sampel juga dapat dilakukan melalui:
- a. Teknik Data Kontinyu Cochran

Jika populasi merupakan data yang kontinu maka rumus Cochran (Cobran, 1953), diaplikasikan sebagai berikut:

Misalkan  $\alpha = 10\%$  maka  $Z\alpha$  adalah 1,96 sedangkan  $s$  dari penelitian pilot ditemukan sebesar 3.200 dengan  $e$  dalam value sebesar 500.

b. Teknik (Hair, et al., 1998)

Rumus Hair berlaku bila metode analisis yang digunakan adalah SEM (Structural Equation Modelling).  $N = 5 \times$  variabel operasional penelitian. Jika menggunakan 5 latent variabel dan masing-masing latent variabel dijabarkan ke dalam 4 variabel operasional, sehingga membentuk  $5 \times 4$  variabel operasional, maka jumlah responden yang dibutuhkan minimal adalah:  $N = 5 \times 20 = 100$  responden.

Selanjutnya menurut Hair, et al., (1998) meskipun tidak ada ukuran sampel yang benar maka menyatakan bahwa ukuran sampel yang sesuai adalah antara 100 - 200 sampel.



# **Bab 6**

## **Probabilitas**

### **6.1 Pendahuluan**

Dalam kehidupan keseharian yang kita jalani ini, banyak kejadian-kejadian yang penuh dengan ketidakpastian. Apakah cuaca besok akan hujan atau cerah?, apakah jadwal kereta Api rute Bandung – Jakarta akan datang tepat waktu?, apakah jika dalam pelemparan sebuah koin mata uang akan memberikan hasil sisi Gambar 6?, bagaimana peluang sebuah tim sepak bola masuk kedalam babak selanjutnya jika tim tersebut mengalami menang, seri atau kalah ? Ketidakpastian dalam kejadian – kejadian di atas sangat sering kita hadapi, namun demikian dengan adanya pendekatan konsep probabilitas akan sangat membantu kita dalam melakukan prediksi dari kejadian yang mungkin akan terjadi, sehingga dapat mendekatkan kita kepada derajat kebenaran suatu kejadian dan pada akhirnya akan membantu kita dalam menentukan suatu keputusan.

Secara sejarah, Blaise Pascal (1623-1662) adalah salah satu ilmuwan yang dimintai bantuan oleh temannya Chevalier de Mere untuk menyelesaikan sebuah persoalan berkaitan dengan probabilitas yang awal mulanya berasal dari masalah perjudian. Pascal tertarik sehingga berdiskusi dengan pakar matematika Pierre de Fermat. Buah pemikiran hasil diskusi antara Pascal dan Fermat menjadi dasar munculnya teori probabilitas dikemudian hari (Sudaryono M.Pd, 2012).



**Gambar 6.1:** Foto Blaise Pascal (Wikipedia, 2019)

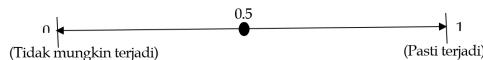
Walaupun berawal dari masalah perjudian, namun teori probabilitas dapat diterapkan dalam berbagai bidang, baik yang bersifat teknik maupun sosial.

## 6.2 Pendekatan Menentukan Probabilitas

Probabilitas atau peluang adalah tingkat keyakinan seseorang untuk menentukan terjadi atau tidak terjadinya suatu kejadian (peristiwa) yang akan terjadi dimasa yang akan datang (Didin Astriani Prasetyowati, 2015). Dalam definisi yang lain probabilitas diartikan sebagai besarnya kesempatan (kemungkinan) suatu peristiwa yang akan terjadi (Sudaryono M.Pd, 2012).

Secara matematika, nilai probabilitas berada dalam rentang nilai 0 dan 1. Suatu kejadian yang mana kejadiannya sudah dipastikan tidak akan pernah terjadi maka probabilitasnya akan mempunyai nilai 0, sebagai contoh probabilitas kejadian manusia bisa terbang tanpa alat bantuan, karena hal tersebut tidak mungkin maka nilai probabilitasnya adalah 0. Sedangkan jika suatu kejadian mempunyai tingkat kejadian yang pasti akan terjadi maka akan mempunyai nilai 1, misalkan probabilitas manusia akan mengalami kematian, karena kejadian tersebut akan pasti terjadi maka nilai probabilitasnya adalah 1. Kasus-kasus yang terjadi dalam kejadian sehari-hari pada umumnya akan bernilai

antara 0 dan 1, sehingga nilai probabilitasnya dapat menjadi acuan dalam melakukan prediksi dan pengambilan keputusan.



**Gambar 6 2:** Peristiwa Probabilitas

Sebagai contoh yang lain dari suatu kejadian A misalkan mempunyai nilai probabilitas 0,95 atau 95%, artinya probabilitas kejadian A akan terjadi adalah 95% dan probabilitas kejadian A tidak akan terjadi adalah 5%. Melalui konsep probabilitas kita akan diberikan petunjuk seberapa besar keyakinan bahwa suatu kejadian akan terjadi. Semakin besar nilai probabilitas, semakin besar terjadinya suatu kejadian. Jika suatu probabilitas didasarkan pada latar belakang ilmiah maka bisa memberikan tingkat keyakinan yang lebih tinggi (Didin Astriani Prasetyowati, 2015).

Probabilitas suatu kejadian A, pada umumnya ditulis  $P(A)$ , dengan memenuhi sifat-sifat berikut (Seymour Lipschutz, 2008):

1. Bernilai antara 0 dan 1.  
 $0 \leq P(A_n) \leq 1$  , untuk setiap n
2. Jumlah probabilitas dari semua kemungkinan hasil eksperimen adalah 1.  
 $P(A_1)+ P(A_2)+ P(A_3)+...+ P(A_n) = 1$

$$\sum_{i=1}^n P(A_i) = 1$$

Sebelum kita membahas lebih dalam tentang probabilitas, maka perlu kita mengenalkan beberapa istilah yang sering berkaitan dengan probabilitas.

### 1. Percobaan

Percobaan adalah aktivitas atau kegiatan pengamatan terhadap suatu kejadian yang memungkinkan terjadi minimal 2 peristiwa atau lebih.

Sebagai contoh misalkan percobaan melempar sebuah koin mata uang, kemungkinan yang bisa muncul adalah sisi Gambar 6 atau angka. Contoh lain, dalam sebuah pertandingan sepakbola yang bisa menghasilkan menang, seri atau kalah. Peristiwa perkiraan cuaca

yang bisa menghasilkan kemungkinan cerah, cerah berawan, hujan ringan, hujan sedang, hujan petir.

## 2. Ruang Sampel

Ruang sampel adalah himpunan semua hasil pada percobaan yang mungkin terjadi.

## 3. Kejadian

Kejadian adalah peristiwa yang terjadi dari sebuah percobaan.

Agar lebih jelas dapat dilihat perbandingan pada tabel 6.1 berikut:

**Tabel 6.1:** Percobaan, Ruang sampel dan Kejadian

Percobaan	Ruang Sampel	Contoh Kejadian
Melempar sebuah koin mata uang	Sisi Gambar 6, sisi angka	Muncul sisi Gambar 6
Melempar sebuah mata dadu	Nilai mata dadu 1,2,3,4,5,6	Muncul mata dadu angka 4
Pertandingan sebuah klub sepak bola	Menang, seri, kalah	Menang
Prakiraan cuaca	cerah, cerah berawan, hujan ringan, hujan sedang, hujan petir	Cerah

## 6.3 Pendekatan Perumusan Probabilitas

Untuk memahami probabilitas, maka diperlukan beberapa pendekatan yang sering digunakan, yaitu pendekatan klasik, pendekatan Relatif, dan pendekatan subjektif.

### 6.3.1 Pendekatan Klasik

Pada pendekatan klasik, semua peristiwa pada suatu percobaan mempunyai kesempatan yang sama untuk terjadi. Probabilitas suatu percobaan dinyatakan

sebagai perbandingan antara jumlah kemungkinan hasil kejadian dengan seluruh kemungkinan hasil pada ruang sampel.

Bila kejadian A terjadi dalam m cara dengan asumsi ruang sampel S, maka probabilitas P(A) dapat dirumuskan sebagai berikut:

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

Keterangan:

$P(A)$  = probabilitas kejadian A

$m$  = kemungkinan cara atau kejadian

$S$  = ruang sampel

Contoh 1.

Sebuah kelas akan mengadakan pemilihan siswa berprestasi, di dalam kelas tersebut terdiri dari 10 siswa laki-laki (A) dan 15 siswa perempuan (B). Berapa probabilitas siswa perempuan akan terpilih ?

Jawab:

Jumlah seluruh siswa sebagai ruang sampel = 25 siswa

Jumlah siswa perempuan = 15 siswa

Maka probabilitas kejadian siswa perempuan terpilih:

$$P(B) = \frac{m}{S} = \frac{15}{25} = 0,6 = 60\%$$

Contoh 2.

Dalam suatu percobaan pelemparan sebuah dadu, berapakah probabilitas P(A) yang terjadi untuk munculnya angka dadu ganjil ?

Jawab:

A = munculnya angka dadu ganjil = {1,3,5} = n(A) = 3

S = ruang sampel dadu = {1,2,3,4,5,6} = n(S) = 6

Maka, probabilitas munculnya angka dadu ganjil:

$$P(A) = \frac{m}{S} = \frac{3}{6} = 0,5 = 50\%$$

### 6.3.2 Pendekatan Frekuensi Relatif

Probabilitas dengan pendekatan frekuensi relatif sering juga disebut dengan probabilitas empiris. Dalam pendekatan frekuensi relatif, peristiwa pada suatu percobaan dianggap tidak mempunyai kesempatan yang sama untuk terjadi, dan probabilitasnya tergantung pada frekuensi dari suatu peristiwa yang terjadi. Hal tersebut yang membedakan antara pendekatan frekuensi relatif dengan pendekatan klasik.

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

$P(A)$  : Probabilitas kejadian A

$n(A)$  : Frekuensi kejadian A

$n(S)$  : Ruang sampel

Contoh 3

Sebuah mata dadu dilempar 100 kali, dalam percobaan tersebut diperoleh hasil sebagai berikut:

**Tabel 6.2:** Percobaan Melempar Dadu

Mata Dadu (x)	Frekuensi (f)
1	16
2	14
3	25
4	15
5	20
6	10

Berapakah probabilitas untuk masing-masing mata dadu dari percobaan yang ditunjukan pada tabel di atas ?

Jawab:

Untuk masing-masing mata dadu, maka diperoleh probabilitasnya adalah sebagai berikut,

$$P(x = 1) = \frac{f_1}{n(S)} = \frac{16}{100} = 0.16$$

$$P(x = 2) = \frac{f_2}{n(S)} = \frac{14}{100} = 0.14$$

$$P(x = 3) = \frac{f_3}{n(S)} = \frac{25}{100} = 0.25$$

$$P(x = 4) = \frac{f_4}{n(S)} = \frac{15}{100} = 0.15$$

$$P(x = 5) = \frac{f_5}{n(S)} = \frac{20}{100} = 0.20$$

$$P(x = 6) = \frac{f_6}{n(S)} = \frac{10}{100} = 0.10$$

#### Contoh 4

Dalam suatu perkuliahan Mata Kuliah Statistik diperoleh bahwa dari 40 mahasiswa, terdapat 15 mahasiswa yang mendapatkan nilai 100. Berapakah besarnya probabilitas seorang mahasiswa mendapat nilai 100 ?

Jawaban:

Berdasarkan kejadian pada soal, maka nilai probabilitas seorang mahasiswa jika ingin mendapatkan nilai 100 adalah

$$P(x = 100) = \frac{f}{n(S)} = \frac{15}{40} = 0.375 = 37.5\%$$

Pada statistik, ada suatu hukum yang banyak mendampingi pembahasan tentang probabilitas yaitu “The Law of large Number” atau Hukum bilangan besar. Hukum ini menyatakan bahwa semakin banyak percobaan yang dilakukan, maka probabilitas empiris dari suatu kejadian akan mendekati probabilitas sesungguhnya (Srava Chrisdes Antoro, 2005).

### Contoh 5

Pada kejadian pelemparan sebuah koin mata uang, maka probabilitas klasik munculnya Gambar 6 adalah  $\frac{1}{2}$ , Sedangkan, pada probabilitas empiris ditemukan kejadian sebagai berikut:

**Tabel 6.3:** Percobaan Pelemparan Sebuah Koin Mata uang

Banyaknya percobaan	Banyaknya Kejadian muncul sisi “Gambar 6”	Frekuensi relatif dari sisi “Gambar 6”
1	0	0
10	3	0.3
50	26	0.52
100	52	0.52
500	236	0.472
1000	494	0.494
10.000	5027	0.5072

Dari tabel 6.3 kita dapat melihat bahwa semakin banyak percobaan dilakukan maka probabilitas empiris semakin mendekati nilai dari probabilitas klasik. Sebagai contoh dari data tabel di atas, pada percobaan 10.000 kali diperoleh frekuensi relatif dari sisi Gambar 6 adalah 0.5072 atau 50,72 %.

### 6.3.3 Pendekatan Subjektif

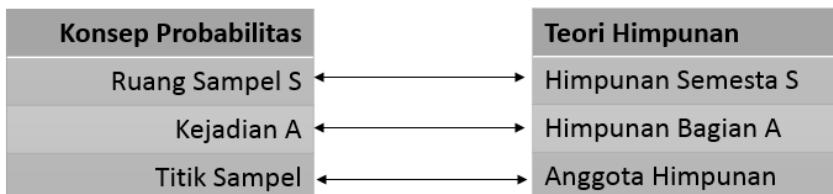
Jika pada dua pendekatan perumusan probabilitas yang sudah kita bahas sebelumnya yaitu pendekatan klasik dan pendekatan frekuensi relatif, di mana pada kedua pendekatan tersebut perumusan probabilitas didasarkan pada sebuah perhitungan matematis yang memberikan nilai keyakinan terhadap suatu probabilitas. Maka pada pendekatan subjektif, untuk menentukan nilai suatu probabilitas kejadian didasarkan pada selera dan keyakinan seseorang berdasarkan pengalaman, analisis pribadi, tingkat pemahaman, pengetahuan dan informasi yang diperoleh.

Sebagai contoh jika besok hari ada pertandingan sepakbola antara Real Madrid dan Barcelona, maka seseorang bisa membuat probabilitas tim mana yang

akan menang dengan melakukan analisis bagaimana hasil dari pertandingan-pertandingan sebelumnya, bagaimana performa dari masing-masing tim yang akan bertanding, bagaimana rekam jejak pelatih yang saat ini menangani, bagaimana kemampuan dari setiap pemain pada setiap tim. Dengan analisis tersebut maka seseorang bisa membuat suatu probabilitas berdasarkan subjektivitas informasi yang diperoleh.

## 6.4 Probabilitas Kejadian Majemuk

Dalam pembahasan konsep probabilitas terdapat di dalamnya keterkaitan dengan teori pada himpunan, sehingga pengetahuan kita tentang teori himpunan akan memberikan dasar yang baik terhadap pemahaman kita pada teori probabilitas ini. Kita uraikan keterkaitannya pada Gambar 6.3 berikut (Sudaryono M.Pd, 2012):



**Gambar 6.3** Hubungan Konsep Probabilitas dan Teori Himpunan

Dalam teori himpunan kita ketahui bahwa

$$n(A \cup B) = n(A) + n(B) - n(A \cap B)$$

Bila kedua ruas kita bagi dengan  $n(S)$ , maka akan menghasilkan

$$\frac{n(A \cup B)}{n(S)} = \frac{n(A)}{n(S)} + \frac{n(B)}{n(S)} - \frac{n(A \cap B)}{n(S)}$$

Maka diperoleh  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$P(A \cap B)$  merupakan probabilitas yang muncul pada kejadian A dan pada kejadian B yang kejadianya terjadi secara bersamaan,  $P(A \cap B)$  disebut juga dengan join probability.

### Contoh 6

Pada percobaan melempar sebuah dadu, berapakah munculnya mata dadu ganjil atau mata dadu bilangan prima ?

Jawab:

$$A = \text{munculnya angka dadu ganjil} = \{1,3,5\} = n(A) = 3$$

$$B = \text{munculnya angka dadu bilangan prima} = \{2,3,5\} = n(B) = 3$$

$$S = \text{ruang sampel dadu} = \{1,2,3,4,5,6\} = n(S) = 6$$

$$\text{Sehingga kita peroleh nilai } P(A) = \frac{n(A)}{n(S)} = \frac{3}{6} = \frac{1}{2} \text{ dan nilai } P(B) = \frac{n(B)}{n(S)} = \frac{3}{6} = \frac{1}{2}.$$

Dari jawaban di atas dapat terlihat bahwa angka 3 dan 5 muncul pada kejadian A dan kejadian B, sehingga  $P(A \cap B) = \frac{2}{6} = \frac{1}{3}$ , maka

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ P(A \cup B) &= \frac{1}{2} + \frac{1}{2} - \frac{1}{3} = \frac{2}{3} \end{aligned}$$

Sehingga kita dapat menyebutkan munculnya mata dadu ganjil atau mata dadu bilangan prima adalah  $\frac{2}{3}$

#### 6.4.1 Kejadian Saling Lepas

Bila kejadian A dan kejadian B dalam suatu ruang sampel S dan berlaku  $A \cap B = \emptyset$  maka kejadian tersebut disebut sebagai kejadian saling lepas atau saling bertentangan atau saling terpisah atau mutually Exclusive (Sudaryono M.Pd, 2012). Dari kejadian  $A \cap B = \emptyset$  di atas dapat diartikan bahwa tidak ada anggota yang sama antara kejadian A dan kejadian B. Sehingga  $P(A \cap B) = P(\emptyset) = 0$  maka probabilitas untuk kejadian saling lepas menjadi:

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

### Contoh 7

Dua buah dadu dilempar, tentukan probabilitas dari muncul jumlah mata dadu 8 dan jumlah mata dadu 10 !

Jawab:

A = munculnya angka dadu jumlah 8

$$= \{(1,7),(2,6),(3,5),(4,4),(5,3),(6,2),(7,1)\} = n(A) = 7$$

B = munculnya angka dadu jumlah 13

$$= \{(3,7),(4,6),(5,5),(6,4),(7,3)\} = n(A) = 5$$

S = ruang sampel kejadian

$$= 36$$

Dari keterangan di atas diperoleh

$$P(A) = \frac{n(A)}{n(S)} = \frac{7}{36}$$

$$P(B) = \frac{n(B)}{n(S)} = \frac{5}{36}$$

Karena  $A \cap B = \emptyset$ , maka artinya kejadiannya saling lepas. Sehingga kita akan memperoleh

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

$$P(A \cup B) = \frac{7}{36} + \frac{5}{36} = \frac{12}{36} = \frac{1}{3} = 0.3333 = 33.33\%$$

#### 6.4.2 Kejadian Saling Bebas

Untuk memahami mengenai kejadian saling bebas, mari kita simak ilustrasi menarik berikut yang diambil dari halaman kemdikbud.go.id tentang dialog antara seorang dokter dengan seorang pasien (Wibowo, 2016):

Dokter: "Bu, penyakit yang ibu derita ini cukup langka dan berbahaya. Satu-satunya jalan harus melalui operasi. Namun, peluang berhasilnya operasi ini cukup kecil, hanya 1 dari 10 operasi yang berhasil. Tapi kabar baiknya, kemarin saya sudah habis melakukan operasi terhadap 9 orang pasien dan semuanya meninggal. Nah ibu adalah pasien ke-10, yang merupakan 1 dari 10 yang akan berhasil itu. Bagaimana, ibu setuju operasi ?

Mari kita cermati satu kutipan dari dialog di atas yang menyatakan bahwa "hanya 1 dari 10 operasi yang berhasil" yang artinya probabilitas keberhasilan dari setiap operasi hanya 1/10. Kejadian di mana 9 pasien yang telah

melakukan operasi dan meninggal dunia tentunya tidak memengaruhi hasil pasien ke 10 yang jika dia melakukan operasi hasilnya akan berhasil, secara matematis kita dapatkan probabilitasnya tidak membesar namun tetap 1/10. Kejadian yang menimpa 9 pasien sebelumnya tidak memengaruhi probabilitas dari pasien selanjutnya. Dialog di atas adalah salah satu ilustrasi dari kejadian saling bebas, di mana kejadian yang satu tidak memengaruhi kejadian yang lain.

Pada dua kejadian saling bebas, berlaku persamaan:

$$P(A \cap B) = P(A) \cdot P(B)$$

Di mana  $P(A \cap B)$  adalah probabilitas kejadian A dan B

Dan sebaliknya jika

$$P(A \cap B) \neq P(A) \cdot P(B)$$

Maka kejadiannya adalah tidak saling bebas.

Contoh 8

Dua buah dadu dilempar secara bersamaan. Berapakah peluang munculnya bilangan prima pada dadu pertama dan munculnya bilangan genap pada dadu kedua.

Jawab:

Kejadian di atas adalah contoh kejadian saling bebas, karena dadu pertama dan dadu kedua tidak saling berkaitan.

A = munculnya bilangan prima pada dadu pertama

$$= \{2,3,5\} = n(A) = 3$$

B = munculnya bilangan genap pada dadu kedua

$$= \{2,4,6\} = n(A) = 3$$

SA = Ruang sampel kejadian A

$$= 6$$

SB = Ruang sampel kejadian B

$$= 6$$

Dari keterangan di atas diperoleh

$$P(A) = \frac{n(A)}{n(S)} = \frac{3}{6} = \frac{1}{2}$$

$$P(B) = \frac{n(B)}{n(S)} = \frac{3}{6} = \frac{1}{2}$$

Maka  $P(A \cap B) = P(A) \cdot P(B)$

$$P(A \cap B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

Contoh 9

Terdapat dua buah kotak. Masing – masing, kotak pertama terdapat 4 bola berwarna hijau dan 3 berwarna kuning, sedangkan pada kotak kedua terdapat 7 bola hijau dan 2 bola berwarna kuning. Berapakah probabilitas terambilnya bola hijau dari kotak pertama dan bola kuning dari kotak kedua !

Jawab:

Kejadian ini adalah kejadian saling bebas, karena masing-masing kotak tidak saling berkaitan.

H1                   = bola hijau di kotak 1,  $n(H_1) = 4$

K2                   = bola kuning di kotak 2,  $n(K_2) = 2$

$S_1$  = Ruang sampel dikotak 1,  $n(S_1) = 7$

$S_2$  = Ruang sampel dikotak 2,  $n(S_2) = 9$

Dari keterangan di atas diperoleh

$$P(H_1) = \frac{n(H_1)}{n(S_1)} = \frac{4}{7}$$

$$P(K_2) = \frac{n(K_2)}{n(S_2)} = \frac{2}{9}$$

Maka  $P(H_1 \cap K_2) = P(H_1) \cdot P(K_2)$

$$P(H_1 \cap K_2) = \frac{4}{7} \cdot \frac{2}{9} = \frac{8}{63}$$

probabilitas terambilnya bola hijau dari kotak pertama dan bola kuning dari kotak kedua adalah  $\frac{8}{63}$

## 6.5 Probabilitas Bersyarat

Probabilitas bersyarat biasa juga disebut dengan conditional probability yaitu probabilitas suatu kejadian A yang terjadi dengan syarat kejadian B terjadi terlebih dahulu, sehingga bisa kita sebutkan kejadian A bersyarat B. Kejadian A bersyarat B dapat kita tuliskan dalam bentuk  $A/B$ , penulisan tanda garis miring “/” tidak menandakan sebagai operasi pembagian. Dengan demikian secara makna penulisan  $A/B$  akan berbeda dengan penulisan  $B/A$ . Penulisan  $A/B$  bermakna kejadian A akan terjadi dengan syarat kejadian B terjadi terlebih dahulu, sedangkan  $B/A$  bermakna kejadian B akan terjadi dengan syarat kejadian A terjadi terlebih dahulu.

Probabilitas bersyarat, dapat dituliskan dalam persamaan berikut:

$$P(A/B) = \frac{P(A \cap B)}{P(B)}, P(B) > 0$$

Atau

$$P(B/A) = \frac{P(B \cap A)}{P(A)}, P(A) > 0$$

Contoh 10

Diberikan data pekerja sesuai dengan jenjang pendidikan dan jenis kelamin pada suatu perusahaan sebagai berikut:

**Tabel 6.4:** Data Pekerja

	S1	D3	Jumlah
Laki-laki	350	250	600
Wanita	150	50	200
Jumlah	500	300	800

Perusahaan tersebut akan menugaskan seorang pegawai untuk membuka cabang baru. Bila yang terpilih adalah pekerja dengan jenjang pendidikan S1, berapakah probabilitasnya bahwa pegawai tersebut adalah Laki-laki ?

Jawab:

Dalam kasus di atas, kejadian terpilihnya laki-laki dengan terlebih dahulu ada

kejadian yang mendahuluinya sebagai syarat yaitu pekerja dengan jenjang S1. Sehingga, jika

A = kejadian terpilihnya pekerja dengan jenjang S1,  $n(A)=500$

B = kejadian terpilihnya pegawai laki-laki

S = ruang sampel semesta,  $n(S)=800$

$$n(A \cap B) = 350$$

Maka dapat kita tentukan

$$P(A \cap B) = \frac{n(A \cap B)}{n(S)} = \frac{n(A \cap B)}{n(S)} = \frac{350}{800}$$

$$P(A) = \frac{n(A)}{n(S)} = \frac{500}{800}$$

$$P(B/A) = \frac{P(B \cap A)}{P(A)} = \frac{\frac{350}{800}}{\frac{500}{800}} = \frac{350}{500} = \frac{7}{10} = 0.7 = 70\%$$

Sehingga dapat kita simpulkan kejadian terpilihnya pegawai laki-laki dengan jenjang S1 adalah sekitar 70%.



## **Bab 7**

# **Data Kelompok dan Analisisnya**

### **7.1 Pendahuluan**

Analisis data sebagai proses yang merinci usaha secara formal untuk menemukan tema dan merumuskan hipotesis (ide) seperti yang disarankan dan sebagai usaha untuk memberikan bantuan dan tema pada hipotesis. Jika dikaji, pada dasarnya definisi pertama lebih menitikberatkan pengorganisasian data sedangkan yang kedua lebih menekankan maksud dan tujuan analisis data. Dengan demikian definisi tersebut dapat disintesiskan bahwa analisis data merupakan proses mengorganisasikan dan mengurutkan data ke dalam pola, kategori dan satuan uraian dasar sehingga dapat ditemukan tema dan dapat dirumuskan hipotesis kerja seperti yang didasarkan oleh data.

Data kelompok, yaitu data yang jumlahnya di atas 30 buah sehingga memerlukan penyajian data dalam tabel distribusi frekuensi di mana data - data tersebut akan dikelompokkan dalam beberapa kelas, dan setiap kelas mempunyai interval nilai tertentu. Data berkelompok adalah cara penyajian data yang menggunakan tabel distribusi frekuensi di mana data tersebut dikelompokkan ke dalam interval tertentu. Sebelum menentukan ukuran pemusatan, letak dan penyebaran suatu data berkelompok, terlebih dahulu akan diuraikan penjelasan dari tabel distribusi frekuensi.

Pada tabel dibawah, data sebanyak 20 buah dibagi dalam 6 kelas, di mana kelas 51 – 55 hingga kelas 76 – 80 berturut-turut disebut kelas pertama, kelas kedua dan

**Tabel 7.1:** Tabel Penilaian

Nilai	f
51 – 55	3
56 – 60	4
61 – 65	5
66 – 70	2
71 – 75	2
76 – 80	4
	20

Pada tiap-tiap kelas terdapat batas bawah dan batas atas, yakni 51, 56, 61, 66, 71 dan 76 dinamakan batas bawah dan 55, 60, 65, 70, 75, dan 80 dinamakan batas atas.

Pada tiap-tiap kelas juga terdapat tepi bawah dan tepi atas, yakni 50,5; 55,5; 60,5; 65,5; 70,5 dan 75,5 berturut-turut dinamakan batas bawah kelas ke 1, 2, 3, 4, 5 dan 6 serta 55,5; 60,5; 65,5; 70,5; 75,5 dan 80,5 berturut-turut dinamakan batas atas kelas ke 1, 2, 3, 4, 5 dan 6.

## 7.2 Jenis-Jenis Analisis Data

Analisis data merupakan salah satu langkah penting dalam rangka memperoleh temuan-temuan hasil penelitian. Hal ini disebabkan, data akan menuntun kita ke arah temuan ilmiah, bila dianalisis dengan teknik-teknik yang tepat. Data yang belum dianalisis masih merupakan data mentah. Dalam kegiatan penelitian, data mentah akan memberi arti, bila dianalisis dan ditafsirkan. Dalam rangka analisis dan interpretasi data, perlu dipahami tentang keberadaan data itu sendiri. Secara garis besar, keberadaan data dapat digolongkan ke dalam dua jenis, yaitu:

### 7.2.1 Data Bermuatan Kualitatif

Data bermuatan kualitatif disebut juga dengan data lunak. Data semacam ini diperoleh melalui penelitian yang menggunakan pendekatan kualitatif, atau penilaian kualitatif. Keberadaan data bermuatan kualitatif adalah catatan lapangan yang berupa catatan atau rekaman kata-kata, kalimat, atau paragraf yang diperoleh dari wawancara menggunakan pertanyaan terbuka, observasi partisipatoris, atau pemakaian peneliti terhadap dokumen atau peninggalan. Untuk memperoleh arti dari data semacam ini melalui interpretasi data, digunakan teknik analisis data kualitatif, seperti yang telah diuraikan pada bab di atas.

### 7.2.2 Data Bermuatan Kuantitatif

Keberadaan data bermuatan kuantitatif adalah angka-angka (kuantitas), baik diperoleh dari jumlah suatu penggabungan ataupun pengukuran. Data bermuatan kuantitatif yang diperoleh dari jumlah suatu penggabungan selalu menggunakan bilangan cacah. Contoh data seperti ini adalah angka-angka hasil sensus, angka-angka hasil tabulasi terhadap jawaban terhadap angket atau wawancara terstruktur. Adapun data bermuatan kuantitatif hasil pengukuran adalah skor-skor yang diperoleh melalui pengukuran, seperti skor tes prestasi belajar, skor skala motivasi, skor timbangan, dan semacamnya.

## 7.3 Teknik Analisis Data Kelompok

### 7.3.1 Distribusi Frekuensi Data Kelompok

Untuk dapat menganalisis data angka yang banyak, terlebih dahulu kita perlu menyusun secara sistematik ke dalam distribusi frekuensi. Distribusi frekuensi adalah mendistribusikan data ke dalam beberapa kelas atau kategori, kemudian menentukan banyaknya individu yang termasuk kelas tertentu yang disebut frekuensi kelas (class frequency).

Data pertama yang diperoleh pada suatu observasi disebut dengan data mentah (rawdata). Data ini belum tersusun secara numerik. Sebagai contoh data mengenai tinggi badan siswa yang penyajiannya masih dalam bentuk presensi kehadiran yang biasanya hanya diurutkan berdasarkan alphabet nama siswa. Terkadang data mentah disajikan berdasarkan urutan naik (ascending) atau

urutan turun (descending). Bentuk penyajian seperti ini disebut array. Selisih antara nilai data terbesar dan terkecil disebut rentang (range).

**CONTOH:** Berikut distribusi frekuensi tinggi badan 100 siswa SMA XYZ

**Tabel 7.2:** Tinggi 100 siswa SMA XYZ

Tinggi Badan (in)	Frekuensi
60 - 62	5
63 - 65	18
66 - 68	42
69 - 71	27
72 - 74	8
Jumlah	100

Jadi dapat disimpulkan bahwa distribusi frekuensi terdiri dari Tinggi Badan (disebut interval kelas) dan frekuensi.

### 7.3.2 Beberapa Istilah Pada Tabel Distribusi Frekuensi

1. Interval Kelas adalah interval yang diberikan dengan menetapkan kelas - kelas dalam distribusi. Pada tabel 1, interval kelasnya adalah 60 - 62, 63 – 65, 66 - 68, 69 - 71, dan 72 - 74. Interval kelas 66-68 secara matematis merupakan interval tertutup [66, 68], ia memuat semua bilangan dari 66 sampai dengan 68. Bilangan 60 dan 62 pada interval 60-62 disebut limit kelas, di mana angka 60 disebut limit kelas bawah dan angka 62 disebut limit kelas atas.
2. Batas Kelas adalah bilangan terkecil dan terbesar sesungguhnya yang masuk dalam kelas interval tertentu. Misalnya

Berdasarkan tabel 1 distribusi frekuensi di atas., maka yang dikatakan batas kelas bawah adalah 60, 63, 66, 69, dan 72. Batas kelas atas adalah 62, 65, 68, 71, dan 7.

### Tepi Kelas (Batas Nyata Kelas)

Untuk mencari tepi kelas maka digunakan rumus berikut ini:

$$\text{Tepi bawah} = \text{Batas kelas bawah} - 0.5$$

$$\text{Tepi atas} = \text{Batas kelas atas} + 0.5$$

Dari tabel 1 di atas maka untuk kelas pertama, tepi kelas bawah 59.5 dan tepi kelas atas 62.5 . Untuk kelas kedua, tepi kelas bawah 62.5 dan tepi kelas atas 65.5

1. Lebar Kelas Interval adalah selisih antara batas atas dan batas bawah batas kelas. Misalnya lebar interval kelas 60-62 adalah  $62.5 - 59.5 = 3$ .
2. Nilai Tengah Kelas adalah titik tengah interval kelas. Diperoleh dengan cara membagi dua jumlah dari batas bawah dan batas atas suatu interval kelas. Contoh nilai tengah kelas untuk kelas interval 66-68 adalah  $(66+68)/2 = 67$ .

#### 7.3.3 Prosedur Umum Membuat Tabel Distribusi Frekuensi

Langkah - langkah dalam membuat tabel distribusi frekuensi adalah:

1. Tetapkan data terbesar dan data terkecil, kemudian tentukan rangenya, range (rentang) = data maks - data min
2. Tentukan banyak kelas interval: Banyak kelas =  $1 + (3.3) * \log \frac{n}{k}$ , dengan n = banyaknya data
3. Tentukan Panjang kelas interval , p = range (rentang) / banyak kelas
4. Starting point, mulailah dengan bilangan batas bawah untuk kelas interval pertama. Dapat dipilih sebagai starting point nya adalah data terkecil dari observasi.
5. Dengan menggunakan limit bawah interval kelas pertama dan lebar interval kelas, tentukan limit bawah interval kelas lainnya
6. Susunlah semua batas bawah interval kelas secara vertikal, kemudian tentukan limit yang bersesuaian
7. Kembalilah data mentah dan bisa menggunakan turus untuk memasukkan data pada interval kelas yang ada

Contoh: Berikut nilai 80 siswa pada ujian akhir mata pelajaran matematika

**Tabel 7.3:** Penilaian siswa

79	80	70	68	90	92	80	70	63	76
49	84	71	72	35	93	91	74	60	63
48	90	92	85	83	76	61	99	83	88
74	70	38	51	73	71	72	95	82	70
81	91	56	65	74	90	97	80	60	66
98	93	81	93	43	72	91	59	67	88
87	82	74	83	86	67	88	71	89	79
80	78	73	86	88	75	81	77	63	75

Langkah - langkah untuk membuat tabel distribusi frekuensi dilakukan sebagai berikut:

1. Perhatikan data, nilai tertinggi dari data di atas adalah 99, dan nilai terendah adalah 35, jadi range (rentang) adalah  $99 - 35 = 64$
2. Tentukan banyak kelas interval: Banyak kelas =  $1 + 3.3 \log 80 = 7.2802$  maka dilakukan pembulatan ke bawah jadi 7
3. Tentukan panjang interval kelas ,  $p = 64 / 7 = 9.14 = 9$  atau 10
4. Pilih  $p = 10$  dengan batas bawah = 35

**Tabel 7.4:** Frekuensi Penilaian

Kelas	Interval Kelas	Frekuensi
1	35 - 44	3
2	45 - 54	3
3	55 - 64	8
4	65 - 74	21
5	75 - 84	21

6	85 - 94	20
7	95 - 104	4
	Jumlah	80

### 7.3.4 Macam - Macam Tabel Distribusi Frekuensi

**Tabel 7.5:** Distribusi Frekuensi Relatif

Nilai Data	Frekuensi Relatif (%)
a - b	g1
c-d	g2
e - f	g3
g - h	g4
i - j	g5
Jumlah	100

dengan frekuensi relatif kelas ke i

$gi = (fi/jumlah) \times 100\%$ , dengan  $fi$  = frekuensi kelas ke -i

Tabel Distribusi Frekuensi Kumulatif

Tabel distribusi kumulatif terdiri dari 2 yaitu:

- Tabel distribusi kumulatif kurang dari
- Tabel distribusi kumulatif lebih dari

**Tabel 7.6:** Frekuensi Kumulatif

Nilai Data	Frekuensi Kumulatif kurang dari	Frekuensi Kumulatif lebih dari
Tepi kelas a	0 f1	f1+f2+f3+f4+f5 f2+f3+f4+f5
Tepi kelas c	f1+f2	f3+f4+f5
Tepi kelas e	f1+f2+f3	f4+f5
Tepi kelas g	f1+f2+f3+f4	f5
Tepi kelas i	f1+f2+f3+f4+f5	0
Tepi kelas k		

Contoh:

Menggunakan tabel diatas, maka distribusi frekuensi kumulatif kurang dari adalah

**Tabel 7.7:** Perhitungan Frekuensi Kumulatif

Tepi Kelas	Frekuensi Kumulatif kurang dari	Distribusi kumulatif lebih dari
59.5	0	100
62.5	5	95
65.5	23	77
68.5	65	35
71.5	92	8
74.5	100	0

### 7.3.5 Pengukuran Nilai Sentral Data Kelompok

Pengukuran nilai sentral menunjukkan pusat suatu kumpulan pengamatan yang merupakan suatu nilai yang mewakili semua nilai pengamatan. Pengukuran nilai sentral yang akan dibahas disini adalah mean, modus, median dan kuartil.

### 7.3.6 Mean Data Kelompok

Menghitung nilai rata-rata suatu data tersebar rumusnya adalah:

$$\text{Rumus } \bar{X} : \frac{\sum X_i}{n}$$

sedangkan untuk data berkelompok:

$$\text{Rumus } \bar{X} : \frac{\sum f_i \cdot x_i}{\sum f_i}$$

Keterangan:

- $\bar{X}$  : rata-rata hitung
- $X_i$  : data ke-i
- $x_i$  : titik tengah masing-masing kelas
- $f_i$  : frekuensi masing-masing kelas
- $n$  : jumlah data

### 7.3.7 Modus Data Kelompok

Modus adalah nilai dari beberapa data yang mempunyai frekuensi tertinggi baik data tunggal maupun data yang berbentuk distribusi atau nilai yang sering muncul dalam kelompok data. Modus merupakan suatu pengamatan dalam distribusi frekuensi yang memiliki jumlah pengamatan di mana jumlah frekuensinya paling besar/paling banyak. jika nilai yang muncul itu hanya ada satu macam saja, maka modus tersebut dinamakan unimodel. Dan jika nilai yang muncul ada dua macam, maka modus tersebut dinamakan bimodal. Jadi

dapat disimpulkan bahwa modus adalah nilai dari beberapa data yang memiliki frekuensi tertinggi baik terbanyak dalam pengamatan.

Modus Data Kelompok adalah nilai yang memiliki frekuensi terbanyak dalam seperangkat data. Modus untuk data yang disusun dalam bentuk kelas interval (data berkelompok) bisa ditentukan berdasarkan nilai tengah kelas interval yang memiliki frekuensi terbanyak.

Namun nilai yang dihasilkan dari nilai tengah kelas interval ini adalah nilai yang kasar. Nilai modus yang lebih halus bisa diperoleh dengan menggunakan rumus di bawah ini.

$$Mo = b + \left( \frac{b_1}{b_1 + b_2} \right) p$$

Mo                   = modus

B                   = batas bawah kelas interval dengan frekuensi terbanyak

p                   = panjang kelas interval

b1                   = frekuensi terbanyak dikurangi frekuensi kelas sebelumnya

b2                   = frekuensi terbanyak dikurangi frekuensi kelas sesudahnya

Sifat-sifat modus:

1. Modus tidak dipengaruhi oleh data yang ekstrim besar atau kecil
2. Bisa digunakan untuk menentukan harga tengah pada distribusi frekuensi
3. Modus tidak unik, karena kadangkala ada sekelompok data yang mempunyai modus lebih dari satu (misalnya modusnya 2 bimodal) atau tidak punya modus. Dalam hal ini modus tidak bisa digunakan sebagai harga tengah.

### 7.3.8 Median Data Kelompok

Dalam pengamatan, biasanya terdapat satu atau dua nilai yang sangat besar atau sangat kecil yang menyebabkan rata-rata hitung tidak mencerminkan keadaan yang sebenarnya. Hal ini diatasi dengan menggunakan harga tengah yang disebut Median. Median adalah nilai atau titik tengah dari seluruh data setelah diurutkan menurut besarnya sehingga untuk data tersebar dapat dicari

dengan mengurutkan data terlebih dahulu kemudian mencari letak nilai tengah sedangkan untuk data berkelompok median dicari dengan:

$$\text{Rumus Me: } b + P \frac{0,5n - F}{f}$$

Keterangan:

- |    |   |
|----|---|
| Me | : Median                                      |
| b  | : batas bawah kelas median                    |
| P  | : panjang interval                            |
| n  | : banyaknya data                              |
| F  | : jumlah semua frekuensi sebelum kelas median |
| f  | : frekuensi kelas median                      |

Sifat-sifat Median:

1. Bersifat unik, hanya ada satu median untuk satu kelompok data
2. Untuk menentukan median, data diurutkan terlebih dahulu dengan urut ascending
3. Median tidak dipengaruhi nilai ekstrim besar atau kecil. Oleh karena itu median merupakan harga tengah yang baik, jika ditemui adanya nilai ekstrim data.

### 7.3.9 Kuartil Data Kelompok

Secara umum kuartil merupakan sekumpulan data yang dibagi menjadi empat bagian yang sama banyak, sesudah disusun menurut urutan nilainya, maka bilangan pembaginya disebut kuartil. Pengertian kuartil di atas menyebutkan, nilai kuartil merupakan nilai dari sekumpulan data yang dibagi menjadi empat bagian yang sama, dan yang membagi data tersebut dinamakan kuartil. Selain itu juga terdapat pengertian lainnya yang menyebutkan kuartil merupakan nilai atau angka yang membagi data terkecil sampai data terbesar atau sebaliknya dari data terbesar sampai data terkecil

Kuartil adalah nilai data yang membagi sejumlah data menjadi 4 bagian yang sama setelah data diurutkan nilainya.

Untuk menentukan kuartil data caranya adalah:

1. Susun data menurut urutan nilainya
2. Tentukan letak kuartil
3. Tentukan nilai kuartil

Untuk data tersebar rumusnya adalah:

$$\text{Ki: data ke } \frac{i(n+1)}{4}$$

Sedangkan untuk data berkelompok:

$$\text{Rumus Ki: } b + P \frac{\frac{in}{4} - F}{f}$$

Keterangan:

- Ki : kuartil data ke-i  
b : batas bawah kelas ke-i  
P : panjang interval  
n : banyaknya data  
F : jumlah semua frekuensi sebelum kelas kuartil

# **Bab 8**

## **Ukuran Pemusatan**

### **8.1 Pendahuluan**

Salah satu aspek terpenting dalam mendeskripsikan distribusi data adalah nilai pusat data. Setiap pengukuran aritmatika yang dimaksudkan untuk mendeskripsikan nilai yang merepresentasikan nilai pusat, atau nilai pusat dari kumpulan data (set observasi), dikenal sebagai ukuran pemusatan (tendensi sentral).

Ada tiga ukuran pemusatan data yang sering digunakan, yaitu:

1. Mean (Rata-rata hitung/rata-rata aritmetik)
2. Median
3. Modus atau Mode

### **8.2 Mean (Rata-Rata)**

Mean atau istilah lainnya nilai rata-rata adalah jumlah keseluruhan data dibagi banyaknya data (datum). Nilai rata-rata dibagi lagi menjadi empat, yaitu sebagai berikut. Mean merupakan metode yang paling banyak digunakan untuk menggambarkan ukuran tendensi sentral. Mean dihitung dengan

menjumlahkan semua nilai data pengamatan kemudian dibagi dengan banyaknya data.

### 8.2.1 Mean dari data tunggal

Mean adalah rasio dari total nilai yang diamati dengan jumlah total observasi. Jika data dari variabel random X sebanyak n buah, yang dapat dilambangkan dengan  $X_1, X_2, X_3 \dots X_n$  maka rata-rata dari data tersebut dapat dituliskan seperti dibawah ini:

$$\underline{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

Keterangan:

$\underline{x}$  = rata-rata /mean

$x_i$  = data ke-I, dengan  $i=1,2,3,\dots,n$ .

Satuan yang digunakan sama dengan satuan asli atau satuan data. Jika datanya menggunakan kilogram, maka rata-rata juga menggunakan kilogram

Contoh soal:

Diketahui nilai ujian akhir mahasiswa mata kuliah Pengenalan Teknologi Informasi sebagai berikut: 80, 88, 52, 60, 77, 95, 55, 72, 93, dan 68, maka rata-rata dari data nilai tersebut adalah:

$$\underline{x} = \frac{80 + 88 + 52 + 60 + 77 + 95 + 55 + 72 + 93 + 68}{10} = \frac{740}{10} = 74$$

#### 8.2.1 Mean dari data distribusi Frekuensi

Variasi lain adalah jika setiap data yang dihitung memiliki frekuensi kemunculan tertentu, sehingga rumus mean sederhana bisa diubah menjadi:

$$\underline{x} = \frac{f_1 + f_2 x_2 + f_3 x_3 + \dots + f_n x_n}{f_1 + f_2 + f_3 + \dots + f_n} = \frac{\sum_{i=1}^n f_i x_i}{\sum_{i=1}^n f_i}$$

Di mana:

$\underline{x}$  = rata-rata/mean

$x_i$  = data ke-I, dengan  $i=1,2,3,\dots,n$

$f_i$  = frekuensi ke-I dengan  $i=1,2,3,\dots,n$

Contoh soal:

**Tabel 8.1:** Frekuensi dan nilai akhir mahasiswa

$x_1$	$f_i$	$f_i x_1$
80	5	400
88	7	616
52	4	208
60	7	420
77	4	308
95	1	95
55	3	165
72	9	648
Jumlah	40	2860

Tabel diatas menunjukkan data nilai ujian akhir mata kuliah Pengenalan Teknologi Informasi dalam satu kelas dengan jumlah 40 orang. Maka rata-rata nilai ujian akhir yang dimiliki dalam kelas adalah:

$$\underline{x} = \frac{400 + 616 + 208 + 420 + 308 + 95 + 165 + 648 + 465 + 340}{5 + 7 + 4 + 7 + 4 + 1 + 3 + 9}$$

$$\underline{x} = \frac{2860}{40} = 71,5$$

### 8.2.2 Mean Aritmatika dari Data Berdaftar Distribusi Frekuensi

Perhitungan rata-rata ini digunakan untuk data dalam jumlah besar tetapi pengulangannya sedikit. Adapun langkah-langkah membuat tabel frekuensi yang berinterval adalah sebagai berikut (Viandari, 2020).

1. Menentukan data terkecil dan terbesarnya.
2. Menentukan jangkauan datanya ( $J$ ). Jangkauan data merupakan hasil pengurangan data terbesar oleh data terkecil ( $J = \text{data terbesar} - \text{data terkecil}$ ).
3. Menentukan banyaknya kelas dengan aturan berikut.  $k = 1 + 3,322 \log n$ , di mana  $n = \text{ukuran sampel}$
4. Menentukan interval kelas

$$l = \frac{J}{k}$$

5. Membuat tabel distribusi frekuensi dengan metode turusmenghitung rata-rata untuk data berinterval dengan cara:
6. Menentukan nilai tengah dari masing-masing kelas, yaitu dengan membagi batas atas + batas bawah dengan 2.
7. Menggunakan rumus rata-rata seperti sebelumnya.

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

Dengan  $x_i$  = nilai tengah kelas.

Contoh Soal:

Banyaknya transaksi pada sebuah toko suatu selama 60 hari ditunjukkan oleh data berikut:

**Tabel 8.2:** Data jumlah transaksi per hari selama 60 hari

75	60	81	82	96	81	103	91	100	102
86	92	85	102	95	926	104	99	90	96
87	64	84	98	90	87	110	93	84	89
84	82	74	96	110	65	87	88	91	98
94	91	89	83	112	88	66	107	97	103
96	112	92	94	84	96	83	101	86	115

Tentukan rata-rata transaksi pada toko tersebut!

Untuk menentukan rata-rata transaksi selama 60 hari, bisa menggunakan cara biasa, namun itu tidak direkomendasikan. Cara paling mudah untuk menentukan rata-ratanya adalah dengan menggunakan tabel distribusi frekuensi. dengan Langkah sebagai berikut.

**Langkah 1:** Tentukan nilai data terkecil dan terbesarnya

$$\text{Data terkecil} = 60$$

$$\text{Data terbesar} = 115$$

**Langkah 2:** Tentukan jangkauannya

$$J = \text{data terbesar} - \text{data terkecil} = 115 - 60 = 55$$

**Langkah 3:** Tentukan banyak kelasnya

$$k = 1 + 3,322 \log n$$

$$= 1 + 3,322 \log 60$$

$$= 6,9$$

Banyaknya kelas dibulatkan menjadi  $k = 7$  kelas.

**Langkah 4:** Tentukan panjang kelas (interval)

$$l = \frac{J}{k} = \frac{55}{7} = 7,9$$

Panjang kelas dibulatkan menjadi 8.

**Langkah 5:** Membuat tabel distribusi frekuensi.

**Tabel 8.3:** Tabulasi distribusi frekuensi

Jumlah Transksi	Turus	Frekuensi
60 - 67	d	4
68 - 75	b	2
76 - 83	ea	6
84 - 91	eed	19
92 - 99	eeeaa	16

100 - 107	ec	8
108 - 155	e	5

**Langkah 6:** tentukan nilai tengah setiap kelas.

**Tabel 8.4:** Distribusi frekuensi dan nilai tengah

Jumlah transaksi	Nilai Tengah	Frekuensi
60 – 67	$\frac{60 + 67}{2} = 63,5$	4
68 – 75	$\frac{68 + 75}{2} = 71,5$	2
76 – 83	$\frac{76 + 83}{2} = 79,5$	6
84 – 91	$\frac{84 + 91}{2} = 87,5$	19
92 – 99	$\frac{92 + 99}{2} = 95,5$	16
100 – 107	$\frac{100 + 107}{2} = 103,5$	8
108 – 115	$\frac{108 + 115}{2} = 111,5$	5

Dengan demikian, rata-rata diperoleh sebagai berikut.

**Tabel 8.5:** Rata rata transaksi pada masing masing kelas yang diperoleh

$x_1$	$f_1$	$f_1x_1$
63,5	4	254
71,5	2	143
79,5	6	477

87,5	19	1662,5
95,5	16	1528
103,5	8	828
111,5	5	557,5
	$\sum f_1 = 60$	$\sum f_1 - x_1 = 5450$

Rata-rata:

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{5450}{60} = 90,83$$

Jadi, rata-rata transaksi per hari selama 60 hari adalah 90,83.

### 8.2.1 Mean data gabungan

Rata-rata gabungan (disebut juga grand mean, pooled mean, atau rata-rata umum) adalah cara yang tepat untuk menggabungkan rata-rata hitung.(Setiawan, 2021) Secara matematis, rata-rata data gabungan dinyatakan sebagai berikut.

$$\begin{aligned}\bar{x} &= \frac{\sum x_1 + \sum x_2}{n} \\ \bar{x} &= \frac{n_1 \underline{x}_1 + n_2 \underline{x}_2}{n_1 + n_2}\end{aligned}$$

#### Contoh Soal 3

Nilai rata-rata PTI laki-laki adalah 68 dan nilai rata-rata PTI adalah 75. Jika rata-rata nilai gabungannya adalah 70, tentukan perbandingan banyaknya laki-laki dan perempuan!

Pembahasan:

Diketahui:

$$x_1 = 68$$

$$x_2 = 75$$

$$\underline{x} = 70$$

Ditanya:  $n_1 : n_2$

Penyelesaian:

Secara matematis, rata-rata nilai gabungan dirumuskan sebagai berikut.

$$\underline{x} = \frac{n_1 x_1 + n_2 x_2}{n_1 + n_2}$$

$$70 = \frac{n_l 68 + n_p 75}{n_1 + n_2}$$

$$70n_1 + 70n_1 = n_l 68 + n_p 75$$

$$2n_1 = 5n_2$$

$$\frac{n_1}{n_2} = \frac{5}{2}$$

Jadi, perbandingan jumlah laki-laki dan perempuan adalah 5: 2.

## 8.3 Median

Median adalah ukuran pemusatan di mana suatu data terbagi menjadi dua sama banyak. Median menentukan letak data setelah data itu disusun menurut urutan nilainya . Median dari sekumpulan data adalah data tengah setelah seluruh data disusun nilainya dari yang terkecil sampai yang terbesar. Median dinotasikan dengan  $Me$  (Hidayati, Handayani and Ines Heidiani, 2019).

### 8.3.1 Median dari Data Tunggal

Median data tunggal ditentukan dengan cara sebagai berikut:

$$M_e = \left\{ \begin{array}{l} \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2} \text{ Jika } n \text{ genap} \\ x \left( \frac{n+1}{2} \right) \text{ Jika } n \text{ ganjil} \end{array} \right.$$

Di mana  $x_1$ , adalah data terkecil dan  $x_n$ , adalah data terbesar, sedangkan  $x_k$  adalah data terkecil ke-k dari data tersusun, untuk  $k= 1, 2, 3, \dots, n$ .

Contoh Soal:

- Median data tunggal dengan banyak datanya ganjil

Misal, dari data: 3, 2, 3, 1, 4, 6, 5, 7, 5 Mediannya adalah ....

Susun data terlebih dahulu menjadi: 1, 2, 3, 3, 4, 5, 5, 6, 7. Maka  $Me = 4$ .

- Median dari data tunggal dengan banyak datanya genap

Misal dari data: 2, 4, 6, 1, 4, 3, 5, 7. Mediannya adalah ...

$$1, 2, 3, 4, 4, 5, 6, 7 . \text{ Maka } Me = \frac{4+4}{2} = 4$$

### 8.3.2 Median dari Data Berdistribusi Frekuensi

$$Me = L_1 + d \left( \frac{\frac{n}{2} \sum F}{f_{median}} \right)$$

Keterangan:

$Me$  = Median

$L_1$  = Batas bawah kelas Median

$d$  = lebar interval kelas

$n$  = banyak data

$\sum F$  = jumlah frekuensi sebelum interval kelas Median

$f_{median}$  = frekuensi kelas Median

Contoh soal:

**Tabel 8.6:** Distribusi nilai 40 mahasiswa dalam suatu mata kuliah:

Rentang Nilai	fi
50-57	6
58-65	7
66-73	9
74-81	6

82-89	7
90-97	5
Jumlah	40

$$\frac{n}{2} = 20$$

Kira-kira berada di rentang nilai 66-73

$$L_1 = 65,5$$

$$d = 8$$

$$\sum F = 6 + 7 = 13$$

$$f_{median} = 9$$

$$M_e = L_1 + d \left( \frac{\frac{n}{2} - \sum F}{f_{median}} \right) = 65,5 + 8 \left( \frac{20 - 13}{9} \right)$$

$$= 65,5 + 6,22 = 71,72$$

## 8.4 Modus

Modus adalah data yang paling sering muncul atau memiliki frekuensi tertinggi dari pengamatan yang diperoleh. Apabila ada satu modus atau satu data yang memiliki frekuensi paling banyak keluar dari data pengamatan, maka disebut sebagai unimodus. Sedangkan bila ada dua data yang memiliki frekuensi paling banyak disebut dengan bimodus, dan seterusnya. Notasi modus yang akan kita gunakan dalam modul ini adalah Mo (Hidayati, Handayani and Ines Heidiani, 2019)

### 8.4.1 Modus dari Data Tunggal

Mencari modus data tunggal dapat dilakukan dengan mudah dan cepat; yaitu hanya dengan memeriksa (mencari) di antara skor yang ada, yang memiliki frekuensi paling banyak. Skor atau nilai yang memiliki frekuensi paling banyak itulah yang disebut sebagai modus (Pratikno, 2020).

Sekumpulan data: 2, 3, 4, 4, 5 Maka modusnya adalah 4 muncul 2 kali.

Sekumpulan data: 3, 3, 3, 4, 4, 5, 5, 5, 6, 9 Maka modusnya adalah 3 dan 5 masing-masing muncul 3 kali.

Sekumpulan data: 1, 1, 2, 3, 4, 4, 5, 6, 7, 7, 8 Maka modusnya adalah 1, 4, dan 7 masing-masing muncul 2 kali d. Sekumpulan data: 3, 4, 5, 6, 7 Maka modusnya tidak ada.

#### 8.4.2 Modus dari Data Berdistribusi Frekuensi

$$M_0 = L_1 + d \left( \frac{\Delta_1}{\Delta_1 + \Delta_2} \right)$$

Keterangan:

$M_0$  = Modus

$L_1$  = Batas bawah kelas Modus

$d$  = lebar interval kelas

$\Delta_1$  = frekuensi kelas modul dikurangi frekuensi kelas sebelumnya

$\Delta_2$  = frekuensi kelas modul dikurangi frekuensi kelas sesudahnya

Contoh soal:

Berapakah Modus dari tabel 8.6.

Lihat Kelas modus 66-73

$L_1 = 65,5$

$d = 8$

$\Delta_1 = 9-7 = 2$

$\Delta_2 = 9-6 = 3$

$$M_o = 65,5 + 8 \left( \frac{2}{2 + 3} \right) = 68,7$$

Jadi, Modus dari data di samping adalah 68,7

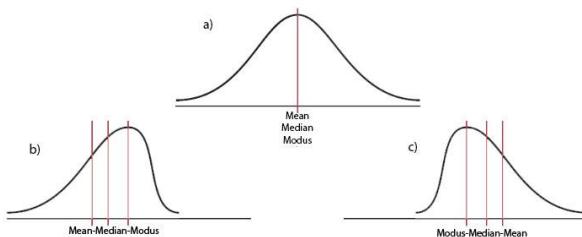
## 8.5 Hubungan antara Mean, Modus, dan Median

Mean, median, dan modus adalah nilai yang digunakan untuk merepresentasikan sekumpulan data. Ketiga nilai ini sering disebut sebagai ukuran kecenderungan terpusat (measure of central tendency). Karena kecenderungan nilai-nilai ini memusat di tengah kumpulan data.

Dalam analisis data, biasanya fokus perhatian bukan pada keseluruhan data, melainkan hanya pada tempat data tersebut dipusatkan. Oleh karena itu, nilai mean, median dan modus sering digunakan untuk merepresentasikan sekumpulan data dalam analisis statistik (Sujalu et al., 2021).

hubungan antara mean, median dan modus sebagai berikut:

$$\text{Mean} - \text{Modus} = 3(\text{Mean} - \text{Median})$$



**Gambar 8.1:** Pola distribusi data (variabel acak kontinu) dengan posisi Mean, Median dan Modus

1. Pola distribusi data Simetris: menunjukkan letak nilai rata-rata hitung, median, dan modus berhimpit (berkisar di satu titik).
2. Pola distribusi data Miring ke kiri: apabila mempunyai nilai modus paling besar dan rata-rata hitung paling kecil.
3. Pola distribusi data Miring ke kanan: apabila mempunyai nilai modus paling kecil dan rata-rata hitung paling besar.

## 8.6 Implementasi Ukuran Pemusatan Menggunakan Python

Python memiliki cukup pustaka (library) dan toolbox dalam membantu pengguna dalam pengolahan data. Salah satu keunggulan Python adalah dukungan dari perusahaan raksasa Google dalam menyediakan Jupyter Notebook online-nya untuk pemrograman, diberikan secara gratis untuk penggunaan standar (Herlawati and Handayanto, 2019). Berikut contoh implementasi bagaimana mencari mean, media dan modus pada python (Zaka, 2019).

### 8.6.1 Mean

Berikut contoh implementasi bagaimana mencari mean pada python tanpa menggunakan library

```
data = [1, 2, 3, 4, 5]
mean = sum(data) / len(data)
print("mean dari data tersebut adalah " + str(mean))
```

mean dari data tersebut adalah 3.0

Berikut contoh implementasi bagaimana mencari mean pada python menggunakan library statistics

```
import statistics as stat
data=[1, 2, 3, 4, 5]
print("mean data tersebut adalah ", stat.mean(data))
mean data tersebut adalah 3
```

### 8.6.2 Median

Berikut contoh implementasi bagaimana mencari median pada python tanpa menggunakan library

```
data = [1, 2, 3, 4, 5]
length = len(data)
```

```
data.sort()
if length % 2 == 0:
    median1 = data[int(length / 2)]
    median2 = data[int(length / 2) - 1]
    median = (median1 + median2) / 2
else:
    median = data[int(length / 2)]
print("median data tersebut adalah " + str(median))
median data tersebut adalah 3
```

Berikut contoh implementasi bagaimana mencari median pada python menggunakan library statistics

```
import statistics as stat
data=[1, 2, 3, 4, 5]
print("median data adalah", stat.median(data))
median data adalah 3
```

### 8.6.3 Modus

Berikut contoh implementasi bagaimana mencari modus pada python tanpa menggunakan library

```
from collections import Counter
n_num = [1, 2, 3, 4, 5, 5]
n = len(n_num)
data = Counter(n_num)
get_mode = dict(data)
mode = [k for k, v in get_mode.items() if v == max(list(data.values()))]
if len(mode) == n:
    get_mode = "modus tidak ditemukan"
```

```
else:  
    get_mode = "modus is " + ','.join(map(str, mode))  
    print(get_mode)  
modus is 5
```

Berikut contoh implementasi bagaimana mencari modus pada python menggunakan library statistics

```
import statistics as stat  
  
data=[1, 2, 3, 4, 5, 5]  
print("modus adalah ", stat.mode(data))  
modus adalah 5
```



## **Bab 9**

# **Ukuran Penyebaran**

### **9.1 Pendahuluan**

Dunia kita dan hidup kita dipenuhi dengan penghitungan dan pengukuran. Statistik membantu kita untuk menghitung dan mengukur, merepresentasikan informasi numerik, terutama informasi tentang kelompok secara efektif, mengeksplorasi gagasan tentang hubungan di antara faktor-faktor yang memengaruhi data kuantitatif, meringkas kumpulan data besar, dan menguji ide-ide. Statistik sering digambarkan sebagai seni mengambil keputusan dibawah ketidakpastian (Kabay, 2019). Statistik adalah disiplin ilmu yang memiliki metode canggih untuk inferensi statistik, prediksi, kuantifikasi ketidakpastian, dan desain eksperimental (Singh, Tiwari and Prasad, 2015).

Statistik merupakan topik atau mata pelajaran esensial dari tingkat sekolah dasar hingga universitas. Namun, seringkali banyak siswa merasa bahwa statistik itu rumit dan rumit (Garfield et al., 2008), karena sebagian besar konsep dalam statistik bersifat abstrak (DelMas, 2004). Karena sifatnya yang abstrak, banyak mahasiswa yang mengalami kesulitan dalam memahami konsep dan hal ini sering menimbulkan kesalahpahaman dalam penalaran statistik, terutama dalam statistik deskriptif dan statistik inferensial. Keadaan seperti itu tentunya akan menghambat pembelajaran statistika mahasiswa dan mengurangi semangat mereka jika masalah ini tidak diatasi. Untuk statistik deskriptif, mahasiswa cenderung membuat kesalahan dalam pengukuran

tendensi sentral, yang terdiri dari mean (Pollatsek, Lima and Well, 2009), mode (Huck, 2009), dan median (Cooper and Shore, 2008). Selain itu, mahasiswa juga mengalami kesulitan dalam mengukur variabilitas (Garfield et al., 2008) yang termasuk standar deviasi (Delmas and Liu, 2005). Secara umum, ada tiga miskonsepsi umum yang sering dibuat oleh mahasiswa tentang probabilitas, salah satu topik statistik inferensial. Dalam hal ini termasuk orientasi hasil (Meletiou, 2000), bias ekiprobabilitas (Fischbein and Schnarch, 1997), dan representatif (Hirsch and O'Donnell, 2001). Selain itu, mahasiswa juga rentan terjadi kesalahan dalam pendistribusian sampel (Castro Sotos et al., 2007), korelasi (Liu, 2010), interval kepercayaan (Castro Sotos et al., 2007), pengujian hipotesis (Castro Sotos et al., 2009), regresi (Huck, 2009). Dengan demikian, untuk memperbaiki kesalahpahaman yang terwujud di kalangan mahasiswa, disarankan agar teknologi informasi digunakan di kelas.

Dalam ilmu statistik, kebutuhan dan peran akan teknologi informasi telah menjadi hal yang sangat vital dan telah dimanfaatkan secara luas di seluruh dunia (Garfield et al., 2008). Integrasi teknologi informasi secara bertahap dapat mengubah konten dan pedagogi dalam pembelajaran statistika (Moore, 1997). Banyak keuntungan yang didapatkan apabila menggunakan teknologi informasi dalam statistika. Salah satu manfaatnya adalah dapat mengurangi waktu dan beban mahasiswa untuk menangani perhitungan yang membosankan dan tidak praktis ketika berhadapan dengan data yang melimpah. Hal ini memungkinkan mahasiswa memiliki waktu yang cukup untuk mengeksplorasi, menganalisis, dan menginterpretasikan data. Manfaat lainnya adalah teknologi informasi dapat membantu mahasiswa dalam memahami ide-ide abstrak statistika. Mahasiswa dapat menampilkan dan memvisualisasikan data dalam berbagai bentuk representasi seperti histogram dan boxplot dengan menggunakan komputer (Garfield et al., 2008).

Disamping itu statistik berhubungan erat dengan informasi. Oleh karena itu, keberhasilan statistik bergantung pada cara kami mengumpulkan, memproses, menjaga, dan menyebarkan informasi. Setelah memasuki abad 21, pengelolaan informasi statistik tidak terbayangkan tanpa dukungan perangkat teknologi informasi modern yang memadai dan dalam dua dekade terakhir terjadi pergerakan penting menuju sistem informasi statistik yang mengintegrasikan tugas-tugas di seluruh aktivitas statistik (Riecan, 2020). Statistik adalah cara yang sangat efisien untuk memperoleh dan menggunakan informasi.

Dalam bab ini akan membahas konsep ukuran penyebaran (dispersi). Dispersi sebenarnya mengacu pada variasi yang ada di dalam dan di antara skor yang

diperoleh suatu kelompok. Kita telah melihat bagaimana rata-rata terjadi konvergensi skor menuju titik tengah dalam distribusi normal. Dalam penyebaran kita akan mencoba dan melihat bahwa bagaimana setiap skor dalam kelompok berbeda dari nilai rata-rata. Semakin besar dispersi, semakin sedikit homogenitas kelompok yang bersangkutan dan jika dispersi lebih kecil berarti kelompok tersebut homogen. Dispersi dalam statistik penting untuk membantu dan mengetahui seberapa jauh populasi sampel bervariasi dari populasi, memberitahu kita tentang kesalahan standar mean dan juga memberi kita indikasi mean dan deviasi standar (Tendencies, 2017).

## 9.2 Konsep Ukuran Penyebaran

Data statistika yang telah dikumpulkan, disajikan, dianalisa secara seksama dan data-data tersebut terkonsentrasi pada daerah sebaran tertentu yang sama. Menyebar berarti ada penyimpangan atau menyimpang dari sebuah acuan sebuah patokan atau yang dijadikan standar, sehingga dalam hal pengukuran tendensi sentral tidak mengungkapkan variabilitas yang ada dalam data (Christopher, 2016). Ukuran tendensi sentral tidak cukup untuk mendeskripsikan data. Dua kumpulan data dapat memiliki mean yang sama tetapi keduanya dapat sepenuhnya berbeda. Jadi untuk mendeskripsikan data, perlu diketahui sejauh mana variabilitasnya. Ini diberikan oleh ukuran dispersi. Rentang, rentang interkuartil, dan deviasi standar adalah tiga ukuran dispersi yang umum digunakan (Manikandan, 2011). Dispersi adalah sebaran rangkaian data secara rata-rata (Shruti, 2021) dan sejauh mana nilai dalam suatu distribusi berbeda dari rata-rata distribusi (Berman and Wang, 2020).

Dispersi didefinisikan sebagai deviasi atau hamburan nilai dari nilai sentralnya yaitu rata-rata (Mean, Median atau Mode tetapi lebih disukai Mean atau Median) (Shah, 2019). Dispersi menemukan variabilitas dalam keseragaman. Dengan kata lain, dispersi mengukur derajat atau sejauh mana nilai variabel menyimpang dari rata-ratanya. Dispersi menunjukkan tingkat heterogenitas antar variabel dan ketika heterogenitas meningkat, dispersi meningkat. Jika semua nilai sama maka setiap ukuran dispersi selalu nol. Semua ukuran dispersi adalah positif. Semua ukuran penyebaran tidak bergantung pada perubahan asal tetapi bergantung pada perubahan skala. Semua prasyarat ukuran yang baik dari tendensi sentral sama-sama berlaku untuk ukuran dispersi yang baik. Dua distribusi mungkin memiliki i) Tendensi sentral yang

sama dan dispersi yang sama; ii) Tendensi sentral berbeda tetapi dispersi sama; iii) Tendensi sentral yang sama tetapi dispersi berbeda; iv) Tendensi sentral berbeda dan dispersi berbeda (Shah, 2019).

Penyebaran adalah suatu nilai yang menunjukkan kearah mana nilai-nilai data statistika itu menyimpang dari suatu acuan yang dijadikan patokan dan masih dapat diterima. Konsentrasi penyimpangan data terhadap suatu acuan patokan nilai dibagi menjadi empat sama banyak disebut kuartil, menjadi sepuluh sama banyak disebut desil, menjadi seratus sama banyak disebut persentil, dan sebagainya. Bentuk penyimpangan terhadap suatu acuan patokan nilai yang digunakan adalah nilai rata-rata disebut simpang rata-rata dan jika berdasarkan nilai baku disebut simpang baku (Latief, 2009). Besarnya nilai simpangan yang diperoleh selanjutnya dapat digunakan untuk menetapkan sebuah patokan lain yang dikenal sebagai angka baku, yaitu dalam bentuk z skore. Angka baku diperoleh dari angka mentah yang diolah dengan cara mengubahnya menggunakan simpangan baku sebagai angka patokannya (Latief, 2009).

Dispersi dalam statistik berarti penyimpangan skor dalam suatu kelompok atau seri, dari skor rata-rata. Mengacu pada penyebaran skor dalam kelompok dalam kaitannya dengan mean. Variasi dapat diukur dan disebut sebagai ukuran dispersi. Ukuran ini mengacu pada banyak metode pengukuran dispersi atau variasi untuk mengukur penyebaran nilai yang berbeda dan terpencar-pencar dalam suatu kelompok (Tendencies, 2017).

### 9.3 Tipe Ukuran Penyebaran

Ada dua jenis ukuran dispersi:

**Tabel 9.1:** Tipe Ukuran Penyebaran (Shah, 2019)

Ukuran Mutlak	Ukuran Relatif
Ukuran dispersi ini akan memiliki unit yang sama dengan variabel	Biasanya dinyatakan sebagai rasio atau persentase dan karenanya bebas unit
Ukuran mutlak terkait dengan distribusi itu sendiri	Ukuran relatif digunakan Untuk membandingkan variabilitas

	antara dua seri atau lebih. Untuk memeriksa keakuratan relatif data
--	--

## 9.4 Fungsi Ukuran Penyebaran

Ukuran Penyebaran digunakan untuk menghitung statistik seperti nilai-F, tingkat korelasi, regresi, dan digunakan untuk membandingkan variabilitas dalam data yang diperoleh. Adapun fungsi ukuran penyebaran adalah:

Untuk mengetahui apakah rata-rata atau median/mode yang dikerjakan dapat diandalkan. Jika variasinya kecil maka kita dapat menyatakan bahwa rata-rata yang dihitung dapat diandalkan, tetapi jika variasi terlalu besar, rata-ratanya mungkin salah; Digunakan dalam deret waktu untuk mengatasi berbagai fluktuasi karena faktor waktu, faktor musiman, dan lain-lain;

Dispersi memberi kita gambaran jika variabilitas berdampak negatif pada data dan dengan demikian membantu dalam mengendalikan variabilitas (Tendencies, 2017).

## 9.5 Pengukuran Dispersi

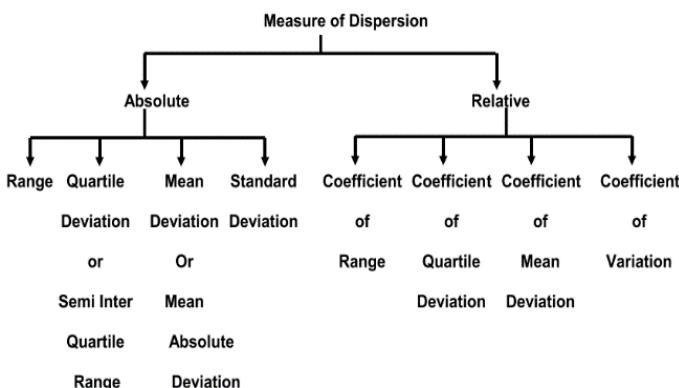
Dalam mengukur dispersi, sangat penting untuk mengetahui jumlah variasi (ukuran absolut) dan derajat variasi (ukuran relatif). Ukuran penyebaran adalah statistik deskriptif yang menggambarkan seberapa mirip satu set skor satu sama lain. Semakin besar kesamaan skor satu sama lain, semakin rendah ukuran dispersi. Semakin sedikit kesamaan skor satu sama lain, semakin tinggi ukuran dispersi. Secara umum, semakin banyak penyebaran suatu distribusi, semakin besar ukuran penyebarannya. Untuk mengatakannya secara ringkas, variasi antara nilai data dalam sampel disebut dispersi (Tendencies, 2017).

Pengukuran dispersi dapat dilakukan secara:

1. Mutlak: Mengukur penyebaran dalam unit asli data.  
Variabilitas di 2 atau lebih distribusi dapat dibandingkan asalkan diberikan dalam unit yang sama dan memiliki rata-rata yang sama;

2. Relatif: Ukuran dispersi bebas dari satuan pengukuran data.

Rasio ukuran dispersi absolut terhadap rata-rata, dari mana deviasi absolut diukur (Kumar, 2020).



**Gambar 9.1:** Pengukuran Dispersi (Shah, 2019)

Ukuran tendensi sentral memang sangat penting keberadaannya, namun kegunaannya agak terbatas. Artinya ukuran tengah bermanfaat tetapi terbatas kegunaannya. Oleh karena itu, seseorang harus memikirkan ukuran lain juga di sekitar pusat dan salah satu ukuran tersebut adalah penyebaran atau variabilitas di tengahnya. Ukuran dispersi yang paling umum digunakan adalah kisaran, persentil, dan deviasi standar. Range adalah perbedaan antara nilai data terbesar dan terkecil. Oleh karena itu, semakin tersebar nilai datanya, semakin besar kisarannya. Namun, jika beberapa pengamatan relatif jauh dari tengah tetapi sisanya relatif dekat ke tengah, kisaran tersebut dapat memberikan ukuran dispersi yang terdistorsi (Tendencies, 2017).

Dispersi adalah sejauh mana nilai dalam suatu distribusi berbeda dari rata-rata distribusi. Untuk mengukur sejauh mana variasi tersebut, ada ukuran tertentu yaitu: (i) Range (ii) Quartile Deviation (iii) Mean Deviation (iv) Standar Deviasi. Selain pengukuran yang memberikan nilai numerik, terdapat metode grafik untuk memperkirakan dispersi. Range dan deviasi kuartil mengukur dispersi dengan menghitung penyebaran di mana nilai-nilai itu berada. Deviasi rata-rata dan deviasi standar menghitung sejauh mana nilai-nilai berbeda dari rata-rata (Berman and Wang, 2020).

## 9.6 Jenis Ukuran Variabilitas/Dispersi

Ada 5 jenis ukuran variabilitas yaitu (Tendencies, 2017):

1. Metode batas
  - Kisaran
  - Kisaran antar-kuatril
  - Kisaran persentil
2. Metode Rata-rata
  - Deviasi kuartil
  - Deviasi rata-rata
  - Deviasi Standar dan
  - Pengukuran lainnya.

Ada empat ukuran dispersi yaitu (Tendencies, 2017):

1. Range
2. Deviasi Kuartil
3. Deviasi Rata-rata
4. Deviasi Standar

## 9.7 Range

Range adalah perbedaan antara pengamatan terbesar dan terkecil dalam data. Keuntungan utama dari ukuran dispersi ini adalah mudah untuk dihitung. Di sisi lain, memiliki banyak kekurangan dan sangat sensitif serta tidak semua menggunakan pengamatan dalam kumpulan data (Swinscow and Campbell, 2002). Range lebih informatif untuk memberikan nilai minimum dan maksimum daripada memberikan kisaran.

Range atau jangkauan adalah ukuran variabilitas yang paling sederhana. Dapat didefinisikan sebagai perbedaan antara skor tertinggi dan terendah dalam distribusi (Tendencies, 2017) atau disebut juga selisih antara nilai data terbesar dan nilai data terkecil (Latief, 2009). Range ( $R$ ) adalah selisih antara nilai

terbesar (L) dan nilai terkecil (S) dalam suatu distribusi (Berman and Wang, 2020).

Jadi,

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

$$R = L - S$$

Nilai jangkauan yang lebih tinggi menunjukkan dispersi yang lebih tinggi dan sebaliknya. Range terlalu dipengaruhi oleh nilai ekstrim yang tidak didasarkan pada semua nilai. Selama nilai minimum dan maksimum tetap dan tidak berubah, maka setiap perubahan pada nilai tidak memengaruhi range dan tidak dapat dihitung sebagai distribusi frekuensi terbuka. Terlepas dari beberapa batasan, jangkauan atau range dipahami dan sering digunakan karena kesederhanaannya. Distribusi terbuka adalah distribusi yang batas bawahnya kelas rendah atau batas atas dari kelas tertinggi atau keduanya tidak ditentukan (Berman and Wang, 2020).

### **Data Tunggal**

Untuk data tunggal atau data tidak berkelompok, range dihitung dengan formula:

$$R = X_{\text{maks}} - X_{\text{min}} = L - S$$

Contoh: Tentukan range dari data berikut ini:

5 7 9 4 3 9 10 6 4 7 9 5

Jawab:

$$R = X_{\text{maks}} - X_{\text{min}} = L - S$$

$$10 - 3 = 7$$

Jadi range dari data di atas adalah 7.

### **Data Berkelompok**

Range untuk data berkelompok adalah selisih antara titik tengah interval terakhir dan titik tengah interval kelas pertama.

Contoh: Tentukan jangkauan dari data berikut ini:

Interval	Frekuensi
24-30	13
31-37	17
38-44	10
45-51	5
52-58	15

Jawab:

Interval	Frekuensi
24-30	27
31-37	34
38-44	41
45-51	48
52-58	55

$$\text{Range} = X_{\text{maks}} - X_{\text{min}} = 55 - 27 = 28$$

Jadi jangkauan = 28

### Batasan Range

1. Penghitungan rentang hanya didasarkan pada dua nilai ekstrem dalam kumpulan data dan tidak mempertimbangkan nilai lain dari kumpulan data. Kadang-kadang nilai ekstrim dari dua kumpulan data yang berbeda mungkin hampir sama, tetapi dua kumpulan data mungkin berbeda dalam penyebaran;
2. Nilainya sensitif terhadap perubahan pengambilan sampel. Kisarannya lebih bervariasi dengan fluktuasi pengambilan sampel,

- yaitu sampel yang berbeda dengan ukuran yang sama dari populasi yang sama mungkin memiliki kisaran yang berbeda;
3. Nilainya dipengaruhi oleh ukuran sampel. Dalam banyak jenis distribusi termasuk distribusi normal, kisarannya bergantung pada ukuran sampel. Jika ukuran sampel besar, nilai kisarannya juga besar (Tendencies, 2017).

## 9.8 Range Interkuartil

Range interkuartil didefinisikan sebagai selisih antara tanggal 25 dan ke-75 persentil (juga disebut kuartil pertama dan ketiga). Oleh karena itu kisaran interkuartil menggambarkan 50% pengamatan tengah. Jika jarak interkuartil besar, itu berarti 50% pengamatan tengah diberi jarak lebar. Keuntungan penting dari jarak interkuartil adalah bahwa hal itu dapat digunakan sebagai ukuran variabilitas jika nilai ekstrim tidak dicatat dengan tepat (seperti dalam kasus interval kelas terbuka dalam distribusi frekuensi) (Sundaram, Dwivedi and Sreenivas, 2010). Fitur menguntungkan lainnya adalah tidak terpengaruh oleh nilai-nilai ekstrim. Kerugian utama dalam menggunakan rentang interkuartil sebagai ukuran dispersi adalah tidak dapat menerima manipulasi matematika.

## 9.9 Deviasi Kuartil

Keberadaan bahkan satu nilai yang sangat tinggi atau rendah dalam suatu distribusi dapat mengurangi kegunaan jangkauan sebagai ukuran dispersi. Dengan demikian, mungkin memerlukan ukuran yang tidak terlalu terpengaruh oleh luaran. Dalam situasi seperti itu, jika seluruh data dibagi menjadi empat bagian yang sama, masing-masing berisi 25% nilai, akan didapatkan nilai kuartil dan median. Kuartil atas dan bawah ( $Q_3$  dan  $Q_1$ ) digunakan untuk menghitung rentang antar kuartil yaitu  $Q_3 - Q_1$ . Range interkuartil didasarkan pada 50% tengah dari nilai dalam suatu distribusi dan, oleh karena itu, tidak terpengaruh oleh nilai ekstrem. Setengah dari range antar kuartil disebut deviasi kuartil (Q.D.) (Berman and Wang, 2020).

Jadi:

$$Q.D. = Q3 - Q1$$

2

Q.D. juga disebut Jangkauan Kuartil Semi-Antar.

Contoh Perhitungan Range dan Q.D. untuk data yang tidak dikelompokkan  
Hitung rentang dan Q.D. dari pengamatan berikut:

20, 25, 29, 30, 35, 39, 41, 48, 51, 60 dan 70

Rentangnya jelas  $70 - 20 = 50$

Untuk Q.D., kita perlu menghitung nilai Q3 dan Q1.

Q1 adalah ukuran nilai  $n + 1$

4

$n$  menjadi 11, Q1 adalah ukuran nilai ke-3.

Karena nilai sudah diatur dalam urutan naik, dapat dilihat bahwa Q1, nilai ke-3 adalah 29.

Q3 adalah ukuran nilai  $3n + 1$

4

yaitu nilai ke-9 yaitu 51. Oleh karena itu  $Q3 = 51$

$$Q.D. = Q3 - Q1$$

2

$$Q.D. = 51 - 29 = 11$$

2

Q.D. adalah perbedaan rata-rata Kuartil dari median. Deviasi kuartil akan memberikan perbedaan rata-rata antara median kedua kelompok, dari median seluruh kelompok. Deviasi kuartil umumnya dapat dihitung untuk distribusi ujung terbuka dan tidak terlalu terpengaruh oleh nilai ekstrem (Berman and Wang, 2020).

### 9.9.1 Sifat Deviasi Kuartil

Deviasi kuartil terkait erat dengan median karena median responsif terhadap jumlah skor yang diletakkan di bawahnya daripada posisi persisnya dan Q1 dan Q3 didefinisikan dengan cara yang sama. Median dan deviasi kuartil memiliki sifat yang sama. Baik median dan deviasi Kuartil tidak dipengaruhi oleh nilai ekstrim (Tendencies, 2017). Dalam distribusi non-simetris, dua kuartil Q1 dan Q3 berada pada jarak yang sama dari median-  $Q1 = Q3 - \text{Median}$ . Jadi Median Quartile  $\pm$  Deviation mencakup tepat 50 persen dari nilai yang diamati dalam data. Jika distribusi terbuka maka Deviasi kuartil adalah satu-satunya ukuran variabilitas yang masuk akal untuk dihitung (Tendencies, 2017).

### 9.9.2 Batasan Penyimpangan Kuartil

1. Nilai deviasi kuartil didasarkan pada nilai tengah 50 persen, tidak berdasarkan semua pengamatan;
2. Nilai deviasi kuartil dipengaruhi oleh fluktuasi sampling;
3. Nilai deviasi kuartil tidak dipengaruhi oleh distribusi nilai individu dalam interval 50 persen tengah nilai observasi (Tendencies, 2017).

## 9.10 Deviasi Rata-Rata

Sebelum membahas deviasi rata-rata terlebih dahulu kita harus mengetahui tentang pengertian deviasi. Skor deviasi mengungkapkan skor yang menunjukkan berapa banyak skor poin yang terletak di atas atau di bawah rata-rata distribusi. Skor deviasi dapat didefinisikan sebagai (X-Mean) yaitu ketika kita mengurangi rata-rata dari masing-masing skor mentah, skor deviasi yang dihasilkan menyatakan posisi skor, relatif terhadap rata-rata.

Perlu diingat bahwa dispersi didefinisikan sebagai sejauh mana nilai-nilai berbeda dari rata-rata. Rentang dan deviasi kuartil tidak berguna dalam mengukur, seberapa jauh nilainya, dari rata-ratanya. Namun, dengan menghitung penyebaran nilai, dapat memberikan gambaran yang baik tentang penyebaran. Dua ukuran yang didasarkan pada deviasi nilai dari rata-ratanya adalah Deviasi Rata-rata dan Deviasi Standar. Karena rata-rata adalah nilai sentral, beberapa penyimpangan bernilai positif dan beberapa negatif. Jika ini

ditambahkan sebagaimana adanya, jumlahnya tidak akan mengungkapkan apa pun. Faktanya, jumlah penyimpangan dari rata-rata selalu nol. Perhatikan dua set nilai berikut (Berman and Wang, 2020).

Set A: 5, 9, 16

Set B: 1, 9, 20

Dapat dilihat bahwa nilai-nilai dalam Set B lebih jauh dari rata-rata dan karenanya lebih tersebar daripada nilai-nilai di Set A. Mean Deviasi mencoba untuk mengatasi masalah ini dengan mengabaikan tanda-tanda deviasi, yaitu menganggap semua deviasi positif. Untuk deviasi standar, pertama-tama deviasi dikuadratkan dan dirata-ratakan, lalu akar kuadrat dari rata-rata ditemukan (Berman and Wang, 2020). Deviasi rata-rata adalah alat statistik yang tepat untuk memperkirakan jarak rata-rata yang ditempuh. Deviasi rata-rata adalah rata-rata aritmatika dari perbedaan nilai dari rata-ratanya. Rata-rata yang digunakan adalah rata-rata aritmatika atau median (Berman and Wang, 2020). Deviasi rata-rata didasarkan pada semua nilai. Perubahan bahkan pada satu nilai akan memengaruhinya. Deviasi mean adalah yang terkecil jika dihitung dari median, yaitu akan lebih tinggi jika dihitung dari mean (Berman and Wang, 2020).

Deviasi Standar adalah akar kuadrat positif dari rata-rata deviasi kuadrat dari mean. Jadi jika ada lima nilai  $x_1, x_2, x_3, x_4$  dan  $x_5$ , pertama mean dihitung. Kemudian penyimpangan nilai dari mean dihitung. Penyimpangan ini kemudian dikuadratkan. Rata-rata dari deviasi kuadrat ini adalah varians. Akar kuadrat positif dari varians adalah simpangan baku (Berman and Wang, 2020).

## 9.11 Standar Deviasi

Standar deviasi (SD) adalah ukuran dispersi yang paling umum digunakan. Ini adalah ukuran penyebaran data tentang mean. SD adalah akar kuadrat dari jumlah deviasi kuadrat dari mean dibagi jumlah observasi (Manikandan, 2011).

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

Rumus ini adalah rumus definisi dan untuk perhitungan, rumus yang lebih mudah digunakan. Rumus komputasi juga menghindari kesalahan pembulatan selama perhitungan (Manikandan, 2011).

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

Dalam kedua rumus ini,  $n - 1$  digunakan sebagai pengganti  $n$  pada penyebut, karena ini menghasilkan perkiraan SD populasi yang lebih akurat. Alasan mengapa SD adalah ukuran dispersi yang sangat berguna adalah, jika pengamatan berasal dari distribusi normal (Gravetter and Wallnau, 2017), maka 68% observasi berada di antara mean  $\pm 1$  SD 95% observasi berada di antara mean  $\pm 2$  SD dan 99,7% observasi berada di antara mean  $\pm 3$  SD.

Keuntungan lain dari SD adalah bahwa bersama dengan mediannya dapat digunakan Address untuk mendeteksi kemiringan. Kerugian dari SD adalah bahwa ini merupakan ukuran dispersi yang tidak tepat untuk data miring. SD digunakan sebagai ukuran dispersi ketika mean digunakan sebagai ukuran tendensi sentral (yaitu, untuk data numerik simetris). Untuk data ordinal atau data numerik miring, median dan rentang interkuartil digunakan (Dawson, 2004).

# **Bab 10**

## **Normalitas dan Homogenitas**

### **10.1 Pendahuluan**

Pada bab ini, akan membahas konsep Normalitas dan Homogenitas dalam statistik Teknologi Informasi. Dalam analisis statistik, semua uji parametrik mengasumsikan beberapa karakteristik tertentu tentang data, yang juga dikenal sebagai asumsi. Pelanggaran asumsi ini mengubah kesimpulan penelitian dan interpretasi hasil. Oleh karena itu semua penelitian, baik untuk artikel jurnal, tesis, atau disertasi, harus mengikuti asumsi ini untuk interpretasi yang akurat. Tergantung pada analisis parametrik, asumsi tersebut bervariasi.

Berikut ini adalah asumsi data yang biasa ditemukan dalam penelitian statistik (Baron, 2013):

1. Asumsi normalitas: Sebagian besar uji parametrik mengharuskan asumsi normalitas dipenuhi. Normalitas berarti distribusi tes berdistribusi normal (atau berbentuk lonceng) dengan mean 0, dengan 1 deviasi standar, dan kurva berbentuk lonceng simetris. Untuk menguji asumsi normalitas, langkah-langkah dan pengujian berikut dapat diterapkan:
  1. Skewness dan Kurtosis: Untuk menguji asumsi distribusi normal, Skewness harus berada dalam kisaran  $\pm 2$ . Nilai Kurtosis harus berada dalam kisaran  $\pm 7$

3. Uji W Shapiro-Wilk: Sebagian besar peneliti menggunakan uji ini untuk menguji asumsi normalitas. Uji Wilk seharusnya tidak signifikan untuk memenuhi asumsi normalitas.
4. Uji Kolmogorov-Smirnov: Dalam kasus sampel yang besar, sebagian besar peneliti menggunakan uji K-S untuk menguji asumsi normalitas. Pengujian ini seharusnya tidak signifikan untuk memenuhi asumsi normalitas.
5. Metode grafis untuk uji normalitas:  
Plot Q-Q: Kebanyakan peneliti menggunakan plot Q-Q untuk menguji asumsi normalitas. Dalam metode ini, nilai yang diamati dan nilai yang diharapkan diplot pada grafik. Jika nilai yang diplot lebih bervariasi dari garis lurus, maka data tidak terdistribusi normal. Jika tidak, data akan didistribusikan secara normal.
6. Asumsi homogenitas varians:  
Uji Levene: Untuk menguji asumsi homogenitas varians digunakan uji Levene. Uji Levene digunakan untuk menilai apakah kelompok memiliki varian yang sama. Pengujian ini seharusnya tidak signifikan untuk memenuhi asumsi persamaan varians
7. Homogenitas asumsi matriks varian-kovarian:  
Uji Box's M: Uji ini digunakan untuk menguji homogenitas multivariat asumsi matriks varians-kovarian. Nilai yang tidak signifikan dari uji Box's M menunjukkan bahwa kelompok tersebut tidak berbeda satu sama lain dan akan memenuhi asumsi
8. Keacakan: Sebagian besar statistik berasumsi bahwa observasi sampel adalah acak. Uji Jalan digunakan untuk menguji asumsi keacakan.
9. Multikolinieritas: Multikolinieritas berarti bahwa variabel-variabel yang diminati berkorelasi tinggi, dan korelasi yang tinggi tidak boleh ada di antara variabel-variabel yang diminati. Untuk menguji asumsi multikolinearitas dapat digunakan indeks VIF dan Condition, khususnya dalam analisis regresi. Nilai  $VIF > 10$  menunjukkan adanya multikolinearitas dan asumsi dilanggar.

Ketika mengambil nama uji normalitas, diperlukan bagaimana menemukan tes terbaik untuk mengetahui apakah suatu variabel berdistribusi normal atau tidak. Karena banyak algoritme bekerja dengan asumsi bahwa data berdistribusi normal, untuk mengeksplorasi ini, digunakan dataset yang memiliki sembilan variabel di mana delapan adalah numerik. Apa Asumsi Normalitas? Asumsi normalitas berarti harus memastikan data secara kasar sesuai dengan bentuk kurva lonceng sebelum menjalankan uji statistik atau regresi tertentu. Pengujian yang membutuhkan data berdistribusi normal antara lain: Uji-t Sampel Independen.

Asumsi normalitas berarti harus memastikan data secara kasar sesuai dengan bentuk kurva lonceng sebelum menjalankan uji statistik atau regresi tertentu. Tes yang membutuhkan data terdistribusi normal meliputi:

1. Uji-t Sampel Independen.
2. Pemodelan Linier Hirarkis.
3. ANCOVA.
4. Goodness of Fit Test.

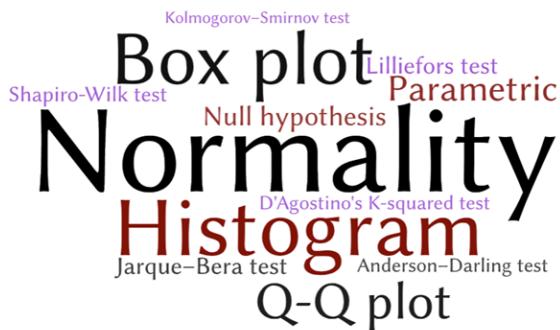
Dalam statistik, homogenitas dan kebalikannya, heterogenitas, muncul dalam mendeskripsikan properti dari sebuah dataset, atau beberapa dataset. Mereka berhubungan dengan validitas dari asumsi yang sering digunakan bahwa properti statistik, dari salah satu bagian dari keseluruhan dataset adalah sama dengan bagian lainnya. Dalam meta-analisis yang menggabungkan data dari beberapa studi, homogenitas mengukur perbedaan atau persamaan antara beberapa studi (lihat juga Studi heterogenitas) (Naik and Kumar, 2016).

Homogenitas dapat dipelajari hingga beberapa derajat kompleksitas. Misalnya, pertimbangan homoskedastisitas memeriksa seberapa besar variabilitas nilai data berubah di seluruh kumpulan data. Namun, pertanyaan homogenitas berlaku untuk semua aspek distribusi statistik, termasuk parameter lokasi. Dengan demikian, studi yang lebih rinci akan mengkaji perubahan pada keseluruhan distribusi marginal. Studi tingkat menengah mungkin beralih dari melihat variabilitas ke mempelajari perubahan kemiringan tersebut. Selain itu, pertanyaan tentang homogenitas juga berlaku untuk distribusi gabungan.

Konsep homogenitas dapat diterapkan dalam banyak cara berbeda dan, untuk jenis analisis statistik tertentu, digunakan untuk mencari properti lebih lanjut yang mungkin perlu diperlakukan sebagai beragam dalam kumpulan data setelah beberapa jenis awal non-homogenitas telah ditangani.

## 10.2 Normalitas

Pendekatan informal untuk menguji normalitas adalah membandingkan histogram data sampel dengan kurva probabilitas normal. Distribusi empiris dari data (histogram) harus berbentuk lonceng dan menyerupai distribusi normal. Ini mungkin sulit untuk melihat apakah sampelnya kecil. Dalam memeriksa normalitas secara visual dengan memplot distribusi frekuensi, juga disebut histogram, dari data dan secara visual membandingkannya dengan distribusi normal (dilapisi merah). Dalam distribusi frekuensi, setiap titik data dimasukkan ke dalam bin diskrit, misalnya (-10, -5], (-5, 0], (0, 5], dll. Power adalah ukuran yang paling sering dari nilai uji normalitas — kemampuan untuk mendeteksi apakah sampel berasal dari distribusi non-normal (11). Beberapa peneliti merekomendasikan uji Shapiro-Wilk sebagai pilihan terbaik untuk menguji normalitas data (11). Secara umum pengujian normalitas dengan grafik dapat menggunakan beberapa media, seperti pada gambar 10.1



**Gambar 10.1:** Visual plot grafik dalam uji normalitas (Mishra, 2020)

Ada dua cara utama untuk menguji normalitas: mengamati grafik, atau menjalankan pengujian yang dirancang khusus untuk menguji normalitas (Kotz and Johnson, 100AD). Data tidak harus normal sepenuhnya. Namun, data yang pasti tidak memenuhi asumsi normalitas akan memberi hasil yang buruk untuk jenis pengujian tertentu (yaitu pengujian yang menyatakan bahwa asumsi tersebut harus dipenuhi!). Seberapa dekat data harus memenuhi uji normalitas?

Ada beberapa cara untuk menguji normalitas fitur / atribut tertentu, tetapi pertama-tama kita perlu mengetahui mengapa penting untuk mengetahui

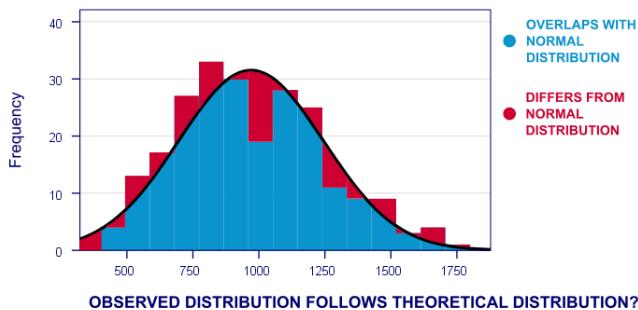
apakah fitur / atribut kita terdistribusi secara normal. Dalam masalah pembelajaran mesin apa pun, kami biasanya memiliki banyak tes (sering dikenal sebagai uji signifikansi statistik) untuk memahami informasi tersembunyi di dalam data seperti apakah usia berdampak pada premi asuransi jiwa dan banyak lagi.

Faktanya adalah banyak pengujian statistik ini memiliki asumsi yang telah ditentukan sebelumnya dan seringkali salah satunya adalah distribusi normal tetapi meskipun demikian harus benar-benar memahami asumsi di balik uji statistik, yang diberikan sebelum menggunakannya, jika tidak maka akan sia-sia di akhir proses.

Lalu bagaimana cara untuk mengetahui normalitas data? Histogram dan plot Q-Q digunakan untuk menilai normalitas distribusi secara tetapi visual ini subjektif untuk interpretasi kita sehingga kita perlu memiliki beberapa signifikansi matematis yang kita semua bisa sepakati. Ada dua jenis uji normalitas yaitu uji Shapiro Wilk dan uji Kolmogorov-Smirnov.

Tes Shapiro Wilk, seperti kebanyakan uji signifikansi statistik, ada hipotesis nol dan hipotesis alternatif, tidak terkecuali uji wilk Shaprio. Dalam pengujian ini, hipotesis nol ( $H_0$ ) menyatakan bahwa data yang diamati tidak berbeda nyata dengan berdistribusi normal sedangkan hipotesis alternatif ( $H_1$ ) adalah berbeda nyata dengan berdistribusi normal.

Penggambaran di bawah ini menunjukkan kurva bentuk lonceng (distribusi normal) telah digambar di atas data yang diamati. Ini kurang teknis tetapi intuisi yang baik untuk memahami tes wilk Shaprio. Ini menghitung persentase tumpang tindih (area biru) untuk mengukur kesamaan antara distribusi normal dan distribusi yang diamati. Nilai-P kemudian menentukan peluang untuk memperoleh persentase tumpang tindih yang ekstrim seperti yang diamati untuk menolak atau menerima hipotesis nol. yaitu apakah data berdistribusi normal atau tidak.

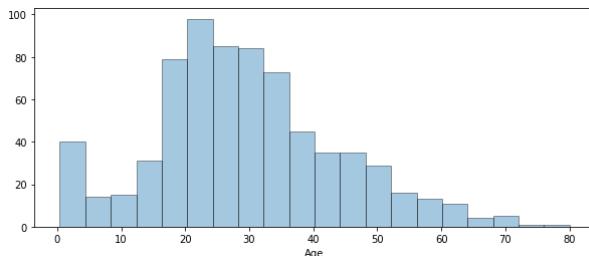


**Gambar 10.2:** Shapiro-Wilk Uji Normal (Farooqui, 2020)

Bilangan tunggal digunakan untuk mengukur kemiripan antara dua distribusi yang dikenal sebagai statistik-W. Nilai-nilai W yang kecil adalah bukti penyimpangan dari normalitas, itu berarti semakin banyak tumpang tindih yang miliki dalam data, statistik W semakin mendekati 1 (nilai maksimum adalah 1 yang berarti distribusi yang diamati dan distribusi normal benar-benar tumpang tindih). Itu dapat dihitung dari rumus yang diberikan.

Akhirnya jika nilai p dari uji Shaprio wilk kurang dari tingkat signifikansi (sering 0,05) maka kita dapat menolak hipotesis nol yaitu data berbeda signifikan dari distribusi normal, sebaliknya data tidak berbeda signifikan dari distribusi normal. Batasan tes wilk Shaprio, Namun, uji wilk Shaprio memiliki batasan terkait ukuran sampel. Sangat tidak mungkin untuk menolak hipotesis nol untuk ukuran sampel kecil sedangkan, hipotesis nol dapat menolak berdasarkan perbedaan yang sangat tidak signifikan dari distribusi normal dalam sampel besar.

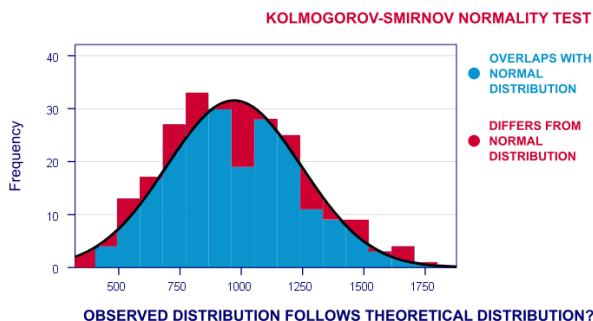
Shapiro Wilk Test dengan python, yaitu dalam python, Scipy menyediakan fungsi untuk tes wilk Shaprio. (scipy.stats.shapiro). Dalam potongan kode berikut, tes shapiro wilk digunakan untuk memverifikasi distribusi normal fitur usia dari kumpulan data titanic. Hal ini dapat dilihat dari histogram bahwa distribusi sedikit lepas dari distribusi normal (bentuk lonceng) maka p-value dan W-statistik secara numerik memverifikasi bahwa distribusi yang diamati berbeda secara signifikan dari distribusi normal (kita dapat menolak hipotesis nol bahwa itu normal. didistribusikan).



**Gambar 10.3:** Tes Kolmogorov-Smirnov (Mishra, 2020)

Hipotesis nol dan alternatif dalam uji Kolmogorov-Smirnov pada gambar 10.3 sama dengan uji Shapiro wilk yaitu  $H_0$  (data tidak berbeda nyata dengan distribusi normal) dan  $H_1$  (data berbeda signifikan dengan distribusi normal).

Intuisi di balik tes Kolmogorov-Smirnov mirip dengan tes wilk Shapiro dengan cara kurva bentuk lonceng (distribusi normal) dapat dilihat pada diagram di bawah ini tetapi dalam kasus tes Kolmogorov-Smirnov kami lebih tertarik untuk menemukannya persentase perbedaan (area merah) antara distribusi normal dan yang diamati.



**Gambar 10.4:** kolmogrov smirnov normality test (Mishra, 2020)

Lebih tepat, ini adalah perbandingan antara distribusi probabilitas kumulatif dari data yang diamati dan probabilitas kumulatif dari distribusi normal. Statistik Kolmogorov-Smirnov juga dikenal sebagai statistik Kolmogorov D pada persamaan 10.1 dapat dihitung dari rumus yang diberikan di bawah ini. Di mana  $F_o(X)$  adalah probabilitas kumulatif dari data yang diamati dan  $F_r(X)$  adalah distribusi normal teoritis. Jika data yang diamati didistribusikan secara normal maka D menyatu dengan 0.

$$D = \text{Maximum}|F_o(X) - F_r(X)| \quad (10.1)$$

Akhirnya nilai kritis dihitung berdasarkan tingkat signifikansi (sering 0,05) dari tabel Kolmogorov-Smirnov yang jika lebih besar dari D-Statistik yang dihitung (seperti nilai kritis D < D) maka kita tidak dapat menolak hipotesis nol yaitu data yang diamati adalah tidak berbeda nyata dengan distribusi normal.

Batasan tes Kolmogorov-Smirnov: Tes ini lebih sensitif di dekat pusat distribusi daripada di bagian ekor. Tes Kolmogorov-Smirnov dengan python yaitu dalam python, Scipy menyediakan fungsi untuk Kolmogorov-Smirnov. (scipy.stats.kstest). Dalam potongan kode berikut, Kolmogorov-Smirnov digunakan untuk memverifikasi distribusi normal fitur usia dari kumpulan data titanic. Untuk memverifikasi apakah d-statistik kurang atau lebih besar dari nilai kritis-d kita perlu melihat ke dalam tabel KS itulah sebabnya kita sering mempertimbangkan nilai-p. Jika p-value lebih kecil dari tingkat signifikansi (0,05) maka hipotesis nol dapat ditolak. Seperti dalam kasus ini usia dalam kumpulan data titanic berbeda secara signifikan dari distribusi normal.

Dalam pengujian normalitas ada beberapa opsi untuk menguji normalitas. Baik pengujian dengan menggunakan tools dan software dalam teknologi Informatika.

1. Uji normalitas chi-square. Dalam hal ini dapat menggunakan uji chi kuadrat untuk normalitas. Keuntungannya adalah relatif mudah digunakan, tetapi ini bukan pengujian yang sangat kuat. Jika memiliki sampel kecil (di bawah 20), itu mungkin satu-satunya tes yang dapat gunakan. Untuk sampel yang lebih besar, lebih baik memilih opsi lain.
2. Tes D'Agostino-Pearson. Ini menggunakan skewness dan kurtosis untuk melihat apakah data cocok dengan data normal. Ini membutuhkan ukuran sampel lebih dari 20.
3. Test Jarque-Bera. Tes umum ini juga relatif mudah. Seperti D'Agostino-Pearson, ide dasarnya adalah menguji kemiringan dan kurtosis data untuk melihat apakah cocok dengan yang diharapkan dari distribusi normal. Semakin besar statistik JB, semakin banyak data yang menyimpang dari biasanya.
4. Tes Kebaikan Kolmogorov-Smirnov. Ini membandingkan data dengan distribusi yang diketahui (yaitu distribusi normal).

5. Uji Lilliefors. Tes Lilliefors menghitung statistik uji T yang dapat bandingkan dengan nilai kritis. Jika statistik pengujian lebih besar dari nilai kritis, itu tandanya data tidak normal. Ini juga menghitung nilai p untuk distribusi, yang bandingkan dengan tingkat signifikansi.
6. Shapiro-Wilk Test Tes ini akan memberi tahu jika sampel acak berasal dari distribusi normal. Tes memberi nilai W; nilai kecil menunjukkan sampel tidak berdistribusi normal.

## 10.3 Homogenitas

Uji kesesuaian dapat digunakan untuk memutuskan apakah suatu populasi cocok dengan distribusi tertentu, tetapi tidak cukup untuk memutuskan apakah dua populasi mengikuti distribusi tidak diketahui yang sama. Uji beda, disebut uji homogenitas, dapat digunakan untuk menarik kesimpulan apakah dua populasi memiliki distribusi yang sama. Untuk menghitung statistik uji untuk uji homogenitas, ikuti prosedur yang sama dengan uji independensi.

Secara umum uji homogenitas adalah metode uji statistik, yang memeriksa apakah dua (atau lebih) dataset berasal dari distribusi yang sama atau tidak. Dalam deret waktu, uji homogenitas diterapkan untuk mendeteksi satu (atau lebih) perubahan / titik putus dalam rangkaian tersebut. Titik henti ini terjadi saat kumpulan data mengubah distribusinya. Banyak analisis statistik membutuhkan kumpulan data yang homogen. Itulah mengapa ini merupakan tes penting dalam analisis statistik. Untuk menganalisis homogenitas kumpulan data maka diperlukan beberapa hal pemahaman terhadap perhitungan mediannya, kurangi median, dari setiap nilai dalam kumpulan data. Hitung berapa kali data akan berjalan di atas atau di bawah median (yaitu, persistensi nilai positif atau negatif). Gunakan tabel signifikansi untuk menentukan ambang batas homogenitas.

Dalam statistik informasi teknologi dikenal uji kemandirian dan homogenitas. Apa perbedaan antara uji kemandirian dan uji homogenitas?

Perbedaannya terletak pada desain. Dalam uji independensi, unit observasi dikumpulkan secara acak dari satu populasi dan dua variabel kategori diamati untuk setiap unit. Dalam uji homogenitas, pengumpulan data dilakukan secara random sampling dari masing-masing sub kelompok secara terpisah.

Uji homogenitas membandingkan proporsi tanggapan dari dua atau lebih populasi berkaitan dengan variabel dikotomis (misalnya, pria / wanita, ya / tidak) atau variabel dengan lebih dari dua kategori hasil. Uji homogenitas chi-square adalah uji nonparametrik yang digunakan dalam situasi di mana variabel dependen bersifat kategoris. Data dapat disajikan dengan menggunakan tabel kontingensi dimana populasi dan kategori variabel adalah label baris dan kolom. Hipotesis nol menyatakan bahwa semua populasi homogen mengenai proporsi kategori variabel kategori. Jika hipotesis nol ditolak, disimpulkan bahwa proporsi di atas berbeda pada populasi yang diamati. Uji chi-square statistik homogenitas dihitung dengan cara yang persis sama dengan uji chi-square untuk statistik independensi. Perbedaan antara kedua pengujian ini terdiri dari pernyataan hipotesis nol, logika yang mendasari, dan prosedur pengambilan sampel.

Dalam statistik, homogenitas dan kebalikannya, heterogenitas, muncul dalam mendeskripsikan properti dari sebuah dataset, atau beberapa dataset. Mereka berhubungan dengan validitas dari asumsi yang sering digunakan bahwa properti statistik, dari salah satu bagian dari keseluruhan dataset adalah sama dengan bagian lainnya. Dalam meta-analisis yang menggabungkan data dari beberapa studi, homogenitas mengukur perbedaan atau persamaan antara beberapa studi (lihat juga Studi heterogenitas).

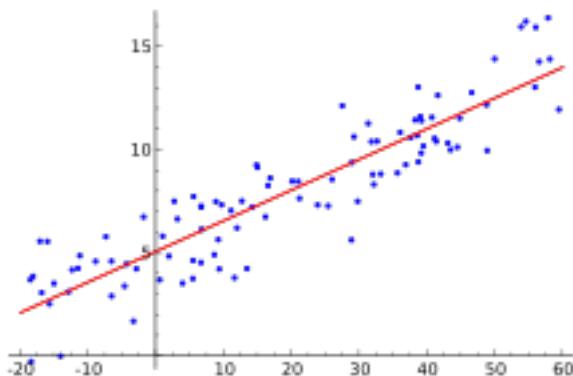
Homogenitas dapat dipelajari hingga beberapa derajat kompleksitas. Misalnya, pertimbangan homoskedastisitas memeriksa seberapa besar variabilitas nilai data berubah di seluruh kumpulan data, (Sinambela and Suryaningsih, 2014). Namun, pertanyaan homogenitas berlaku untuk semua aspek distribusi statistik, termasuk parameter lokasi. Dengan demikian, studi yang lebih rinci akan mengkaji perubahan pada keseluruhan distribusi marginal. Studi tingkat menengah mungkin beralih dari melihat variabilitas ke mempelajari perubahan kemiringan tersebut. Selain itu, pertanyaan tentang homogenitas juga berlaku untuk distribusi gabungan.

Konsep homogenitas dapat diterapkan dalam banyak cara berbeda dan, untuk jenis analisis statistik tertentu, digunakan untuk mencari properti lebih lanjut yang mungkin perlu diperlakukan sebagai beragam dalam kumpulan data setelah beberapa jenis awal non-homogenitas telah ditangani dengan:

### 1. Regression

Perbedaan nilai tipikal di seluruh kumpulan data mungkin awalnya ditangani dengan membangun model regresi menggunakan variabel penjelas tertentu untuk menghubungkan variasi dalam nilai tipikal dengan

kuantitas yang diketahui. Kemudian harus ada tahap analisis selanjutnya untuk memeriksa apakah kesalahan dalam prediksi dari regresi berperilaku dengan cara yang sama di seluruh kumpulan data. Dengan demikian, pertanyaannya menjadi salah satu homogenitas distribusi residual, karena variabel penjelas berubah pada gambar 10.5.



**Gambar 10.5:** regression analysis

2. Deret waktu

Tahapan awal dalam analisis deret waktu mungkin melibatkan penggambaran nilai terhadap waktu untuk menguji homogenitas deret dalam berbagai cara: stabilitas lintas waktu sebagai lawan tren; stabilitas fluktuasi lokal dari waktu ke waktu.

3. Menggabungkan informasi di seluruh situs

Dalam hidrologi, rangkaian data di sejumlah lokasi yang terdiri dari nilai tahunan aliran sungai maksimum tahunan dalam satu tahun dianalisis. Model yang umum adalah distribusi nilai-nilai ini sama untuk semua situs selain faktor penskalaan sederhana, sehingga lokasi dan skala ditautkan dengan cara yang sederhana. Kemudian ada pertanyaan untuk memeriksa homogenitas di seluruh situs distribusi nilai yang diskalakan.

4. Menggabungkan sumber informasi

Dalam meteorologi, kumpulan data cuaca diperoleh dalam catatan selama bertahun-tahun dan, sebagai bagian dari ini, pengukuran di stasiun tertentu mungkin berhenti sesekali, sementara pada waktu yang sama, pengukuran dapat dimulai di lokasi terdekat. Kemudian

ada pertanyaan apakah, jika rekaman digabungkan untuk membentuk satu rangkaian rekaman yang lebih panjang, rekaman tersebut dapat dianggap homogen dari waktu ke waktu. Contoh pengujian homogenitas kecepatan angin dan arah data dapat ditemukan di Romanić et al., 2015.

##### 5. Homogenitas dalam populasi

Survei populasi sederhana dapat dimulai dari gagasan bahwa tanggapan akan homogen di seluruh populasi. Menilai homogenitas populasi akan melibatkan melihat apakah tanggapan subpopulasi teridentifikasi tertentu berbeda dari, yang lain. Misalnya, pemilik mobil mungkin berbeda dengan bukan pemilik mobil, atau mungkin ada perbedaan antara kelompok usia yang berbeda.

Dalam hal ini, homogenitas sangat diperlukan dalam pengolahan dan analisa data. Homogenitas menjadi asumsi varians penting agar estimasi gabungan dapat digunakan. Penyatuan varians dilakukan karena varians diasumsikan sama dan mengestimasi kuantitas yang sama (variens populasi) terlebih dahulu.

# **Bab 11**

## **Analisis Korelasi**

### **11.1 Pendahuluan**

Di dalam kehidupan sehari-hari setiap individu pasti berhubungan dengan orang lain. Sama seperti hubungan pada manusia, banyak kejadian di antara kita berhubungan dengan kejadian lain. Misalnya nilai yang didapat seseorang pada saat ujian pasti berhubungan dengan jumlah jam yang digunakan untuk belajar atau mempersiapkan diri untuk ujian. Kemahiran seorang mahasiswa komputer mengetik di keyboard komputer berhubungan dengan banyaknya waktu yang digunakan untuk berlatih mengetik.

Dalam statistik hubungan suatu kejadian dengan kejadian lain disebut korelasi. Sesuai arah hubungan, korelasi dibagi atas dua bagian yaitu korelasi positif dan negatif. Pada korelasi positif, pertambahan nilai suatu variabel akan diikuti dengan pertambahan nilai variabel lain yang berhubungan dengannya. Dengan kata lain kedua variabel berubah pada arah yang sama. Sebagai contoh, semakin rajin belajar maka semakin baik nilai seorang mahasiswa. Sebaliknya pada korelasi negatif kedua variabel bergerak dalam arah yang berbeda, di mana pertambahan nilai suatu variabel justru akan membuat pengurangan nilai pada variabel yang dikorelasikan dengannya. Misalnya, bila stok barang melimpah di pasar maka harga akan semakin turun.

Besaran statistik berhubungan dengan arah dan intensitas hubungan adalah koefisien korelasi. Jika nilainya positif maka korelasinya positif artinya kedua variabel X dan Y bergerak/berubah dengan arah yang sama. X dan Y sama-sama bertambah atau berkurang. Jika nilainya negatif maka korelasinya negatif, di mana kedua variabel X dan Y berubah dengan arah yang berbeda. X bertambah namun Y berkurang atau sebaliknya X berkurang tetapi Y justru bertambah. Hubungan antara X dan Y dalam analisis korelasi bukanlah hubungan causation atau hubungan sebab-akibat (cause-effect), di mana variabel X menyebabkan terjadinya Y atau sebaliknya, tetapi hubungan covariance (kovarianasi) di mana perubahan X dapat dihubungkan dengan perubahan Y.

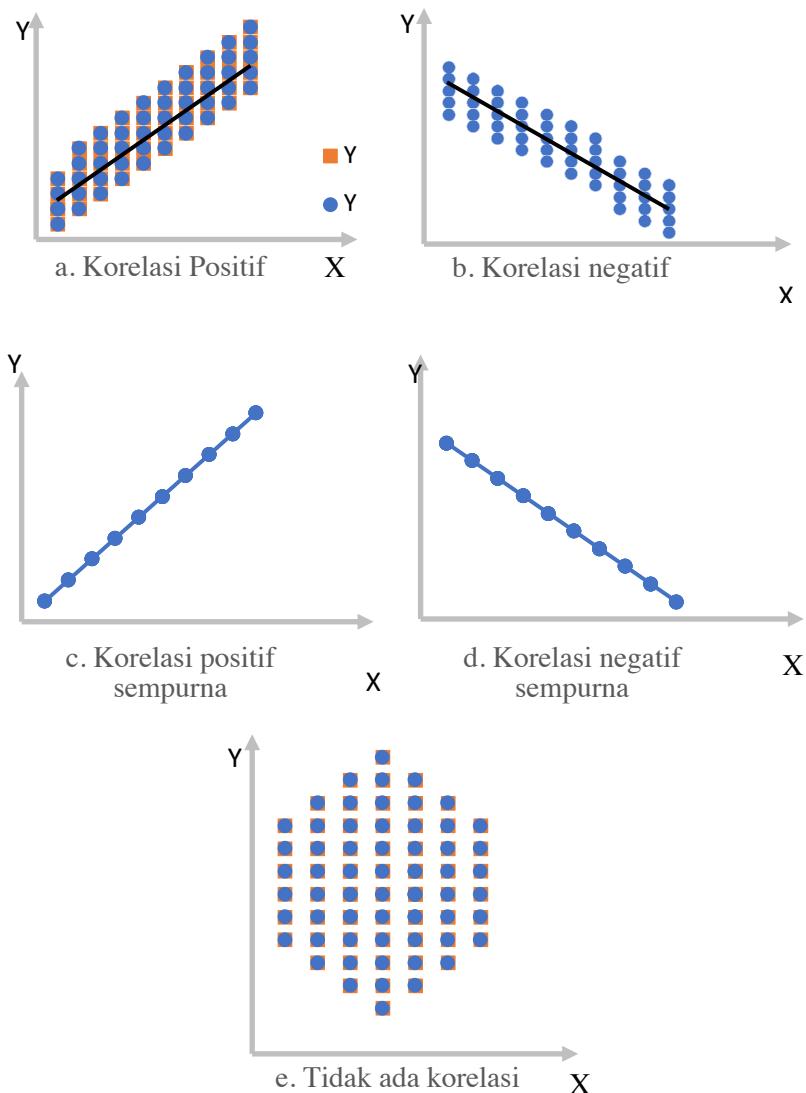
## 11.2 Menentukan Koefisien Korelasi

Untuk menentukan arah dan kekuatan korelasi maka perlu ditentukan koefisien korelasi. Ada banyak cara menghitung koefisien korelasi tergantung kepada tipe dan sajian data. Pada bab ini dibahas hanya beberapa teknik yang umum digunakan untuk mengukur korelasi antara variabel-variabel, antara lain:

1. Menggambar diagram pencar antara data-data X dan Y
2. Menghitung koefisien korelasi Pearson,
3. Menghitung koefisien korelasi data berkelompok,
4. Menghitung korelasi pangkat Spearman.

### 11.2.1 Diagram Pencar

Dengan melihat diagram pencar dapat diketahui arah korelasi antara variabel X dengan Y. Pada gambar 11.1 berikut ditunjukkan ada lima kemungkinan arah korelasi antara X dan Y: (a) berkorelasi positif, (b) berkorelasi negatif, (c) berkorelasi positif sempurna, (d) berkorelasi negatif sempurna atau (e) tidak ada korelasi. Berkorelasi positif berarti jika X bertambah maka Y cenderung bertambah, sebaliknya berkorelasi negatif berarti jika X bertambah maka Y berkurang. Berkorelasi positif sempurna atau negatif sempurna berarti semua titik data terletak pada garis lurus. Tidak ada korelasi berarti tidak dapat disimpulkan arah hubungan, dan derajat hubungannya adalah nol.



**Gambar 11.1:** Lima kemungkinan korelasi linier untuk data X dan Y (Burns, 1996)

### 11.2.2 Koefisien Korelasi Pearson

Dengan diagram pencar dapat ditentukan arah hubungan antara X dan Y, namun tidak dapat ditentukan derajat (kuat atau lemahnya) hubungan tersebut. Untuk memastikan apakah hubungan tersebut kuat atau lemah maka perlu dihitung koefisien korelasi. Bilamana data X dan Y adalah ordinal, maka dapat ditentukan koefisien korelasi menggunakan rumus Pearson. Koefisien korelasi Pearson dikenal juga sebagai korelasi produk momen.

Jika sejumlah data untuk variabel X dengan rerata  $\bar{X}$ , untuk variabel Y dengan rerata  $\bar{Y}$ ,  $(X - \bar{X})$  adalah deviasi data X terhadap rerata  $\bar{X}$ , dan  $(Y - \bar{Y})$  adalah deviasi Y terhadap rerata  $\bar{Y}$ , maka koefisien korelasi dihitung dengan rumus (Asuero, Sayago, & Gonzales, 2006):

$$r = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sqrt{(X-\bar{X})^2(Y-\bar{Y})^2}} \text{ atau} \quad (1)$$

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

Contoh 1.

Tentukan koefisien korelasi untuk data variabel X dan Y berikut

X	Y	$(X - \bar{X})$	$(Y - \bar{Y})$	$(X - \bar{X})(Y - \bar{Y})$	$(X - \bar{X})^2$	$(Y - \bar{Y})^2$
10	30	-2,8	-3,4	9,52	7,84	11,56
12	32	-0,8	-1,4	1,12	0,64	1,96
13	34	0,2	0,6	0,12	0,04	0,36
14	35	1,2	1,6	1,92	1,44	2,56
15	36	2,2	2,6	5,72	4,84	6,76
$\Sigma X = 64$	$\Sigma Y = 167$			$\Sigma = 18,4$	$\Sigma = 14,8$	$\Sigma = 23,2$
$\bar{X} = 12,8$	$\bar{Y} = 33,4$					

Dari tabel di atas didapat:

$$\Sigma X = 64;$$

$$\Sigma Y = 167;$$

$$\Sigma(X - \bar{X})(Y - \bar{Y}) = 18,4;$$

$$\Sigma(X - \bar{X})^2 = 14,8;$$

$$\Sigma(Y - \bar{Y})^2 = 23,2$$

Maka koefisien korelasi menurut rumus (1) adalah:

$$r = \frac{18,4}{\sqrt{14,8}\sqrt{23,2}} = 0,993$$

Hasilnya akan sama jika rumus (2) digunakan untuk menghitung r.

X	Y	XY	X <sup>2</sup>	Y <sup>2</sup>
10	30	300	100	900
12	32	384	144	1024
13	34	442	169	1156
14	35	490	196	1225
15	36	540	225	1296
$\Sigma X = 64$	$\Sigma Y = 167$	$\Sigma XY = 2156$	$\Sigma X^2 = 834$	$\Sigma Y^2 = 5601$

Jika koefisien korelasi dihitung menggunakan rumus (2) jawabannya akan sama. Dengan n=5;  $\Sigma X = 64$ ;  $\Sigma Y = 167$ ,  $\Sigma XY = 2156$ ;  $\Sigma X^2 = 834$ ;  $\Sigma Y^2 = 5601$ ; maka koefisien korelasi menurut rumus (2) adalah:

$$r = \frac{5(2156) - (64)(167)}{\sqrt{5(834) - (64)^2}\sqrt{5(5601) - (167)^2}} = 0,993$$

**Tabel 11.1:** Arti magnitudo koefisien korelasi (Burns, 1996)

Koefisien korelasi	Arti
0,90 - 1,00	Korelasi sangat tinggi (Hubungan sangat kuat)
0,70 - 0,90	Korelasi tinggi (Hubungan kuat)
0,40 - 0,70	Korelasi sedang (Hubungan sedang)
0,20 - 0,40	Korelasi rendah (Hubungan lemah)
< 0,20	Korelasi sangat rendah (hubungan sangat lemat, dapat diabaikan)

Koefisien korelasi menunjukkan arah dan kekuatan hubungan antara dua variabel. Arah dari hubungan ditunjukkan oleh tanda (+ atau -) dan kekuatan hubungan dinyatakan oleh besar absolut dari koefisien itu, yaitu seberapa dekat koefisien tersebut ke +1,0 atau -1,0. Pada tabel 11.1 diberikan besar koefisien korelasi dan artinya. Jika  $r = \pm 1$  berarti korelasi sempurna, dan jika  $r = 0$  berarti tidak ada korelasi.

### 11.2.3 Koefisien korelasi data berkelompok

Jika data X dan Y ada dalam bentuk distribusi frekuensi atau berkelompok maka koefisien korelasi dihitung dengan rumus (Supranto, 2000):

$$r = \frac{n(\sum uvf) - (\sum uf_u)(\sum vf_v)}{\sqrt{n(u^2 f_u) - (\sum uf_u)^2} \sqrt{n(v^2 f_v) - (\sum vf_v)^2}} \quad (3)$$

Di mana:

$n$  = jumlah sampel (data),

$u$  dan  $v$  = skala baru untuk interval data X dan Y (untuk menyederhanakan data, menggantikan nilai titik tengah interval data X dan Y), biasanya  $u$  atau  $v$  dibuat 0 untuk interval data paling tengah.

$f$  adalah frekuensi data untuk kelompok data tertentu.

Contoh 2.

Diberikan tabel frekuensi antara nilai Pemrograman dan Logika Matematika. Ada mengatakan bahwa nilai logika matematika berkorelasi positif terhadap nilai pemrograman. Tentukan koefisien korelasi untuk data pada tabel berikut

Logika	Pemrograman				
	1-20	21-40	41-60	61-80	81-100
1-20	2	4	3		
21-40		5	4	3	
41-60		2	6	8	3
61-80			2	4	2
81-100			1	3	5

Untuk menghitung koefisien korelasi r maka dibuat tabel korelasi sebagai berikut

	v	fv	vfv	v <sup>2</sup> fv	uvf
2	4	3		-2	9
	5	4	3	-1	12
	2	6	8	3	0
		2	4	2	1
		1	3	5	2
u	-2	-1	0	1	2
fu	2	11	16	18	10
ufu	-4	-11	0	18	20
u <sup>2</sup> f	8	11	0	18	40
u	8	13	0	7	24
uvf					52

Contoh perhitungan untuk  $\Sigma uvf$  pada kolom akhir:

Baris 1 ( $v=-2$ ):  $uvf = (-2)(-2)(2) + (-1)(-2)(4) + (0)(-2)(3) = 16$

Baris 2 ( $v=-1$ ):  $uvf = (-1)(-1)(5) + (0)(-1)(4) + (1)(-1)(3) = 2$

Baris 3 ( $v=0$ ):  $uvf = (-1)(0)(2) + (0)(0)(6) + (1)(0)(8) + (2)(0)(3) = 0$

Baris 4 ( $v=1$ ):  $uvf = (0)(1)(2) + (1)(1)(4) + (2)(1)(2) = 8$

Baris 5 ( $v=5$ ):  $uvf = (0)(2)(1) + (1)(2)(3) + (2)(2)(5) = 26$

Sehingga jumlah untuk kolom terakhir nilai  $\Sigma uvf = 16+2+0+8+26 = 52$ .

Hasil perhitungan akan sama untuk  $\Sigma uvf$  dari sisi baris terakhir:

Kolom 1 ( $u=-2$ ):  $uvf = (-2)(-2)(2) = 8$

Kolom 2 ( $u=-1$ ):  $uvf = (-1)(-2)(4) + (-1)(-1)(5) + (-1)(0)(2) = 13$

Kolom 3 ( $u=0$ ):  $uvf = (0)(-2)(3) + (0)(-1)(4) + (0)(0)(6) + (0)(1)(2) + (0)(2)(1) = 0$

Kolom 4 ( $u=1$ ):  $uvf = (1)(-1)(3) + (1)(0)(8) + (1)(1)(4) + (1)(2)(3) = 7$

Kolom 5 ( $u=2$ ):  $uvf = (2)(0)(3) + (2)(1)(2) + (2)(2)(5) = 24$

Sehingga jumlah untuk baris terakhir atau  $\Sigma uvf = 8+13+0+7+24 = 52$ .

Sehingga didapat  $n=43$ ,  $\Sigma ufu=23$ ,  $\Sigma vfv=-4$ ,  $\Sigma u^2 fu=77$ ,  $\Sigma v^2 fv=92$  dan  $\Sigma uvf=52$ , maka koefisien korelasi dihitung menggunakan rumus (3)

$$r = \frac{43(52) - (23)(-4)}{\sqrt{43(77) - (23)^2} \sqrt{43(92) - (-4)^2}}$$

$$r = 0,703$$

#### 11.2.4 Koefisien Korelasi Peringkat (Rank) Spearman

Koefisien korelasi yang dihitung dengan rumus (1) atau (2) digunakan jika data adalah ordinal. Jika data berupa peringkat (rank) maka koefisien korelasi dihitung menggunakan rumus Spearman sebagai berikut (Mukaka, 2012):

$$r = 1 - \frac{6 \sum d}{n(n^2 - 1)} \quad (4)$$

Di mana  $n$  adalah banyaknya pasangan peringkat,  $d$  adalah selisih dari pasangan peringkat.

Contoh 3.

Ada sebuah hipotesa yang mengatakan semakin cepat menyelesaikan suatu soal ujian komputer akan semakin bagus skor ujiannya. Diberikan data urutan 10 tercepat mahasiswa yang menyelesaikan ujian komputer dan skor ujiannya yang diberikan pada tabel berikut

Urutan selesai ujian (a)	Skor ujian	Ranking skor ujian (b)	$d$ (a-b)	$d^2$
1	90	1	0	0
2	85	3	1	1
3	88	2	1	1
4	78	5	1	1
5	80	4	1	1
6	76	6	0	0
7	74	7	0	0
8	72	8	0	0
9	68	10	1	1
10	70	9	1	1
				$\Sigma d^2 = 41,5$

Maka koefisien korelasi Spearman adalah

$$r = 1 - \frac{6(6)}{10(10^2-1)} = 1 - \frac{36}{990} = 0,964$$

Contoh 4.

Ada hipotesis menyatakan bahwa semakin lama waktu mahasiswa untuk melakukan praktek akan semakin mahir. Pada tabel berikut diberikan data jumlah jam yang dilakukan latihan mengetik dan kemahiran mengetik dari 12 mahasiswa IT. Hitunglah koefisien korelasi peringkat dari data tersebut.

Jumlah jam praktek	Skor kemahiran	Peringkat jam	Peringkat kemahiran	d	d <sup>2</sup>
24	75	1	4	3	9
23	83	2	2	0	0
19,5	98	3	1	2	4
18	80	4	3	1	1
17	74	5	5	0	0
16,5	69	6	8,5	2,5	6,25
16	71	7	6	1	1
15	69	8	8,5	0,5	0,25
14,5	59	9	11	2	4
14	62	10	10	0	0
13,5	70	11	7	4	16
10	54	12	12	0	0
					$\Sigma d^2 = 41,5$

Maka koefisien korelasi Spearman adalah

$$r = 1 - \frac{6(41,5)}{12(12^2 - 1)} = 1 - \frac{249}{1716} = 0,855$$

## 11.3 Koefisien determinasi

Koefisien determinasi dalam satu segi lebih berguna dari koefisien korelasi dalam hal di mana itu dapat memberikan penjelasan statistik yang lebih masuk akal perihal hubungan antara variabel X dan Y. Koefisien determinasi disimbolkan oleh  $r^2$ , dengan kata lain itu adalah kuadrat dari koefisien korelasi ( $r$ ). Koefisien korelasi menentukan intensitas hubungan antara variabel X dan Y, namun koefisien determinasi memberikan variabilitas dalam Y dapat dijelaskan oleh variabilitas dalam X. Dengan kata lain nilai koefisien determinasi adalah besar (dalam persentase) variabel X dapat menjelaskan perubahan atau varians dalam variabel Y.

Pada tabel 11.2 diberikan perbandingan antara koefisien korelasi ( $r$ ) dan determinasi ( $r^2$ ). Saat koefisien korelasi = 0,7 (korelasi tinggi) namun koefisien determinasi = 0,49. Ini berarti bahwa sekalipun korelasinya tinggi, namun hanya 49% variabilitas dalam variabel Y dapat dijelaskan oleh variabilitas X, selebihnya ditentukan variabel atau faktor lain.

**Tabel 11.2:** Hubungan besar koefisien korelasi dan determinasi (Holcomb, 1998).

Koefisien korelasi Pearson ( $r$ )	Koefisien Determinasi ( $r^2$ )	Persentase
0,00	0,00	0%
0,10	0,01	1%
0,20	0,04	4%
0,30	0,09	9%
0,40	0,16	16%
0,50	0,25	25%
0,60	0,36	36%
0,70	0,49	49%
0,80	0,64	64%

0,90	0,81	81%
1,00	1,00	100%

Koefisien korelasi maupun determinasi dapat dengan cepat ditentukan menggunakan program computer Excell maupun SPSS atau program pengolahan data statistik lainnya.

Dengan Excell: Klik Data, kemudian Data Analisis, kemudian klik Correlation, maka akan didapat jawaban yang sama dengan hasil hitung di atas.

Dengan SPSS: Klik Analyze, kemudian Correlate, Bivariate, kemudian pilih Pearson, hasilnya akan sama dengan hasil hitung dengan excel atau perhitungan manual.

## 11.4 Uji statistik uji Korelasi

Di atas telah dijelaskan cara menentukan koefisien korelasi dan apa arti dari hasil perhitungan tersebut. Selain menghitung koefisien korelasi perlu juga dilakukan uji terhadap koefisien korelasi tersebut. Uji statistik untuk koefisien korelasi dilakukan dengan menghitung nilai t sebagai berikut (Sudjana, 2001):

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

(5)

Di mana n adalah jumlah sampel dan r adalah koefisien korelasi.

Hipotesis penelitian adalah

$H_0: \rho = 0$ , tidak ada hubungan yang signifikan antara variabel X dan Y.

$H_a: \rho > 0$ , ada hubungan yang positif antara variabel X dan Y, atau

$H_a: \rho < 0$ , ada hubungan yang negatif antara variabel X dan Y, atau

$H_a: \rho \neq 0$ , ada hubungan antara variabel X dan Y.

Untuk menguji hipotesis maka nilai thitung dibandingkan dengan nilai ttabel.

Untuk uji dua pihak dan nilai  $\alpha$  yang diambil maka ttabel.

H<sub>0</sub> ditolak jika:

$$-\alpha(n-2) < t_{hitung} < \alpha(n-2) \{uji dua pihak\}$$

$$t_{hitung} > t_{tabel} = \alpha/2(n-2) \text{ atau } -t_{hitung} < t_{tabel} = \alpha/2(n-2) \{uji satu pihak\}$$

Jika t-tabel tidak diketahui maka pengujian hipotesis dapat dilakukan dengan membandingkan nilai signifikansi (sig.), yang diperoleh dari pengolahan statistik menggunakan Excell, SPSS atau program komputer yang lain, dengan nilai  $\alpha$  (uji dua pihak) atau  $\alpha/2$  (uji satu pihak). Kriteria pengujian adalah H<sub>0</sub> ditolak jika:

$$\text{Sig.} < \alpha \text{ (uji dua pihak) atau}$$

$$\text{Sig.} < \alpha/2 \text{ (uji satu pihak).}$$

Contoh 5.

Ujilah hipotesis yang mengatakan bahwa tidak hubungan yang positif antara skor mata kuliah Kalkulus dengan Logika matematika untuk data skor nilai 12 mahasiswa IT berikut:

Kalkulus	88	90	78	65	80	45	70	96	52	60	78	84
Logika	68	84	72	60	78	70	68	80	76	65	82	80

Dengan menggunakan program komputer SPSS diperoleh hasil pengolahan Correlation baik untuk uji dua pihak (2-tailed) maupun uji satu pihak (1-tailed) sebagai berikut:

Correlations

		Skor Kalkulus	Skor Logika
Skor Kalkulus	Pearson Correlation	1	0,509
	Sig. (2-tailed)		0,091
	N	12	12
Skor Logika	Pearson Correlation	0,509	1
	Sig. (2-tailed)	0,091	
	N	12	12

## Correlations

		Skor Kalkulus	Skor Logika
Skor Kalkulus	Pearson Correlation	1	0,509*
	Sig. (1-tailed)		0,045
	N	12	12
Skor Logika	Pearson Correlation	0,509*	1
	Sig. (1-tailed)	0,045	
	N	12	12

\*. Correlation is significant at the 0,05 level (1-tailed).

Dari tabel di atas dilihat bahwa koefisien korelasi adalah 0,509, dan nilai signifikansi untuk uji dua pihak (2-tailed) adalah 0,091 dan untuk uji satu pihak (1-tailed) adalah 0,045. Untuk uji dua pihak, nilai  $\alpha=0,05$  lebih kecil dari sig., demikian halnya untuk uji satu pihak, nilai  $\alpha/2=0,025$  lebih kecil dari 0,045, maka  $H_0$  diterima, ini berarti bahwa tidak korelasi yang signifikan antara Skor Logika dan Kalkulus.

# **Bab 12**

## **Analisis Regresi**

### **12.1 Pendahuluan**

Sejarah regresi dimulai pertama dari tahun 1877 dari ide yang dikembangkan oleh Sir Francis Galton. Ide regresi berawal dari studinya tentang hubungan antara tinggi orangtua dan anak. Orangtua yang tinggi cenderung mempunyai anak yang tinggi pula sedangkan orang tua pendek cenderung mempunyai anak yang pendek. Studi Galton ini menjelaskan bahwa tinggi anak sebagai variabel dependen dipengaruhi oleh tinggi orangtua sebagai variabel independen. Dari ide Galton ini kemudian regresi dikembangkan untuk menjelaskan hubungan variabel dari dalam berbagai disiplin ilmu (Agus Widarjono, 2015)

Keterkaitan banyak bidang ilmu selalu berhubungan dengan peramalan atau meramalkan suatu peristiwa dan hal-hal yang terjadi. Metode peramalan biasanya dikaitkan dengan ilmu matematika. Persamaan matematik yang digunakan untuk meramalkan hubungan dari nilai-nilai suatu variabel disebut persamaan regresi (Ronald, E. Walpole, 1995). Persamaan regresi merupakan persamaan matematis yang menginterpretasikan hubungan pengaruh (casual relationship) antara dua variabel yaitu varibel yang memengaruhi (variabel independen) dan variabel dipengaruhi (variabel dependen) (Algifari, 2016). Pembahasan analisis regresi pada bab ini meliputi Analisis regresi sederhana

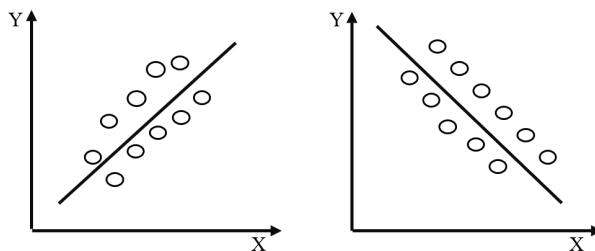
yaitu model persamaan garis lurus dan Analisis Regresi Berganda yaitu perpanjangan dari analisis regresi sederhana.

## 12.2 Jenis Hubungan Antar Variabel

Variabel dependen adalah variabel yang diestimasi atau diprediksi berdasarkan nilai variabel lain yaitu variabel independen. Variabel independen adalah variabel yang digunakan untuk memprediksi/ mengestimasi varibel dependen (Agus Widarjono, 2015).

Hubungan pengaruh antara dua variabel yang digambarkan pada suatu persamaan garis lurus menunjukkan bahwa nilai variabel dependen berubah dalam besaran yang tetap setiap terjadi perubahan suatu unit nilai variabel independen (Algifari , 2016). Jika variabel dependen dan variabel independen bergerak dalam arah yang sama (direct relationship) maka hubungan antara kedua varabel adalah hubungan positif sedangkan jika kedua variabel bergerak dalam arah yang berlawanan (Inverse Relationship) maka hubungan antara kedua variabel adalah hubungan negatif. (Agus Widarjno, 2015)

Gambaran hubungan positif dan negatif antara variabel dalam dilihat pada diagram berikut ini:



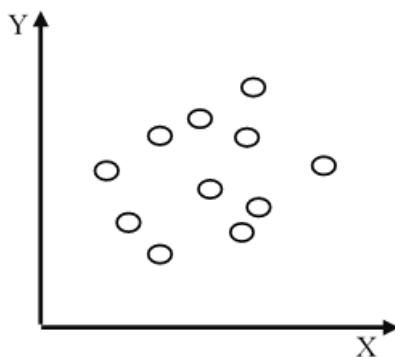
**Gambar 12.1:** Hubungan positif dan negatif Antar Variabel dalam Regresi

1. Hubungan positif (kiri): Pertambahan nilai sumbu X (garis horizontal) umumnya diikuti dengan pertambahan nilai sumbu Y (garis vertikal)

2. Hubungan negatif (kanan): Pertambahan nilai sumbu X (garis horizontal) umumnya diikuti dengan pengurangan nilai sumbu Y (garis vertikal)

Persamaan garis yang terlihat menunjukkan hubungan positif dan negatif dari suatu persamaan garis regresi. Garis regresi adalah garis lurus di antara titik-titik pada diagram tebar yang menggambarkan hubungan linier antara variabel pada diagram tebar tersebut. Garis ini sering juga disebut dengan garis kesesuaian terbaik (line of best-fit) di mana merupakan garis terdekat dengan titik tebar pada diagram (Johan Harlan, 2018).

Sementara itu, jika titik-titik pada diagram berpencar dengan tidak mengikuti atau mendekati garis regresi, maka tidak ada hubungan yang terjadi antara variabel dependen dengan variabel independent yang ditunjukkan oleh diagram berikut ini:



**Gambar 12.2:** Diagram tidak ada hubungan antara Variabel dalam Regresi

## 12.3 Analisis Regresi Sederhana

### 12.3.1 Pengertian Regresi Linier Sederhana

Regresi linear adalah teknik analisis data untuk menginterpretasikan hubungan antara satu variabel dependen dengan satu atau lebih variabel independen. Regresi linier sederhana (Simple Linier Regression) adalah analisis yang

menginterpretasikan hubungan antara satu variabel dependen dengan satu variabel independen (Johan Harlan, 2018).

Regresi Linier Sederhana digambarkan dalam persamaan matematis yang menggambarkan hubungan linier antara dua variabel di mana salah satu variabel diprediksi memberikan pengaruh terhadap variabel yang lain (Suyono, 2018).

Analisis regresi memiliki tiga kegunaan, yaitu:

1. Menggambarkan kondisi data atau kasus yang sedang diteliti, melalui analisis regresi kita dapat menjelaskan kondisi data melalui terbentuknya suatu model hubungan yang bersifat numerik.
2. Digunakan dalam pengendalian (kontrol) pada suatu kasus atau hal-hal yang sedang diamati dengan menggunakan model regresi yang diperoleh.
3. Model regresi yang diperoleh dari hasil analisis regresi juga dapat dimanfaatkan untuk memprediksi/meramalkan perilaku variabel dependen (Agus Tri Basuki, 2015).

### 12.3.2 Model Analisis Regresi Linier Sederhana

Model Regresi Linier Sederhana diinterpretasikan oleh persamaan garis lurus yang dihasilkan dari hasil penggunaan Metode Kuadrat Terkecil (Least Square Method) sebagai berikut:

$$Y' = a + bX$$

Di mana:

- $a$  = konstanta
- $b$  = kemiringan dari garis regresi atau koefisien regresi
- $X$  = nilai tertentu dari variabel bebas
- $Y'$  = nilai yang diukur/dihitung pada variabel tidak bebas

Persamaan regresi digambarkan pada diagram pencar yang mana akan menghasilkan garis linier yang disebut garis regresi. Garis regresi terletak sedemikian rupa agar menggambarkan penyimpangan positif dan penyimpangan negatif pada titik-titik pencar pada garis seimbang sehingga

penyimpangan pada garis tersebut adalah nol yang memenuhi sifat matematis berikut:

$$\sum (Y - Y') = 0 \text{ dan}$$

$$\sum (Y - Y')^2 = \text{nilai terkecil atau terendah}$$

Agar diperoleh penyimpangan yang seimbang dalam menentukan persamaan regresi digunakan rumus berikut ini:

$$b = \frac{\sum x_i y_i}{\sum x_i^2}$$

Atau

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

$$a = \underline{Y} - b \underline{X}$$

(J.Supranto, 2000)

### 12.3.3 Menghitung Regresi Linier dengan Aplikasi

Regresi linier dapat dihitung dengan menggunakan aplikasi Komputer. Pada pembahasan ini akan dijelaskan tentang perhitungan menggunakan Aplikasi SPSS.

Contoh Kasus:

Sebuah perusahaan produsen batik ingin mengetahui hubungan antara biaya produksi dengan jumlah produksi. Dengan menggunakan data pada tabel 12.1 berikut ini:

**Tabel 12.1:** Data Biaya dan Jumlah Produksi

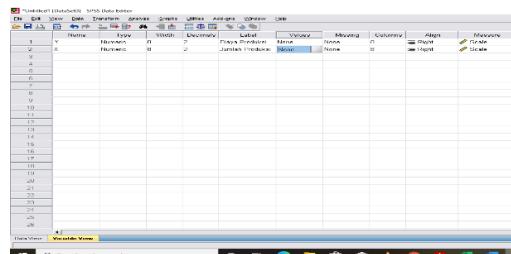
Biaya Produksi (Y)	Jumlah Produksi (X)
64	20
61	16
84	34

70	23
88	27
92	32
72	18
77	22

Tentukanlah persamaan regresi dari hubungan biaya produksi dengan jumlah produksi di mana biaya produksi (Y) tergantung dari jumlah produksi (X). (Nuryadi,dkk, 2017)

## **Langkah Pertama:**

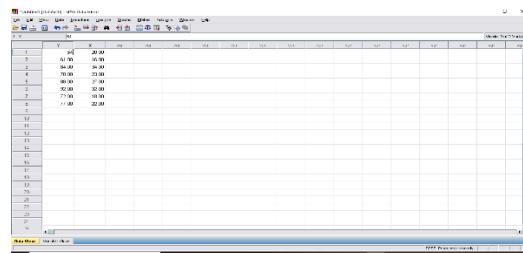
Buka aplikasi SPSS pada komputer dan lihat bagian bagian sudut kiri bawah pilih “Variable View” dan isi seperti gambar di bawah:



**Gambar 12.3:** Variabel View

### **Langkah Kedua:**

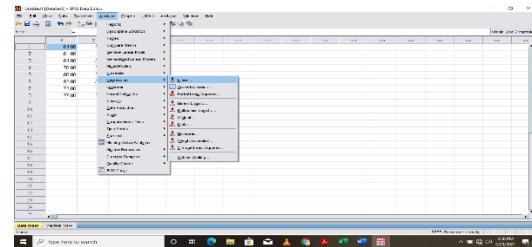
Selanjutnya pada bagian sudut kiri bawah pilih “Data View” dan input data pada kolom yang sesuai seperti gambar dibawah ini:



**Gambar 12.4:** Data View

### Langkah Ketiga:

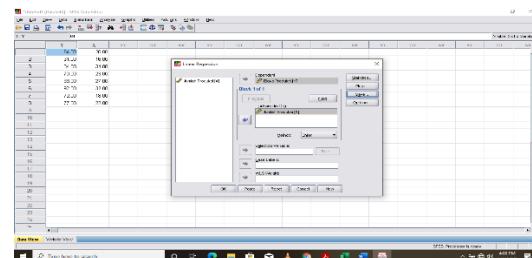
Lalu lihat bagian atas dan pilih menu Analyze-Regression-Linear... akan muncul seperti pada gambar dibawah ini:



Gambar 12.5: Menu Analyze-Regression-Linear

### Langkah Keempat:

Kemudian akan muncul kotak dialog “Linear Regressions”, input “Biaya Produksi” ke dalam kotak “dependent” dan “Jumlah Produksi” ke dalam kotak “Independent” seperti pada gambar dibawah ini:

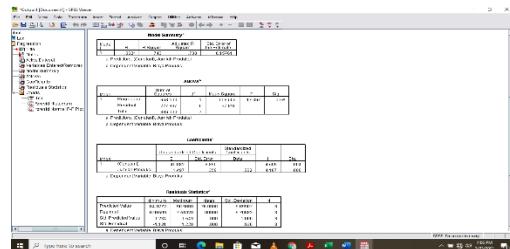


Gambar 12.6: Kotak Dialog Linear Regressions

Klik bagian “Plot” dan centang bagian “Histogram” dan “Normal Probability Plot” lalu klik continue, kemudian klik bagian “Save” dan centang “Unstandardized” pada kolom Residual lalu klik continue, kemudian klik OK.

### Langkah Kelima:

Akan muncul halaman output hasil analisis seperti pada gambar berikut ini:



**Gambar 12.7:** Output Analisis Regresi

Persamaan regresi ditunjukkan oleh tabel Coefficients nilai Constant = 40.082 dan Jumlah Produksi = 1.497 sehingga Persamaan Regresi dapat dituliskan menjadi

$$Y = 40.082 + 1.497 X$$

Dari persamaan regresi dapat disimpulkan bahwa akan terjadi kenaikan Biaya Produksi sebesar 40.082 dengan kenaikan jumlah produksi 1.497.

## 12.4 Analisis Regresi Linier Berganda

### 12.4.1 Pengertian Regresi Linier Berganda

Regresi berganda adalah perpanjangan dari regresi linier sederhana. Analisis Regresi Linier Sederhana menghasilkan persamaan yang bersifat linier yaitu jika variabel X mengalami perubahan maka variabel Y akan mengikuti secara permanen. Sementara itu, Analisis Regresi Linier Berganda digunakan saat kita ingin memprediksi hubungan antara satu variabel independen berdasarkan nilai dari dua atau lebih variabel dependen. Untuk itu, penyelesaian model regresi berganda dapat dilakukan secara sistematis melalui proses penyelesaian dengan aturan matriks. Oleh karena itu, jika Regresi Linier sederhana menggunakan Metode Kuadrat Terkecil dalam penyelesaian maka analisis regresi berganda lebih mudah diselesaikan dengan menggunakan metode matriks.

### 12.4.2 Model Regresi Linier Berganda

Analisis regresi linier berganda juga menghasilkan persamaan matematis. Jika ada lebih dari dua variabel maka hubungan linier dapat dinyatakan dalam persamaan regresi linier berganda sebagai berikut:

$$Y' = b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n$$

$$Y'_i = b_0 + b_1X_{1i} + b_2X_{2i} + \dots + b_nX_{ni}$$

Di mana:

$Y$  = nilai-nilai hasil pengamatan

$Y'$  = nilai regresi

$I = 1, 2, 3, \dots, k$ .

Pada persamaan di atas ada satu variabel dependen, yaitu  $Y'$  dan ada  $n$  variabel independen yaitu  $X_1, X_2, \dots, X_n$ .

Untuk menentukan nilai  $b_0, b_1, \dots, b_n$  dapat menggunakan persamaan normal berikut yang diperoleh dari penggunaan metode kuadrat terkecil (Least Square Method):

$$b_0n + b_1 \sum X_1 + b_2 \sum X_2 + \dots + b_n \sum X_n = \sum Y$$

$$\begin{aligned} b_0 \sum X_1 + b_1 \sum X_1^2 + b_2 \sum X_1X_2 + \dots + b_n \sum X_1X_n \\ = \sum X_1Y \end{aligned}$$

$$\begin{aligned} b_0 \sum X_2 + b_1 \sum X_2X_1 + b_2 \sum X_2^2 + \dots + b_n \sum X_2X_n \\ = \sum X_2Y \\ \vdots \vdots \vdots \vdots \end{aligned}$$

$$\begin{aligned}
 b_0 \sum X_n + b_1 \sum X_n X_1 + b_2 \sum X_n X_2 + \dots + b_n \sum X_n^2 \\
 = \sum X_n Y
 \end{aligned}$$

Jika persamaan-persamaan di atas diselesaikan akan kita peroleh nilai  $b_0, b_1, \dots, b_n$  yang dapat digunakan untuk membentuk persamaan regresi. Untuk meramalkan nilai Y maka nilai X1, X2, ..., Xn sebagai variabel independen sudah diketahui.

Sebagai contoh, untuk n = 2,  $Y' = b_0 + b_1 X_1 + b_2 X_2$ , satu variabel dependen (Y), dan dua variabel independen (X1 dan X2) maka  $b_0, b_1, \dots, b_n$  dihitung dari persamaan normal berikut:

$$b_0 n + b_1 \sum X_1 + b_2 \sum X_2 = \sum Y$$

$$b_0 \sum X_1 + b_1 \sum X_1^2 + b_2 \sum X_1 X_2 = \sum X_1 Y$$

$$b_0 \sum X_2 + b_1 \sum X_2 X_1 + b_2 \sum X_2^2 = \sum X_2 Y$$

Ketiga persamaan di atas dapat diubah kedalam bentuk matriks berikut:

$$\underbrace{\begin{bmatrix} n & \sum X_1 & \sum X_2 \\ \sum X_1 & \sum X_1^2 & \sum X_1 X_2 \\ \sum X_2 & \sum X_2 X_1 & \sum X_2^2 \end{bmatrix}}_A \underbrace{\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}}_b = \underbrace{\begin{bmatrix} \sum Y \\ \sum X_1 Y \\ \sum X_2 Y \end{bmatrix}}_H$$

Dengan:

$A = \text{matriks (diketahui)}$

$b = \text{vector kolom (diketahui)}$

$H = \text{vector kolom (diketahui)}$

Variabel b dapat diketahui dengan cara sebagai berikut:

$$Ab = H$$

$$b = A^{-1}H$$

di mana  $A^{-1}$  adalah kebalikan (invers) dari matriks A. (J. Supranto, 2000)

### 12.4.3 Menghitung Regresi Linier dengan Aplikasi

Regresi linier berganda dapat dihitung dengan menggunakan aplikasi Komputer. Pada pembahasan ini akan dijelaskan tentang perhitungan menggunakan Aplikasi SPSS.

Contoh Kasus:

Seorang dosen Statistika ingin mengetahui pengaruh frekuensi belajar dan nilai Kalkulus secara bersama-sama terhadap prestasi Statistika mahasiswa berdasarkan data di bawah ini.

**Tabel 12.2:** Daftar Nilai Statistika Mahasiswa

ID Mahasiswa	X1	X2	Y
1	15	70	7
2	10	50	7
3	10	60	7
4	5	60	4
5	8	50	6

6	7	80	6
7	5	50	5
8	5	50	5
9	5	50	5
10	5	70	6
11	20	80	9
12	3	55	4
13	12	70	8
14	3	45	4
15	7	60	7

Keterangan:

X1 = Frekuensi Belajar (0-20) > dalam jam/minggu

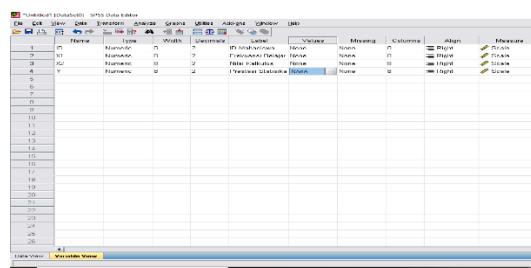
X2 = Nilai Kalkulus (0-100)

Y = Prestasi Statistika (0-10)

(Nuryadi,dkk, 2017)

### Langkah Pertama:

Buka aplikasi SPSS pada komputer dan lihat bagian bagian sudut kiri bawah pilih “Variable View” dan isi seperti gambar di bawah:



Gambar 12.8: Variabel View

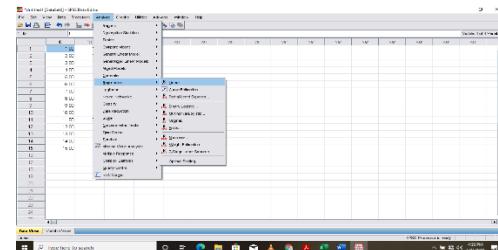
**Langkah Kedua:**

Selanjutnya pada bagian sudut kiri bawah pilih “Data View” dan masukkan data pada kolom yang sesuai seperti gambar dibawah ini:

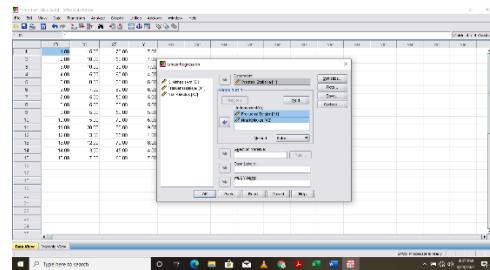
ID	X1	X2	X3	X4	X5	X6	X7	X8	X9
1	100	85	75	80	80	70	70	70	70
2	120	85	75	80	80	70	70	70	70
3	700	85	75	80	80	70	70	70	70
4	100	85	75	80	80	70	70	70	70
5	100	85	75	80	80	70	70	70	70
6	100	85	75	80	80	70	70	70	70
7	1000	85	75	80	80	70	70	70	70
8	1000	85	75	80	80	70	70	70	70
9	1000	85	75	80	80	70	70	70	70
10	1000	85	75	80	80	70	70	70	70
11	1000	85	75	80	80	70	70	70	70
12	100	85	75	80	80	70	70	70	70
13	100	85	75	80	80	70	70	70	70
14	100	85	75	80	80	70	70	70	70
15	100	85	75	80	80	70	70	70	70
16	100	85	75	80	80	70	70	70	70
17	100	85	75	80	80	70	70	70	70
18	100	85	75	80	80	70	70	70	70

**Gambar 12.9: Data View****Langkah Ketiga:**

Lalu pilih menu Analyze-Regression-Linear... seperti pada gambar dibawah ini:

**Gambar 12.10: Menu Analyze-Regression-Linear****Langkah Keempat:**

Kemudian akan muncul kotak dialog “Linear Regressions” dan masukkan “Prestasi Statistik” ke dalam kotak “dependent” dan “Frekuensi Belajar” dan “Nilai Kalkulus” ke dalam kotak “Independent” seperti pada gambar dibawah ini:

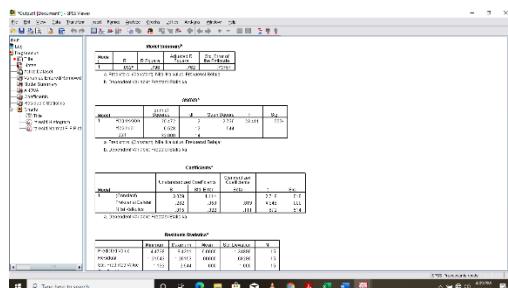


**Gambar 12.11:** Kotak Dialog Linear Regressions

Klik bagian “Plot” dan centang bagian “Histogram” dan “Normal Probability Plot”, kemudian klik bagian “Save” dan centang “Unstandardized” pada kolom Residual, lalu klik OK.

#### Langkah Kelima:

Selanjutnya akan muncul halaman output hasil analisis seperti pada gambar berikut ini:



**Gambar 12.12:** Output Analisis Regresi Linier Berganda

Persamaan regresi ditunjukkan oleh tabel Coefficients nilai Constant = 3.029 , Nilai Frekuensi Belajar= 0.262 dan nilai Kalkulus = 0.015 sehingga Persamaan Regresi dapat dituliskan menjadi

$$Y = 3.029 + 0.262 X_1 + 0.015 X_2$$

Dari persamaan regresi dapat disimpulkan bahwa akan terjadi kenaikan efektivitas belajar mahasiswa sebesar 3.029 dengan kenaikan kenaikan frekuensi belajar sebesar 0.262 dan nilai kalkulus sebesar 0.015.

# **Bab 13**

# **Aplikasi Statistika**

## **13.1 Penerapan Statistik**

Pemahaman akan pentingnya statistik sangatlah berguna bagi kita dalam hidup ini. Dengan mempeljari hasil output dari aplikasi, kita bisa melihat trend naik dan turunnya data, mengamati pola dan kebiasaan object, yang dapat digunakan untuk membuat perencanaan yang lebih matang.

Ilmu statistika banyak diaplikasikan ke dalam berbagai bidang yaitu, antara lain ilmu sosial, bidang bisnis, bidang ekonomi, dan bidang industri. Dalam bidang pemerintahan statistika digunakan untuk sensus penduduk, voting atau polling (contoh pemilihan dan penghitungan suara dalam pemilihan umum atau quick count), pada bidang komputasi statistika diterapkan pada pola kecerdasan buatan (Ulty, 2020).

Berikut ini adalah kegunaan statistik:

- Statistika digunakan sebagai media penyusun, meringkas atau penyempurnaan data. Data dapat diperoleh dari sebuah penelitian dengan menggunakan topik tertentu yang tidak mencakup dalam jumlah besar hanya saja berisi banyak informasi.
- Statistika untuk membantu dalam penelitian membangun aktivitas atau survey eksperimen yang bisa meminimalkan biaya untuk informasi

dalam jumlah yang ditentukan.

- Statistika ini digunakan sebagai penetapan metode yang baik untuk penarikan kesimpulan atau ringkasan sesuai dengan sampel tertentu.
- Terakhir, statistika digunakan untuk mengukur baik atau tidaknya sebuah kesimpulan atau evaluasi.

Statistik juga sangat dibutuhkan dalam penelitian karena memiliki peranan (Komunitas, 2018).

- Statistik memungkinkan pencatatan secara lengkap dari data penyelidikan
- Statistik memampukan seorang peneliti untuk bekerja secara berurutan dari awal sampai akhir
- Statistik menyediakan cara-cara meringkas data ke dalam bentuk yang lebih gampang mengerjakannya
- Statistik memberikan dasar-dasar melalui proses-proses yang mengikuti aturan yang dapat diterima oleh ilmu pengetahuan
- Statistik memberikan landasan untuk meramalkan secara ilmiah tentang bagaimana sesuatu gejala akan terjadi dalam kondisi-kondisi yang telah diketahui
- Statistik memungkinkan peneliti menganalisa, menguraikan sebab akibat yang kompleks dan rumit yang tanpa statistic akan merupakan persoalan yang membingungkan serta kejadian yang tak teruraikan
- Menggambarkan data dalam bentuk tak tentu
- Menyederhanakan data yang komplek menjadi data yang mudah dimengerti
- Merupakan teknik untuk membuat perbandingan
- Memperluas pengalaman individu
- Menentukan tingkat hubungan atau peranan antar variable
- Mengukur besar besaran variable
- Dasar untuk merumuskan kebijakan perusahaan atau pemerintahan

## 13.2 Aplikasi Statistik

Aplikasi statistik adalah program komputer khusus yang membantu kita dalam mengumpulkan, mengatur, menganalisis, menafsirkan, dan merancang data secara statistik. Ada dua teknik statistik yang sering digunakan dalam menganalisis data, yaitu: statistik deskriptif dan statistik inferensial (Teknozone.ID, 2020). Dalam memudahkan kita mengolah data dengan analisa statistika, kita dapat menggunakan berbagai macam alat bantu atau software. Adapun software statistika yang dapat digunakan sangatlah banyak namun tidak semuanya memiliki keakuratan yang baik. Ada beberapa software statistika yang sering digunakan baik dalam dunia pendidikan atau pun pekerjaan.

## 13.3 Jenis-Jenis Aplikasi Statistik

Kita melakukan penelitian untuk menguji hipotesis, dan kita melakukannya dengan cara mendapatkan data terlebih dahulu. Dengan harapan, jika eksperimen yang kita lakukan telah direncanakan dan dieksekusi dengan benar, maka kita akan mendapatkan hasil yang bagus yang dapat menunjukkan sesuatu yang unik di dunia ini. Terdapat berbagai macam alat bantu atau software dalam memudahkan kita untuk mengolah data menggunakan analisa statistika. Cukup banyak software statistika yang dapat kita gunakan namun tidak semua memiliki keakuratan yang baik.

Berikut ini merupakan software statistika yang sering digunakan dalam dunia pendidikan maupun dalam dunia pekerjaan :

### 1. Statistical Package for Social Science (SPSS)

SPSS merupakan singkatan dari Statistical Package for Social Science. SPSS sendiri pertama kali dibuat pada tahun 1968. Dan pada tahun 2009 tepatnya pada tanggal 28 Juli, SPSS dibeli oleh IBM seharga \$ 1,2 milyar. Setelah dibeli IBM, SPSS berkembang cukup pesat dan tidak hanya untuk statistik dasar, namun sampai proses analytic juga.



**Gambar 13.1:** SPSS

## 2. MINITAB

Minitab merupakan salah satu program komputer yang banyak digunakan untuk mempermudah pengolahan data statistik. Keunggulan dari program tersebut adalah dapat mengolah data statistika untuk tujuan sosial dan teknik. Hasil analisis di program MINITAB dapat ditampilkan dalam histogram, plot, dan angka. Dengan hanya memberikan satu atau dua perintah, bahkan dapat digabungkan dengan program pengolah data lain seperti Microsoft Office.



**Gambar 13.2:** Logo Minitab

## 3. Microsoft Excel

Meskipun bukan solusi mutakhir untuk analisis statistik, MS Excel memang menawarkan berbagai alat untuk visualisasi data dan statistik sederhana. Sangat mudah untuk menghasilkan ringkasan dan gambar dan angka yang dapat

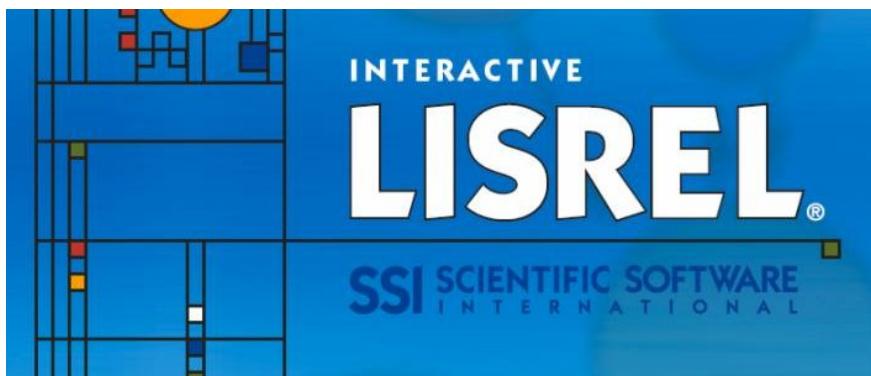
disesuaikan, menjadikannya alat yang dapat digunakan bagi banyak orang yang ingin melihat interpretasi sederhana dari data. Karena banyak individu dan perusahaan memiliki dan mengetahui cara menggunakan Excel, itu juga menjadikannya pilihan yang dapat diakses bagi mereka yang pemula.



**Gambar 13.3:** Logo Microsoft Access

#### 4. Linear Structural Relationship (LISREL)

Lisrel (Linear Structural Relationship) adalah software statistic yang ketiga yang paling sering digunakan dalam bidang akademis maupun praktisi yang merupakan salah satu program komputer yang dapat mempermudah analisis untuk menyelesaikan masalah-masalah yang tidak dapat diselesaikan oleh perangkat analisis konvensional. Lisrel banyak digunakan dalam berbagai ilmu pengetahuan sosial, sehingga dalam versi terbarunya dibuat lebih interaktif dan mudah, serta memiliki banyak fitur yang cukup lengkap. Penerapannya juga makin luas di bidang ilmu sosial dan ilmu perilaku untuk pengambilan keputusan.



**Gambar 13.4:** Logo LISREL

### 5. Statistical Analysis System (SAS)

SAS adalah singkatan dari Statistical Analysis System yang disediakan oleh SAS Institute Inc. SAS pertama kali dikembangkan oleh Anthony J. Barr pada tahun 1966. Seorang mahasiswa lulusan dari North Carolina State University lulus 1962-1964.

SAS menggunakan bahasa pemrograman, dan memungkinkan programmer melakukan entri data, analisis statistik, peramalan untuk mendukung keputusan riset operasi, peningkatan kualitas pengembangan aplikasi data dan lain sebagainya. Pemograman dalam SAS dikategorikan dalam 2 jenis, yaitu data step untuk membuat, membaca atau pun memanipulasi data, dan proc step (procedure step) digunakan untuk menganalisa, meringkas atau pun membuat tabulasi dari sebuah data. SAS dibangun sejak versi 6.0 hingga versi 9.2.

Adapun analisa yang dapat dilakukan dengan menggunakan SAS adalah :

1. Base SAS – Basic Procedure and Data Management
2. SAS/STAT – Statistical Analysis
3. SAS/GRAFH – Graphic and presentation
4. SAS/OR – Operation research
5. SAS/ETS – Econometric and Time Series
6. SAS/IML – Interactive matrix language
7. SAS/AF – Application facility
8. SAS/QC – Quality control
9. SAS/INSIGHT – Data mining

10. SAS/PH – Clinical trial analysis
11. Enterprise Miner – Data mining



**Gambar 13.5:** Logo SAS

## 6. Eviews

Eviews adalah singkatan dari Econometric Views, yang pertama kali dirilis versi 1 nya pada tahun 1994. Eviews termasuk salah satu program yang baru namun populer. Eviews sendiri dikembangkan oleh Quantitative Micros Software (QMS). Sayangnya aplikasi Eviews ini hanya tersedia untuk sistem operasi windows, jadi pengguna OS linux dan Mac belum bisa menikmati aplikasi ini.



**Gambar 13.6:** Logo Eviews

## 7. STATA

STATA adalah singkatan dari Statistika dan Data yaitu program komputer yang digunakan untuk analisa statistika dan dibuat oleh StataCorp tahun 1985. Dirancang untuk keperluan ekonomi, sosiologi dan epidemiologi dengan berbagai fitur manajemen data, analisis statistika, grafik, simulasi dan pemrograman.

Masih banyak software statistika lainnya yang biasa digunakan dalam analisis data. Untuk software bisa didownload versi trialnya. Analisis data yang pernah saya gunakan adalah SPSS, MINITAB, AMOS, EVIEWS dan yang paling sering digunakan adalah SPSS dan MINITAB.



**Gambar 13.7:** Logo STATA

#### 8. Analysis of Moment Structure (AMOS)

AMOS adalah singkatan dari Analysis of Moment Structure yang merupakan salah satu software yang digunakan untuk mengestimasi model pada model SEM (Structural Equation Model). Dengan menggunakan Amos maka perhitungan rumit dalam SEM akan jauh lebih mudah dilakukan dibandingkan dengan menggunakan perangkat lunak lainnya sehingga untuk orang-orang awam sekalipun yang bukan ahli statistik akan dapat menggunakan dan memahaminya dengan mudah. Selain itu, penggunaan Amos akan mempercepat dalam membuat spesifikasi, melihat serta melakukan modifikasi model secara grafik dengan menggunakan tool yang sederhana. AMOS diambil alih oleh Microsoft untuk disesuaikan dengan versi SPSS saat ini.



**Gambar 13.8:** Logo AMOS

## 9. R-Software

Software ini adalah aplikasi gratis dibawah lisensi GNU. Software ini dapat digunakan untuk menganalisis ilmu sosial maupun eksakta. Beberapa fiturnya bisa dirubah namun aplikasi ini menggunakan bahasa pemrograman sehingga terkadang menyulitkan bagi pengguna baru.



**Gambar 13.9:** R-Software

## 10. Statistics Open For All (SOFA)

SOFA atau Statistics Open For All adalah software analisis statistik open source dan gratis untuk Windows. Di dalamnya, Anda dapat membuat dan mengedit lembar data yang kompleks dan juga menganalisis statistik, melakukan tes statistik dasar, membuat bagan, membuat tabel proyek yang menarik, dan banyak lagi.

Penekanan utama dari software ini adalah pada kemudahan penggunaan dan itu pasti unggul pada bagian itu. Plus, untuk pengguna pemula, ini menyediakan bagian Memulai yang mengarahkan Anda ke halaman bantuan online yang berisi semua langkah bersama screenshot untuk menggunakan software ini.



**Gambar 13.10** Logo SOFA

# Daftar Pustaka

- Adytya, B., (2019). Perbedaan Populasi dan Sampel, Ketahui Pengertian Hingga Contohnya. [Online] Available at: <https://www.merdeka.com/trending/perbedaan-populasi-dan-sampel-ketahui-pengertian-hingga-contohnya-kln.html?page=2>[Accessed 18 Maret 2021].
- Algifari. (2016) "Statistika Induktif untuk Ekonomi dan Bisnis ". Yogyakarta : UPP STIM YKPN
- Amirullah, (2017). Populasi dan Sampel (pemahaman, jenis dan teknik). [Online] Available at: <https://zenodo.org/record/825326/files/POPULASI%20DAN%20SAMP EL.pdf>.[Accessed Maret 2021].
- Amrhein, V., Trafimow, D. and Greenland, S. (2018) ‘Inferential statistics as descriptive statistics: there is no replication crisis if we don’t expect replication’, PeerJ Preprints, 6, p. e26857v4.
- Anwar, H. (2017) <https://www.statistikian.com/2017/03/perbedaan-uji-normalitas-dan-homogenitas.html>. Available at: <https://www.statistikian.com/2017/03/perbedaan-uji-normalitas-dan-homogenitas.html>.
- Arikunto, S., (2006). Prosedur Penelitian: Suatu Pengantar Praktik. Jakarta: Rineka Cip.
- Asuero, A. G., Sayago, A., & Gonzales, A. G. (2006). The Correlation Coefficient: An Overview. Critical Reviews in Analytical Chemistry, 41–59.

- Aunuddin (2005) Statistika: rancangan dan analisis data. Bogor: Bogor IPB Press. Tersedia pada: <http://kikp.pertanian.go.id/pustaka/opac/detail-opac?id=52846>.
- Babbie, E., (2004). The practice of social research. Belmont CA: Wadsworth.
- Baron, M. (2013) Probability and Statistics for Computer Scientists, Probability and Statistics for Computer Scientists. doi: 10.1201/b14800.
- Berman, E. and Wang, X. (2020) ‘Measures of Dispersion’, Exercising Essential Statistics, pp. 39–44. doi: 10.4135/9781506348933.n7.
- Brynjolfsson, E., Rock, D. and Syverson, C. (2018) ‘Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics’, in The economics of artificial intelligence: An agenda. University of Chicago Press, pp. 23–57.
- Burns, R. B. (1996). Introduction to Research Methods. Melbourne: Addison Wesley.
- Castro Sotos, A. E. et al. (2007) ‘Students’ Misconceptions of Statistical Inference: A Review of the Empirical Evidence from Research on Statistics Education’, Educational Research Review, 2(2), pp. 98–113. doi: 10.1016/j.edurev.2007.04.001.
- Castro Sotos, A. E. et al. (2009) ‘How Confident are Students in Their Misconceptions about Hypothesis Tests?’, Journal of Statistics Education, 17(2). doi: 10.1080/10691898.2009.11889514.
- Christopher, A. N. (2016) Interpreting and Using Statistics in Psychological Research. SAGE Publications.
- Cobran, G. W., 1953. Sampling Techniques. 2nd ed. New York: John Wiley and Sons Inc.
- Cooper, L. L. and Shore, F. S. (2008) ‘Students’ Misconceptions in Interpreting Center and Variability of Data Represented Via Histograms and Stem-and-Leaf Plots’, Journal of Statistics Education, 16(2). doi: 10.1080/10691898.2008.11889559.
- Dawson, B. (2004) ‘Basic and Clinical Biostatistics’. New York: McGraw-Hill. Available at: <http://accessmedicine.mhmedical.com/book.aspx?bookId=356>.

- Delmas, R. and Liu, Y. (2005) ‘Exploring Students’ Conceptions of the Standard Deviation’, *Statistics Education Research Journal*, 4(1), pp. 55–82. Available at: [https://www.researchgate.net/publication/241601603\\_Exploring\\_student\\_s'\\_conceptions\\_of\\_the\\_standard\\_deviation](https://www.researchgate.net/publication/241601603_Exploring_student_s'_conceptions_of_the_standard_deviation).
- DelMas, R. C. (2004) A Comparison of Mathematical and Statistical Reasoning: The Challenge of Developing Statistical Literacy, Reasoning, and Thinking. Edited by D. B.-Z. J. Garfield. Dordrecht, The Netherlands: The Netherlands: Kluwer Academic Publishers. Available at: <https://experts.umn.edu/en/publications/a-comparison-of-mathematical-and-statistical-reasoning>.
- Didin Astriani Prasetyowati (2015) MODUL KULIAH STATISTIKA PROBABILITAS. Palembang: UNIVERSITAS INDO GLOBAL MANDIRI.
- Farooqui, S. (2020) Normality test with python. Why it is important in the first place?, <https://medium.com/analytics-vidhya/normality-test-with-python-cb3ef0d7fb2>. Available at: <https://medium.com/analytics-vidhya/normality-test-with-python-cb3ef0d7fb2> (Accessed: 8 April 2021).
- Fischbein, E. and Schnarch, D. (1997) ‘The Evolution with Age of Probabilistic, Intuitively Based Misconceptions’, *Journal for Research in Mathematics Education*. US: National Council of Teachers of Mathematics, 28(1), pp. 96–105. doi: 10.2307/749665.
- Frankel, J. R. & Norman, E. W., (1993). How to design and Evaluate Research in Education. 2nd ed. New York: McGraw hill Inc.
- Friedman, J. H. (1998) ‘Data Mining and Statistics: What’s the connection?’, *Computing science and statistics*, 29(1), pp. 3–9.
- Furqoni, M. R. (2021) Penyajian Data. Tersedia pada: <https://nilaimutlak.id/penyajian-data/>.
- Garfield, J. B. et al. (2008) Developing Students’ Statistical Reasoning: Connecting Research and Teaching Practice, *Developing Students’ Statistical Reasoning: Connecting Research and Teaching Practice*. doi: 10.1007/978-1-4020-8383-9.

- Gay, L. R. & Diehl, P. L., (1992). Research Methods for Business and Management. 2nd ed. New York: MacMillan Publishing Company.
- Gibbons, J. D. and Chakraborti, S. (2020) Nonparametric statistical inference. CRC press.
- Gravetter, F. J. and Wallnau, L. B. (2017) Statistic for the Behavioral Science. Available at: [https://vulms.vu.edu.pk/Courses/PSY516/Downloads/Statistics\\_for\\_the\\_Behavioral\\_Sciences\\_by\\_Frederick\\_J.\\_Gravetter,\\_Larry\\_B.\\_Wallnau\\_\(z-lib.org\).pdf](https://vulms.vu.edu.pk/Courses/PSY516/Downloads/Statistics_for_the_Behavioral_Sciences_by_Frederick_J._Gravetter,_Larry_B._Wallnau_(z-lib.org).pdf).
- Gunawan,Imam. (2016). Pengantar Statistika Inferensial, Jakarta : Rajawali Pres.
- Hair, J. F., Anderson, R. E., Thatam, R. L. & Black, W. C., (1998). Multivariate Data Analysis. Englewood Cliffs: Prentice Hall.
- Hanlon , B. & Large, B., (2011). Samples and Populations, s.l.: Department of Statistics University of Wisconsin—Madison.
- Harlan, Johan (2007) "Analisis Regresi Linier". Jakarta: Penerbit Gunadarma
- Hayati, R., (2020). Jenis Populasi dan Sampel dalam Penelitian. [Online] Available at: <https://penelitianilmiah.com/jenis-populasi-dan-sampel/> [Accessed 22 Maret 2021].
- Herlawati and Handayanto, R. T. (2019) 'Penggunaan Matlab dan Python dalam Klasterisasi Data', Jurnal Kajian Ilmiah (JKI), 20(1), pp. 103–118. Available at: <http://ejurnal.ubharajaya.ac.id/index.php/JKI>.
- Hidayati, T., Handayani, I. and Ines Heidiani, I. (2019) STATISTIKA DASAR.
- Hirsch, L. S. and O'Donnell, A. M. (2001) 'Representativeness in Statistical Reasoning: Identifying and Assessing Misconceptions', Journal of Statistics Education, 9(2). doi: 10.1080/10691898.2001.11910655.
- Holcomb, Z. C. (1998). Fundamentals of Descriptive Statistics. USA: Pyrczak Publishing.
- Huck, S. W. (2009) Statistical Misconceptions. New York: Psychology Press, Taylor & Francis. Available at: <https://www.taylorfrancis.com/books/mono/10.4324/9780203889282/statistical-misconceptions-schuyler-huck>.

- I Ketut Swarjana (2016) Statistik Kesehatan. Diedit oleh A. A. C. Yogyakarta: Andi. Tersedia pada: <https://opac.perpusnas.go.id/DetailOpac.aspx?id=1176760>.
- Jarmin, R. S. (2019) ‘Evolving measurement for an evolving economy: thoughts on 21st century US economic statistics’, *Journal of Economic Perspectives*, 33(1), pp. 165–184.
- Kabay, M. E. (2019) ‘Statistics in Business, Finance, Management and Information Technology’. Available at: [http://www.mekabay.com/courses/academic/norwich/qm213/statistics\\_t\\_ext.pdf](http://www.mekabay.com/courses/academic/norwich/qm213/statistics_t_ext.pdf).
- Kaur, P. et al. (2018) ‘Descriptive statistics’, *International Journal of Academic Medicine*, 4(1), p. 60.
- Komunitas (2018) Penerapan Statistika Matematika Dalam Kehidupan Sehari-hari | Komunitas Blogger Matematika. Tersedia pada: <http://komunitasbloggermatematika.blogspot.com/2018/01/penerapan-statistika-matematika-dalam.html> (Diakses: 12 April 2021).
- Kotz, S. and Johnson, N. L. (100AD) ‘Encyclopedia of Statistical Scis. [Vol. 09] Oakes’s Test of Concordance to Preference Functions’, 1.
- Kumar, R. (2020) ‘Safety Statistics and Accident Inspection’, PMIR Department, Patna University.
- Kurniawan, A., (2021). Pengertian Populasi Menurut Para Ahli. [Online] Available at: <https://www.gurupendidikan.co.id/populasi-adalah/> [Accessed 18 Maret 2021].
- Kurniawati, G. N. (2020) Kursus Belajar Data: Yuk Kenali Skala Pengukuran Sebelum Melangkah ke Proses Analisis Data! Available at: <https://dqlab.id/kursus-belajar-data-kenali-skala-pengukuran-data> (Accessed: 2 April 2021).
- Latief, K. A. (2009) ‘Ukuran Penyebaran’, pp. 1–7.
- Lindgren, B. (2017) Statistical theory. Routledge.
- Liu, T.-C. (2010) ‘Developing Simulation-based Computer Assisted Learning to Correct Students’ Statistical Misconceptions based on Cognitive Conflict Theory, using “Correlation” as an Example’, *Journal of Educational Technology & Society. International Forum of Educational*

- Technology & Society, 13(2), pp. 180–192. Available at: <http://www.jstor.org/stable/jeductechsoci.13.2.180>.
- Malhotra, K. N., (1993). Marketing Research and Applied Orientation. 2nd ed. USA: Prentice Hall International.
- Manikandan, S. (2011) ‘Measures of Dispersion’, Journal of Pharmacology and Pharmacotherapeutics, 2(4), pp. 315–316. doi: 10.4103/0976-500X.85931.
- McCormick, K. and Salcedo, J. (2017) SPSS statistics for data analysis and visualization. John Wiley & Sons.
- Meletiou, M. M. (2000) Developing Students’ Conceptions of Variation: An Untapped Well in Statistical Reasoning, Dissertation. The University of Texas at Austin. Available at: <https://iaseweb.org/documents/dissertations/00.Meletiou.pdf>.
- Miller, J. D. (2017) Statistics for Data Science: Leverage the power of statistics for Data Analysis, Classification, Regression, Machine Learning, and Neural Networks. Packt Publishing Ltd.
- Mishra, S. (2020) Methods for Normality Test with Application in Python, Towards Data Science. Available at: <https://towardsdatascience.com/methods-for-normality-test-with-application-in-python-bb91b49ed0f5> (Accessed: 8 April 2021).
- Momoh, O., (2021). Population. [Online] Available at: <https://www.investopedia.com/terms/p/population.asp> [Accessed 18 Maret 2021].
- Moore, D. S. (1997) ‘New Pedagogy and New Content: The Case of Statistics’, International Statistical Review, 65(2), pp. 123–137. doi: 10.1111/j.1751-5823.1997.tb00390.x.
- Mukaka, M. M. (2012). Statistics Corner: A Guide to Appropriate use of Correlation Coefficient in Medical Research. Malawi Medical Journal, 69-71.
- Naik, L. and Kumar, S. D. K. (2016) ‘Basic Statistics for Libraries and for Analysis of Research Data’, Proceedings of the National Conference on Research Methodology in Library and Information Science, (September).

- Niningprastiwi, (2017). Pengertian Populasi dalam Penelitian. [Online] Available at: <http://niningprastiwi27.blogspot.com/2017/02/pengertian-populasi-dalam-penelitian.html>[Accessed 22 Maret 2021].
- Nuryadi (2017) Dasar-Dasar Statistik Penelitian. Yogyakarta: Sibuku Media. Tersedia pada: eprints.mercubuana-yogya.ac.id.
- Nuryadi, dkk (2017) " Dasar- Dasar Statistik Penelitian". Yogyakarta: SiBuku Media.
- Nusar Hajarisman (no date) 'PEMERIKSAAN DATA BERPENGARUH DALAM MODEL REGRESI GAMMA', Universitas Islam Bandung, pp. 79–91.
- Otok, B. W. (2016) Pengumpulan dan Penyajian Data.
- Oussous, A. et al. (2018) 'Big Data technologies: A survey', Journal of King Saud University-Computer and Information Sciences, 30(4), pp. 431–448.
- Pollatsek, A., Lima, S. and Well, A. D. (2009) 'Concept or Computation: Students' Understanding of the Mean', Educational Studies in Mathematics, 12(2), pp. 191–204. Available at: <https://www.srri.umass.edu/sites/srri/files/PollatsekLimaWell1981/index.pdf>.
- Pratikno, A. S. et al (2020) 'Pemetaan Ukuran Pemusatan Data'. Available at: <https://osf.io/v3n9h/>.
- putri Rika Dwiana (2020) PERBANDINGAN KEKUATAN UJI METODE KOLMOGOROV-SMIRNOV, ANDERSON-DARLING, DAN SHAPIRO-WILK UNTUK MENGUJI NORMALITAS DATA. Sanata Dharma.
- Rahman, S. A. (2020) BIOSTATISTIK DESKRIPTIF. Yogyakarta: Penerbit Samudra Biru.
- Rasdihan (2009) Metode Statistik Deskriptif untuk Umum. Jakarta: Grasindo.
- Ratner, B. (2017) Statistical and Machine-Learning Data Mining:: Techniques for Better Predictive Modeling and Analysis of Big Data. CRC Press.
- Rayat, C. S. (2018) 'Measures of Central Tendency', in Statistical Methods in Medical Research. Springer, pp. 33–46.

- Retnawati, H., (2017). Teknik Pengambilan Sampel. [Online] Available at: <http://staffnew.uny.ac.id/upload/132255129/pengabdian/15-Teknik%20Penyampelan%20alhamdulillah.pdf>[Accessed 22 Maret 2021].
- Riecan, J. (2020) 'Role of the Information Technology in Official Statistics', pp. 1–9. Available at: <https://www.psa.gov.qa/ar/media/events/Documents/using information technology in official statistics.pdf>.
- Ronal E. Walpole., (1995) "Pengantar Statstika" , Jakarta: PT. Gramedia Pustaka Utama,
- Roscoe, J. T., (1975). Fundamental research statistics for the behavioural sciences. 2nd ed. New York: Holt Rinehart & Winston: s.n.
- Saddoen, A., (2021). Pengertian Populasi : Ciri-Ciri, Jenis-Jenis dan Faktornya. [Online] Available at: <https://moondoggiesmusic.com/pengertian-populasi/#gsc.tab=0>[Accessed 25 Maret 2021].
- Selvamuthu, D. and Das, D. (2018) 'Descriptive Statistics', in Introduction to Statistical Methods, Design of Experiments and Statistical Quality Control. Springer, pp. 63–96.
- Sethna, J. (2021) Statistical mechanics: entropy, order parameters, and complexity. Oxford University Press, USA.
- Setiawan, A. (2021) UKURAN PEMUSATAN DATA: MEAN - MEDIAN - MODE, SMARTSTAT. Available at: <https://www.smartstat.info/materi/statistika/statistik-deskriptif/ukuran-pemusatan-data-mean-median-mode.html>.
- Seymour Lipschutz, M. L. (2008) Matematika Diskrit. edisi keti. Edited by S.. Lemeda Simarmata. Erlangga.
- Shah, J. K. (2019) 'Measures of Dispersion', pp. 320–341. Available at: <https://castudyweb.com/wp-content/uploads/2019/06/measures-of-dispersion.pdf>.
- Sharon (2020) Macam-Macam Teknik Pengumpulan Data: Skala Pengukuran Data yang Dapat Kamu Implementasikan Setelah Melakukan Teknik Pengumpulan Data. Available at: <https://www.dqlab.id/macam-macam-teknik-pengolahan-data-dalam-data-science> (Accessed: 2 April 2021).

- Shruti (2021) Measures of Dispersion: A Close View. Available at: <https://www.economicsdiscussion.net/dispersion/measures-of-dispersion-a-close-view/12173>.
- Sinambela, M. and Suryaningsih, E. (2014) ‘A COMPARISON OF METHODS FOR TESTING HOMOGENEITY OF AVERAGE TEMPERATURE AND PRECIPITATION SERIES’, pp. 978–602.
- Singh, R. K., Tiwari, N. and Prasad, R. C. (2015) ‘Significant Role of Statistics in Computational Sciences’, International Journal of Computer Applications Technology and Research, 4(12), pp. 952–955. doi: 10.7753/ijcatr0412.1014.
- Singpurwalla, D. (2018) ‘A handbook of statistics: An overview of statistical methods’.
- Slovin, J. M., (1960). Sampling. 1st ed. New York: Simon and Schuster Inc.
- Srava Chrisdes Antoro, M. S. (2005) Konsep Probabilitas. Available at: [http://srava\\_chrisdes.staff.gunadarma.ac.id/Downloads/folder/0.3](http://srava_chrisdes.staff.gunadarma.ac.id/Downloads/folder/0.3) (Accessed: 22 March 2021).
- Subchan (2015) Matematika SMP/MTs Kelas IX. Jakarta: Pendidikan, Kementerian Kebudayaan, D A N Indonesia, Republik.
- Sudaryono M.Pd (2012) Statistika Probabilistik. Edited by N. WK. Yogyakarta: Penerbit Andi Yogyakarta.
- Sudibyo, N. (2020) “Penyajian Data,” (August). Tersedia pada: [https://www.researchgate.net/publication/343883588\\_Penyajian\\_Data](https://www.researchgate.net/publication/343883588_Penyajian_Data).
- Sudjana. (2001). Metoda Statistika. Bandung: Tarsito.
- Sugiyono (2016) Statistik untuk Penelitian. Bandung: Alfabeta. Tersedia pada: <https://onesearch.id/Record/IOS6863.ai:slims-355>.
- Sugiyono, (2005). Statistika Untuk Penelitian . Bandung: Penerbit Alfabeta
- Sugiyono, (2019). Metode Penelitian Kuantitatif, Kualitatif, dan R&D. 2 ed. Bandung: CV. Alfabeta.
- Suharsimi, A., (2007). Prosedur Penelitian Suatu Pendekatan Praktik. Jakarta: Rineka Cipta.
- Sujalu, A. P. et al. (2021) STATISTIK EKONOMI 1. Zahir Publishing. Available at: <https://books.google.co.id/books?id=RIAjEAAAQBAJ>.

- Sujarweni, W. V., (2015). Metodologi Penelitian Bisnis dan Ekonomi. In: Pertama ed. Yogyakarta: Pustaka Baru Press.
- Sundaram, K. R., Dwivedi, S. N. and Sreenivas, V. (2010) Medical Statistics: Principles and Methods. Tunbridge Wells; New Delhi: Anshan ; BI Publications Pvt. Ltd.
- Supardi, (1993). Laporan Penelitian: Populasi Dan Sampel Penelitian. UNISIA, TAHUNXIIITRIWULAN VI(17), pp. 100-108.
- Supranto (2000) Statistik : Teori dan Aplikasi. Jakarta: Erlangga.
- Supranto, J. (2000). Statistik: Teori dan Aplikasi, Jilid 1. Jakarta: Erlangga.
- Susilana, R., (2020). Modul 6 Populasi dan Sampel. [Online] Available at: [http://file.upi.edu/Direktori/DUAL-MODES/PENELITIAN\\_PENDIDIKAN/BBM\\_6.pdf](http://file.upi.edu/Direktori/DUAL-MODES/PENELITIAN_PENDIDIKAN/BBM_6.pdf)[Accessed Maret 2021].
- Suyono. (2018) "Analisis Regresi untuk Penelitian" Yogyakarta: Deepublish
- Swinscow, T. D. V. and Campbell, M. J. (2002) Statistics at Square One. Bmj London.
- Syukra Alhamda (2016) Buku Ajar Metlit dan Statistik. Yogyakarta: Deepublish. Tersedia pada: <https://penerbitbukudeepublish.com/shop/buku-ajar-metlit/>.
- Teknozone.ID (2020) 15 Aplikasi Analisis Statistik Terbaik 2020. Tersedia pada: <https://teknozone.id/aplikasi-analisis-statistik-terbaik/> (Diakses: 10 April 2021).
- Tendencies, C. (2017) Concept of Dispersion. IGNOU. Available at: <http://www.egyankosh.ac.in/bitstream/123456789/23455/1/Unit-3.pdf>.
- Tri Basuki, Agus (2016) "Analisis Regresi dalam Penelitian Ekonomi dan Bisnis". Jakarta: Rajawali PERS
- Triyono, (2003). Teknik Sampling dalam Penelitian. [Online].
- Ulty (2020) Penerapan Statistika Dalam Kehidupan Sehari-Hari | Lancang Kuning. Tersedia pada: <https://lancangkuning.com/post/15120/penerapan-statistika-dalam-kehidupan-sehari-hari.html> (Diakses: 12 April 2021).

- Uma , S. & Bougie, R., (2017). Metode Penelitian untuk Bisnis:Pendekatan Pengembangan-Keahlian. Edisi 6, Buku 1, Cetakan Kedua ed. Jakarta: Salemba Empat.
- Viandari, E. (2020) Mean, Median, dan Modus, Quipper. Available at: <https://www.quipper.com/id/blog/mapel/matematika/mean-median-dan-modus/>.
- Weihls, C. and Ickstadt, K. (2018) ‘Data science: the impact of statistics’, International Journal of Data Science and Analytics, 6(3), pp. 189–194.
- Wibowo, S. (2016) Peluang Kejadian Bersyarat. Available at: [https://sumberbelajar.belajar.kemdikbud.go.id/sumberbelajar/tampil/PEL\\_UANG-KEJADIAN-BERSYARAT-2016/menu3.html](https://sumberbelajar.belajar.kemdikbud.go.id/sumberbelajar/tampil/PEL_UANG-KEJADIAN-BERSYARAT-2016/menu3.html) (Accessed: 23 March 2021).
- Widarjono, Agus (2015) “Statistika Terapan dengan Excel & SPSS. Yogyakarta: UPP STIM YKPN
- Wikipedia (2019) Blaise Pascal. Available at: [https://id.wikipedia.org/wiki/Blaise\\_Pascal](https://id.wikipedia.org/wiki/Blaise_Pascal) (Accessed: 22 March 2021).
- Wild, C. J., Utts, J. M. and Horton, N. J. (2018) ‘What is statistics?’, in International handbook of research in statistics education. Springer, pp. 5–36.
- Zaka, M. N. (2019) ‘Implementasi Ukuran Pemusatan Data Menggunakan Python’. Available at: [https://github.com/yogjay/StatistikaTerapan/blob/master/Implementasi\\_Ukuran\\_Pemusatan\\_Data\\_Menggunakan\\_Python.ipynb](https://github.com/yogjay/StatistikaTerapan/blob/master/Implementasi_Ukuran_Pemusatan_Data_Menggunakan_Python.ipynb).



# Biodata Penulis



**Akbar Iskandar** dilahirkan di Jepuru pada tanggal 21 bulan Juli tahun 1986, jepuru terletak di desa Padang Kec. Gantarang kabupaten bulukumba, beralamat tinggal di Jalan Lembo dengan Nomor 175 Kota Makassar, pekerjaan nya sehari-hari menjadi salah satu dosen di STMIK AKBA – Makassar, riwayat pendidikan berawal di SD Negeri Nomor 232 Dampang (Bulukumba), melanjutkan ke SMP Negeri Nomor 5 Gangking (Bulukumba), dan selanjutnya ke Sekolah Menengah Kejuruan di SMK Negeri 1 Bulukumba, beranjak remaja maka ia lanjutkan Pendidikan Tinggi pada Tahap Strata 1 (S1) di Jurusan Pendidikan Teknik Elektronika Universitas Negeri Makassar, dan Strata 2 (S2) diselesaikan di Jurusan Penelitian dan Evaluasi Pendidikan Universitas Negeri Makassar. Kemudian ia kembali melanjutkan pendidikan S2 yang kedua kalinya pada bidang Sistem Komputer di makassar (STMIK Handayani). Selanjutnya melanjutkan pendidikan S3 di salah satu Universitas Negeri di Yogyakarta yaitu UNY. Selain itu, ia juga aktif diberbagai kegiatan ilmiah/penelitian dan menulis buku. Beberapa Artikel Ilmiah nya sudah terindeks di Scopus dan WOS, aktif menjadi Reviewer diberbagai Jurnal baik Nasional maupun Internasional.

Selain itu ia aktif dlm kegiatan Konferensi Nasional dan Tingkat internasional, berkolaborasi dengan berbagai peneliti baik Nasional maupun Internasional sudah menjadi gaya hidupnya dalam melaksanakan Penelitian, ia juga aktif dalam berbagai kegiatan Pengabdian pada Masyarakat.



**Muttaqin, S.T., M.Cs**Lahir dan besar di Aceh. Pendidikan TK hingga SMA diselesaikan di Kabupaten Aceh Utara dan Kabupaten Bireun Provinsi Aceh. Menyelesaikan Pendidikan D3 Instrumentasi & Komputasi di Universitas Syiah Kuala, S1 Teknik Informatika Sekolah Tinggi Teknik Bina Cendikian Banda Aceh, dan S2 Ilmu Komputer Universitas Gadjah Mada. Mengajar mata kuliah Sistem Operasi Komputer, Kecerdasan Buatan, Sistem Informasi Geografis, Pemodelan Sistem Informasi, Teknik Digital, Pemograman C++, Sistem Basis Data, E-Commerce. Sukses menerbitkan buku kolaborasi sebanyak 12 buku dengan judul: Biometrika Teknologi Identifikasi, E-Commerce: Implementasi, strategi & Inovasinya, Panduan Belajar Manajemen Referensi dengan Mendeley, Etika Profesi: Membangun Profesionalisme Diri, Keamanan Data dan Informasi, MOOC: Platform Pembelajaran Daring di Abad 21, Pembelajaran Daring untuk Pendidikan Teori dan Penerapan, Pengenalan Teknologi Informasi, Sistem Pendukung Keputusan: Metode & Implementasi, Sistem Informasi Manajemen, Tren Teknologi Masa Depan, Pengantar Forensik Teknologi Informasi. Semuanya diterbitkan oleh Penerbit Kita Menulis. Email penulis muttaqin.ugm@gmail.com, Hp. +6285260409204.



**Sarini Vita Dewi** Lulusan S1 Teknik Elektro tahun 2010, bidang Teknologi Informasi di Universitas Syiah Kuala Banda Aceh dengan gelar S.T. Menyelesaikan pendidikan S2 di Universitas Gadjah Mada pada Prodi Teknik Elektro Magister Teknologi Informasi (MTI) pada tahun 2014. Dari tahun 2015 hingga saat ini menetap di banda aceh dan berprofesi sebagai dosen di Fakultas Ilmu Komputer Universitas Ubudiyah Indonesia, Banda Aceh



**Jamaludin, M.Kom**, seorang praktisi dan akademisi yang lahir di Bah Jambi, 11 Januari 1973 memiliki latar belakang sarjana teknik informatika dari Sekolah Tinggi Poliprofesi Medan dan magister komputer dari Universitas Sumatera Utara dengan peminatan komputer. Saat ini bertugas sebagai dosen di Politeknik Ganesha Medan sejak tahun 2013 sampai sekarang. Aktif dalam penelitian dan pengabdian kepada masyarakat untuk merealisasikan kerja dosen dalam Tri Dharma Perguruan Tinggi.

Mulai aktif menulis buku sejak September 2019 sampai sekarang. Kemudian aktif juga menulis artikel di media cetak/online mulai sejak September 2020 sampai sekarang. Tema yang digemari dalam penulisan buku adalah komputer, bisnis online, technopreneurship dan pendidikan.



**Dr. Irawati HM, S.E.,M.Ak.**, lahir di Pare-pare pada tanggal 28 Desember 1987. Ia tercatat sebagai lulusan S1 dan S2 Universitas Muslim Indonesia Jurusan Akuntansi dan S3 Universitas Hasanuddin Makassar. Wanita yang kerap disapa Ira ini adalah anak dari pasangan H.Muin (ayah) dan Hj. Saeja (ibu). Irawati HM berprofesi sebagai Tenaga Pendidik (Dosen). Pada 2019 lalu, Ira berhasil menyelesaikan studi S3 di Universitas Hasanuddin Makassar dan meraih gelar Doktor.



**Cahyo Prianto**, Pria kelahiran 27 juli 1984 ini telah menempuh pendidikan di jurusan Pendidikan Fisika Universitas Pendidikan Indonesia dan dilanjutkan di Sekolah Teknik Elektro dan Informatika, Institut Teknologi Bandung. Saat ini, penulis beraktivitas sebagai Dosen di Program Studi Teknik Informatika Politeknik Pos Indonesia, dengan mengampu beberapa mata kuliah seperti Data Mining, Statistika dan Aljabar Linear. Penulis dapat dihubungi melalui email cahyoprianto@poltekpos.ac.id.



**Rosmita Sari Siregar, S.Pd.I., M.Pd** lahir di Medan 25 Desember 1991 dari pasangan yang bernama Alm, Syahnan Siregar dan Ibu Hasmarida Harahap. Anak ketiga dari lima bersaudara.

Menempuh pendidikan SD tamat tahun 2003, melanjutkan ke MTs Negeri 2 Medan tamat tahun 2006, kemudian menyelesaikan Madrasah Aliyah Negeri 1 Medan tamat tahun 2009. Pendidikan Strata satu diselesaikan pada tahun 2013 pada Jurusan Pendidikan Matematika IAIN Sumatera Utara Medan. Magister Pendidikan dari Universitas Negeri

Medan dengan program studi Pendidikan Matematika pada tahun 2015. Riwayat pekerjaan pada tahun 2009-2014 sebagai Guru Bidang Studi Matematika Yayasan Pendidikan Islam Ar-Ridha, tahun 2015 hingga sekarang sebagai Dosen Tetap di Universitas Prima Indonesia.

Beberapa artikel sudah terpublikasi melalui jurnal nasional tak terakreditasi dan jurnal nasional terakreditasi sinta 1- 6. Pada tahun 2019 mendapatkan Dana Hibah untuk penelitian dosen pemula sebagai ketua peneliti. Buku ini merupakan buku kolaborasi keempat bagi penulis di Yayasan Kita Menulis. Semoga buku ini banyak memberikan kontribusi dan manfaat bagi para pembaca dan dapat menjadi inspirasi penulis lainnya.

Email : [rosmitasarisiregar@gmail.com](mailto:rosmitasarisiregar@gmail.com)



**Muhammad Noor Hasan Siregar, ST, M.Kom** Lulusan S1. Teknik Industri (ST) dari Universitas Andalas Padang dan melanjutkan pendidikan di Teknik Informatika Universitas Putra Indonesia “YPTK” Padang (M.Kom). Sekarang bekerja sebagai Dosen di Universitas Graha Nusantara Padangsidimpuan dari tahun 2011. Informasi lebih lanjut dapat dilihat melalui blog di alamat <http://hasan.dosen.ugn.ac.id> dan bisa dihubungi melalui email noor.siregar@gmail.com



**Dina Chamidah, S.Pd., SH., M.Si., M.Kn., C. STMI** Dosen di Program Studi Pendidikan Biologi, Fakultas Bahasa dan Sains, Universitas Wijaya Kusuma Surabaya. Memperoleh gelar Sarjana Pendidikan Biologi dari FPMIPA Universitas Negeri Surabaya, juga Sarjana Hukum dari Fakultas Hukum Universitas Kartini, Surabaya. Magister Sains dari Fakultas Kedokteran Hewan Universitas Airlangga, juga Magister Kenotariatan dari Universitas Surabaya, Pendidikan Khusus Advokat PERADI dari Universitas Surabaya, Pendidikan Sekolah

Trainer dan Motivator Indonesia di Yogyakarta dan sekarang masih belajar di Magister Hukum di Fakultas Hukum Universitas Wijaya Kusuma Surabaya dan juga sebagai Awardee LPDP pada Program Doktoral Pendidikan Biologi di Universitas Negeri Malang. Anggota Organisasi: PBI, IDRI, PDRI, FDI, Kodepena, KPII, ADI, FKDI, Divisi Pelatihan TAPLAI 2 Lembaga Ketahanan Nasional Republik Indonesia (LEMHANNAS RI), Jakarta, Indonesia, pernah menjadi calon legislatif DPR RI dari Partai PERINDO, pernah menjabat sebagai Notaris Pengganti di Kota Mojokerto dan juga pernah menjabat sebagai Ketua Program Studi Pendidikan Biologi di Universitas Wijaya Kusuma Surabaya, juga sebagai anggota EURASIA (TERA, STRA, SSHRA, HBSRA), dan dari tahun 2019 menjadi International Referee Board dari TAFFD's. Mulai tahun 2020 sebagai Team Peer Reviewer Jurnal Teknosains Kodepena, Member dari IAN (International Association of Neuroscience), ISDR (International Society for Dermatology Research), Ikatan Ilmuwan Indonesia International (i4), Ambassador of Sustainability, Wakil Ketua Divisi Akademik dan Publikasi Ilmiah Mata Garuda 2.0, Pengurus Divisi Temu Ilmiah Neurosaintis Muda Indonesia dan juga sebagai Founder, Owner dan Advokat di DC Law Firm. Minat utama saya adalah Biologi, Pendidikan Biologi, Ilmu Biologi Reproduksi, Zoologi, Pendidikan, Manajemen Pendidikan, Manajemen, Teknologi Pendidikan, Hukum dan Kenotariatan.



**Marzuki Sinambela**, lahir di Tarutung pada tanggal 7 Desember 1983. Ia menyelesaikan kuliah dan mendapat gelar Diploma Ahli Madya pada September 2005 di Akademi Meteorologi dan Geofisika (AMG) Jakarta, Sarjana Teknik Informatika pada Oktober 2008 di STMIK Budidarma, Magister Teknik pada Agustus 2011 dan Doktor Ilmu Fisika pada November 2020 di Universitas Sumatera Utara dalam bidang keilmuan Physics Computation, Instrument dan Machine Learning,. Saat ini aktif mengajar di Universitas Methodist Indonesia (UMI) dan Peneliti Muda di Badan Meteorologi Klimatologi dan Geofisika (BMKG).



**Dr. Albinur Limbong** dilahirkan di Kabupaten Simalungun Sumatera Utara pada tanggal 30 Oktober 1965. Penulis menamatkan pendidikan sarjana di IKIP Medan dalam bidang Pendidikan Fisika tahun 1990. Kemudian menyelesaikan pendidikan Magister kemudian Doktor dalam bidang Fisika di Macquarie University Sydney Australia tahun 2000. Penulis adalah dosen Matematika dan Fisika di Fakultas Teknologi Informasi UNAI, dan sekarang mendapat tugas tambahan sebagai Wakil Rektor Bidang Akademik sejak tahun 2014.



**Yusra Fadhillah, S.Kom., M.Kom** lahir di Jakarta pada tanggal 18 Juni 1991. Menyelesaikan kuliah dan mendapat gelar Sarjana Teknik Informatika pada 02 Februari 2016. Dan merupakan salah satu alumni Jurusan Teknik Informatika pada Fakultas Ilmu Komputer Universitas Putra Indonesia “YPTK” Padang. Dan pada tahun yang sama, bekerja di Telkom Akses sebagai NOC (Network Operation Center) atau OMC (Operation and Maintenance Center) di Kota Padang. Dan pada tahun 2017 melanjutkan Program Magister Teknologi Informasi dan lulus pada tahun 24 September 2018 dari Universitas

Putra Indonesia “YPTK” Padang. Pada tahun 2019 diangkat menjadi Dosen Universitas Graha Nusantara Padangsidiimpuan dan ditempatkan di Fakultas Teknik pada Program Studi Ilmu Komputer atau Informatika.



**Janner Simarmata**, Sarjana Teknik Informatika dari STMIK Bandung, Magister Ilmu Komputer dari Universitas Gadjah Mada (UGM) dan Doktor Pendidikan Teknologi Kejuruan (PTK) diperoleh dari Universitas Pendidikan Indonesia (UPI) Bandung bidang kajian Blended Learning.

Menulis buku sejak tahun 2005. Dosen di Pendidikan Teknologi Informatika dan Komputer (PTIK) Fakultas Teknik Universitas Negeri Medan.



# Statistika

## Bidang Teknologi Informasi

Statistika salah satu disiplin ilmu yang paling penting untuk menyediakan alat dan metode untuk menemukan struktur dan memberikan wawasan yang lebih dalam tentang data dan disiplin yang paling penting untuk menganalisis dan mengukur ketidakpastian. Maka dari itu, metode statistik diterapkan di semua bidang yang melibatkan pengambilan keputusan untuk membuat kesimpulan yang akurat dari kumpulan data, sehingga menghasilkan beberapa informasi yang berarti tentang data itu atau membuat keputusan dalam menghadapi ketidakpastian berdasarkan metodologi statistik.

Buku ini membahas:

- Bab 1 Konsep Dasar Statistika
- Bab 2 Jenis Data dan Skala Pengukuran
- Bab 3 Pemeriksaan Data
- Bab 4 Teknik Penyajian Data
- Bab 5 Populasi dan Sampel
- Bab 6 Probabilitas
- Bab 7 Data Kelompok dan Analisisnya
- Bab 8 Ukuran Pemusatan
- Bab 9 Ukuran Penyebaran
- Bab 10 Normalitas dan Homogenitas
- Bab 11 Analisis Korelasi
- Bab 12 Analisis Regresi
- Bab 13 Aplikasi Statistika



YAYASAN KITA MENULIS  
[press@kitamenulis.id](mailto:press@kitamenulis.id)  
[www.kitamenulis.id](http://www.kitamenulis.id)

ISBN 978-623-342-047-1

9 786233 420471