



BIOSTATISTIKA: ANALISIS DATA UNTUK KESEHATAN DAN PENELITIAN

PENULIS:

- Ryryn Suryaman Prana Putra
- Rahmawati
- Asruria Sani Fajriah
- Resti Arian
- Suryani Mansyur
- Muhammad Al Azhary
- Nurpatwa Wilda Ningsi
- Nurul Hidayah
- Dwi Yulia Maritasari

BIOSTATISTIKA: ANALISIS DATA UNTUK KESEHATAN DAN PENELITIAN

Ryryn Suryaman Prana Putra

Rahmawati

Asruria Sani Fajriah

Resti Ariani

Suryani Mansyur

Muhammad Al Azhary

Nurpatwa Wilda Ningsi

Nurul Hidayah

Dwi Yulia Maritasari



GET PRESS INDONESIA

BIOSTATISTIKA: ANALISIS DATA UNTUK KESEHATAN DAN PENELITIAN

Penulis :

Ryryn Suryaman Prana Putra
Rahmawati
Asruria Sani Fajriah
Resti Arian
Suryani Mansyur
Muhammad Al Azhary
Nurpatwa Wilda Ningsi
Nurul Hidayah
Dwi Yulia Maritasari
ISBN : 978-634-255-733-4

Editor : Mila Sari, M. Si.

Desain Sampul dan Tata Letak : Aulia Syaharani, S.Tr.Kes.

PENERBIT : GET PRESS INDONESIA

Anggota IKAPI No. 033/SBA/2022
Jl. Palarik RT 01 RW 06 Kelurahan Air Pacah
Kecamatan Koto Tangah Padang Sumatera Barat
website: www.getpress.co.id
email: adm.getpress@gmail.com

Cetakan pertama, September 2026

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

KATA PENGANTAR

Puji dan syukur kami panjatkan ke hadirat Tuhan Yang Maha Esa, yang telah memberikan rahmat dan karunia-Nya sehingga buku berjudul *Biostatistika: Analisis Data untuk Kesehatan dan Penelitian* ini dapat diselesaikan. Isi buku ini mencakup pembahasan tentang pendahuluan dan capaian pembelajaran, konsep dasar biostatistika, pengumpulan dan manajemen data kesehatan, penyajian data: tabel, grafik, dan diagram, ukuran pemusatan dan penyebaran data, distribusi probabilitas dan teori dasar, uji hipotesis parametrik dan nonparametrik, analisis korelasi dan regresi, dan biostatistik dalam epidemiologi dan kesehatan masyarakat. Akhir kata, semoga buku ini dapat menjadi referensi dan dapat memberikan manfaat serta inspirasi bagi semua pembaca.

Padang, Juli 2026

Penulis

DAFTAR ISI

KATA PENGANTAR.....	i
DAFTAR ISI.....	ii
DAFTAR GAMBAR	vii
DAFTAR TABEL.....	viii
BAB 1 PENDAHULUAN DAN CAPAIAN PEMBELAJARAN.....	1
1.1 Hakikat Biostatistik dalam Ilmu Kesehatan dan Penelitian....	1
1.2 Peran Biostatistika dalam Pengambilan Keputusan Kesehatan Masyarakat.....	4
1.3 Ruang Lingkup Biostatistika dalam Analisis Data Kesehatan	7
1.4 Kedudukan Buku Ini dalam Pembelajaran Biostatistika untuk Kesehatan dan Penelitian.....	10
1.5 Capaian Pembelajaran Dasar Biostatistika, Jenis Data, dan Skala Pengukuran	13
1.6 Capaian Pembelajaran Pengumpulan, Manajemen, dan Penyajian Data Kesehatan	15
1.7 Capaian Pembelajaran Ukuran Statistik Deskriptif dan Distribusi Probabilitas.....	18
1.8 Capaian Pembelajaran Statistik Inferensial dan Uji Hipotesis	20
1.9 Capaian Pembelajaran Analisis Hubungan, Prediksi, ANOVA, dan Uji Lanjut.....	23
1.10 Capaian Pembelajaran Penerapan Biostatistika dalam Epidemiologi, Kesehatan Masyarakat, dan Software Statistik.....	26
1.11 Integrasi Capaian Pembelajaran Bab 1 dengan Struktur Keseluruhan Buku	29
DAFTAR PUSTAKA	33
BAB 2 KONSEP DASAR BIOSTATISTIKA	39
2.1 Konsep Biostatistika: Pengertian dan Ruang Lingkup.....	42

2.1.1 Konsep Dasar Biostatistika.....	42
2.1.2 Biostatistika sebagai Ilmu dan Alat Analisis.....	43
2.1.3 Ruang Lingkup Biostatistika dalam Penelitian Kesehatan.....	43
2.1.4 Ruang Lingkup Biostatistika dalam Praktik Kesehatan.....	44
2.1.5 Keterkaitan Biostatistika dengan Ilmu Lain.....	47
2.2 Peran Biostatistika dalam Bidang Kesehatan.....	51
2.2.1 Peran dalam Penelitian Klinis	51
2.2.2 Peran dalam Epidemiologi.....	53
2.2.3 Peran dalam Kesehatan Masyarakat	54
2.2.4 Peran dalam Diagnostik dan Laboratorium Klinik	55
2.2.5 Peran dalam Pengambilan Kebijakan Kesehatan	56
DAFTAR PUSTAKA	58
BAB 3 PENGUMPULAN DAN MANAJEMEN DATA KESEHATAN	61
3.1 Peran Pengumpulan dan Manajemen Data dalam Biostatistika.....	63
3.1.1 Data Sebagai Fondasi Analisis Statistik	63
3.1.2 Hubungan Pengumpulan Data, Manajemen Data, dan Biostatistika	64
3.1.3 Konsep " <i>Garbage In, Garbage Out</i> " Dalam Penelitian Kesehatan.....	65
3.1.4 Dampak Kualitas Data Terhadap Hasil Analisis	66
3.2 Metode Pengumpulan Data Kesehatan	67
3.2.1 Wawancara	67
3.2.2 Kuesioner.....	68
3.2.3 Observasi.....	69
3.2.4 Telaah Dokumen.....	71
3.3 Instrumen Pengumpulan Data Kesehatan	72

3.4 Manajemen Data Kesehatan	75
3.4.1 Siklus Manajemen Data	75
3.4.2 Strategi Meminimalkan Bias Pengumpulan Data	78
3.4.3 Dokumentasi dan Metadata Data Kesehatan	79
DAFTAR PUSTAKA	80
BAB 4 PENYAJIAN DATA: TABEL, GRAFIK, DAN DIAGRAM	91
4.1 Tujuan Penyajian Data	92
4.2 Prinsip Umum Penyajian Data	93
4.2.1 Penyajian Data dalam Bentuk Tabel	94
4.3 Penyajian Data dalam Bentuk Grafik	100
4.4 Penyajian Data dalam Bentuk Diagram	109
4.5 Bentuk Penyajian Berdasarkan Jenis Data	112
4.6 Kesalahan Umum dalam Penyajian Data	113
4.7 Etik dalam Penyajian Data Kesehatan	116
DAFTAR PUSTAKA	117
BAB 5 UKURAN PEMUSATAN DAN PENYEBARAN DATA	119
5.1 Ukuran Pemusatan	119
5.1.1 <i>Mean</i>	120
5.1.2 Median	122
5.1.3 Modus	125
5.2 Ukuran Penyebaran Data (ukuran variasi / dispersi)	127
5.2.1 <i>Range</i> (jangkauan)	127
5.2.2 Varians (Keragaman)	128
5.2.3 Standar Deviasi (Simpangan Baku)	132
5.2.4 Koefisien Keragaman	133
Keterangan: KV = koefisien keragaman	133
DAFTAR PUSTAKA	135
BAB 6 DISTRIBUSI PROBABILITAS DAN TEORI DASAR	137
6.1 Pendekatan Probabilitas	138

6.1.1 Klasik.....	138
6.1.2 Frekuensi.....	138
6.1.3 Subyektif.....	139
6.2 Konsep Dasar.....	140
6.3 Distribusi Probabilitas dalam Kesehatan.....	144
6.3.1 Konsep Dasar Distribusi Normal.....	144
6.3.2 Analisis Distribusi Binomial dan Poisson dalam Kesehatan.....	144
6.3.3 Implementasi Distribusi Probabilitas dalam Sektor Kesehatan.....	145
6.4 Konsep Dasar Probabilitas dalam Kesehatan.....	146
6.4.1 Konsep Probabilitas dan Risiko dalam Ranah Kesehatan.....	146
6.4.2 Estimasi Probabilitas Kejadian Penyakit dalam Populasi.....	146
6.4.3 Signifikansi Interpretasi Probabilitas dalam Praktik Klinis.....	147
6.5 Permutasi dan Kombinasi	148
6.5.1 Menyusun Objek dalam Satu Garis (Permutasi)	148
6.5.2 Konsep Kombinasi	149
DAFTAR PUSTAKA	150
BAB 7 UJI HIPOTESIS PARAMETRIK DAN NONPARAMETRIK.....	151
7.1 Statistik Parametrik.....	153
7.1.1 Pengujian Satu Sampel (<i>One Sample T-Test</i>)	154
7.1.2 Pengujian Dua Sampel Bebas (<i>Independent Sample t-Test</i>).....	156
7.1.3 Pengujian Dua Sampel Berkaitan (<i>Paired Sample T-Test</i>).....	158
7.2 Statistik Nonparametrik	160
7.2.1 Uji Wilcoxon Untuk Sampel Berpasangan	161

7.2.2 Uji Mann Whitney.....	164
7.2.3 Uji Kruskal Wallis.....	166
7.2.4 Uji Chi-Square	167
DAFTAR PUSTAKA	171
BAB 8 ANALISIS KORELASI DAN REGRESI	173
8.1 Korelasi dan Regresi.....	174
8.1.1 Korelasi.....	177
8.1.2 Uji Regresi	183
DAFTAR PUSTAKA	193
BAB 9 BIOSTATISTIK DALAM EPIDEMIOLOGI DAN KESEHATAN MASYARAKAT	195
9.1 Konsep Dasar Biostatistik	196
9.2 Hubungan Biostatistik dalam Epidemiologi.....	198
9.2.1 Epidemiologi sebagai Ilmu Dasar Kesehatan Masyarakat.....	199
9.2.2 Peran Biostatistik dalam Epidemiologi	200
9.2.3 Biostatistik dalam Evidence-Based Public Health.....	201
9.3 Data dalam Epidemiologi.....	203
9.4 Ukuran Epidemiologi dalam Biostatistik.....	204
9.4.1 Pengukuran Frekuensi Penyakit	205
9.4.2 Insidens dan Prevalens	206
9.4.3 Ukuran Mortalitas dan Morbiditas	207
9.4.4 Pengukuran Risiko Penyakit.....	209
DAFTAR PUSTAKA	211
BIODATA PENULIS	215

DAFTAR GAMBAR

Gambar 2. 1 Manajemen Data Kesehatan yang Berkualitas	77
Gambar 4. 1 Boxplot Perbandingan Tekanan Darah Sistolik Berdasarkan Status Merokok.....	107
Gambar 8. 1 Bagan Keputusan Pemilihan Uji Korelasi dan Regresi.....	182
Gambar 8. 2 Asumsi Klasik Regresi Linier dan Cara Pengujiannya	184

DAFTAR TABEL

Tabel 4. 1 Distribusi Responden Menurut Status Merokok pada Penelitian Hipertensi di Puskesmas X Tahun 2026.....	96
Tabel 4. 2 Distribusi Responden Berdasarkan Kelompok Umur	97
Tabel 4. 3 Hubungan Status Merokok dengan Kejadian Hipertensi.....	98
Tabel 4. 4 Karakteristik Numerik Responden Penelitian.....	99
Tabel 4. 5 Hasil Analisis Faktor Risiko Kejadian Hipertensi.....	100
Tabel 6. 1 Hubungan Antar Peristiwa.....	141
Tabel 6. 2 Kombinasi 2 Koin.....	142
Gambar 7. 1nKategori, Asumsi, dan Jenis Data Statistika.....	152
Tabel 7. 1 Kategorisasi Level Effect Size (d).....	156
Tabel 7. 2 Kategorisasi Persentase Pengaruh dalam Uji Perbedaan <i>Independent Samples t-Test</i>	158
Tabel 8. 1 Perbedaan Analisis Korelasi dan Regresi.....	175
Tabel 8. 2 Jenis Uji Korelasi Parametrik	177
Tabel 8. 3 Jenis Uji Korelasi Non Parametrik.....	178
Tabel 8. 4 Panduan Interpretasi Uji Korelasi.....	180
Tabel 8. 5 Jenis Model Regresi Alternatif Berdasarkan Tipe Variabel Dependen (Y).....	186

BAB 1

PENDAHULUAN DAN CAPAIAN PEMBELAJARAN

Oleh Ryryn Suryaman Prana Putra

1.1 Hakikat Biostatistik dalam Ilmu Kesehatan dan Penelitian

Biostatistika merupakan cabang ilmu statistik yang digunakan untuk memahami, mengolah, menganalisis, dan menafsirkan data yang berkaitan dengan kehidupan, kesehatan, penyakit, pelayanan kesehatan, dan penelitian biomedis. Ilmu ini tidak hanya membahas angka, tetapi juga membantu peneliti menjawab pertanyaan ilmiah secara terukur. Dalam kesehatan masyarakat, biostatistika berfungsi sebagai alat untuk membaca pola masalah kesehatan pada populasi. Data tentang kejadian penyakit, faktor risiko, status gizi, kematian, kelahiran, perilaku kesehatan, dan mutu layanan kesehatan membutuhkan pendekatan statistik agar dapat ditafsirkan secara tepat. Oleh karena itu, biostatistika menjadi dasar penting dalam penelitian kesehatan karena membantu mengubah data mentah menjadi informasi yang dapat dipertanggungjawabkan secara ilmiah (Daniel & Cross, 2018; Sullivan, 2022; Dawson & Trapp, 2020).

Hakikat biostatistika dapat dipahami melalui tiga unsur utama, yaitu data, variasi, dan ketidakpastian. Data menggambarkan fakta empiris yang diperoleh melalui pengukuran, observasi, pencatatan, atau survei. Variasi

muncul karena manusia, lingkungan, perilaku, dan kondisi biologis tidak selalu sama antara satu individu dengan individu lain. Ketidakpastian muncul karena peneliti sering bekerja dengan sampel, bukan seluruh populasi. Biostatistika membantu peneliti mengelola variasi dan ketidakpastian tersebut melalui metode estimasi, pengujian hipotesis, analisis hubungan, dan penarikan kesimpulan secara rasional (Rosner, 2016; Motulsky, 2018; Peacock & Peacock, 2020).

Dalam sejarah perkembangan ilmu statistik, Ronald A. Fisher memberikan dasar penting bagi rancangan penelitian dan inferensi statistik. Fisher memperkenalkan prinsip eksperimental seperti randomisasi, replikasi, dan pengendalian variasi melalui desain percobaan. Randomisasi bertujuan mengurangi bias dalam pembagian kelompok penelitian. Replikasi membantu memastikan bahwa hasil penelitian tidak muncul secara kebetulan. Prinsip ini masih digunakan dalam penelitian klinis, laboratorium, epidemiologi, kesehatan lingkungan, dan intervensi kesehatan masyarakat, meskipun penerapannya saat ini telah berkembang sesuai konteks penelitian modern (Fisher, 1935/1990; Hulley *et al.*, 2022; Portney, 2020).

Selain Fisher, Jerzy Neyman dan Egon Pearson juga memberi kontribusi besar melalui teori pengujian hipotesis. Teori Neyman-Pearson menjelaskan bahwa pengambilan keputusan statistik perlu mempertimbangkan hipotesis nol, hipotesis alternatif, tingkat signifikansi, kesalahan tipe I, kesalahan tipe II, dan kekuatan uji. Dalam penelitian kesehatan, teori ini membantu peneliti menentukan apakah perbedaan atau hubungan yang ditemukan pada sampel cukup kuat untuk mendukung kesimpulan ilmiah. Misalnya, peneliti dapat menguji apakah terdapat perbedaan tekanan darah antara kelompok intervensi dan kelompok kontrol. Kerangka ini menjadi fondasi dalam banyak uji statistik, seperti uji-t, chi-square, ANOVA, korelasi, regresi, dan

berbagai uji nonparametrik (Neyman & Pearson, 1933; Daniel & Cross, 2018; Sullivan, 2022).

Biostatistika juga berkaitan erat dengan teori probabilitas karena banyak peristiwa kesehatan bersifat tidak pasti. Risiko seseorang mengalami penyakit tertentu tidak dapat dipastikan secara mutlak, tetapi dapat diperkirakan berdasarkan pola data dan faktor risiko yang tersedia. Probabilitas membantu peneliti memahami peluang kejadian, distribusi data, kesalahan pengambilan sampel, dan batas keyakinan hasil penelitian. Dalam epidemiologi, konsep probabilitas digunakan untuk menghitung risiko relatif, odds ratio, attributable risk, dan ukuran asosiasi lainnya. Dengan demikian, biostatistika memberikan bahasa kuantitatif untuk menjelaskan risiko kesehatan secara lebih terukur dan tidak hanya berdasarkan dugaan subjektif (Kestenbaum, 2019; Rothman *et al.*, 2021; Gordis, 2019).

Dalam penelitian kesehatan, biostatistika tidak boleh dipahami sebagai proses mekanis memilih uji statistik semata. Peneliti perlu memahami pertanyaan penelitian, desain studi, jenis data, skala pengukuran, distribusi data, asumsi analisis, dan konteks kesehatan yang sedang dikaji. Kesalahan memilih metode statistik dapat menghasilkan kesimpulan yang keliru, meskipun data yang digunakan tampak lengkap. Karena itu, biostatistika menuntut kemampuan berpikir kritis, bukan hanya kemampuan menjalankan perangkat lunak statistik. Hakikat biostatistika terletak pada kemampuan menghubungkan masalah kesehatan, data, metode analisis, dan interpretasi ilmiah secara utuh (Hernán & Robins, 2020; Vittinghoff *et al.*, 2012; Agresti, 2018).

Biostatistika menjadi semakin penting karena penelitian kesehatan saat ini menghasilkan data dalam jumlah besar dan dengan bentuk yang semakin beragam. Data kesehatan dapat berasal dari survei rumah tangga, rekam medis elektronik, sistem surveilans, laboratorium, klaim asuransi kesehatan,

aplikasi digital, dan penelitian lapangan. Perkembangan ini menuntut peneliti untuk memiliki literasi data yang baik. Literasi data mencakup kemampuan memahami sumber data, kualitas data, bias, validitas, reliabilitas, serta etika penggunaan data. Tanpa pemahaman biostatistika, data kesehatan yang besar dapat menghasilkan keputusan yang salah karena dianalisis tanpa landasan metodologis yang memadai (Harris *et al.*, 2021; Wickham *et al.*, 2023; WHO, 2020).

1.2 Peran Biostatistika dalam Pengambilan Keputusan Kesehatan Masyarakat

Kesehatan masyarakat membutuhkan keputusan yang berbasis bukti karena kebijakan dan program kesehatan memengaruhi kehidupan banyak orang. Biostatistika membantu pengambil keputusan memahami besarnya masalah, kelompok yang paling terdampak, faktor risiko utama, dan efektivitas intervensi. Tanpa analisis statistik, keputusan kesehatan dapat hanya bergantung pada persepsi, tekanan politik, atau pengalaman terbatas. Data yang dianalisis dengan baik dapat menunjukkan apakah suatu masalah benar-benar prioritas atau hanya tampak menonjol karena pemberitaan. Oleh karena itu, biostatistika berperan sebagai dasar rasional dalam perencanaan, pelaksanaan, pemantauan, dan evaluasi program kesehatan masyarakat (Friis & Sellers, 2021; Detels *et al.*, 2021; Brownson *et al.*, 2018).

Salah satu teori yang relevan dalam pengambilan keputusan kesehatan adalah evidence-based medicine yang dipopulerkan oleh David Sackett dan kolega. Teori ini menekankan bahwa keputusan klinis perlu menggabungkan bukti ilmiah terbaik, keahlian profesional, dan nilai atau kebutuhan pasien. Dalam kesehatan masyarakat, prinsip ini berkembang menjadi evidence-based public health.

Pendekatan tersebut menuntut kebijakan kesehatan disusun berdasarkan data epidemiologis, hasil evaluasi program, analisis biaya, dan konteks sosial masyarakat. Dengan demikian, biostatistika berperan sebagai instrumen utama untuk menilai kualitas bukti yang digunakan dalam keputusan kesehatan (Sackett *et al.*, 1996; Brownson *et al.*, 2018; Detels *et al.*, 2021).

Biostatistika juga mendukung proses identifikasi prioritas masalah kesehatan. Dalam praktik kesehatan masyarakat, tidak semua masalah dapat diselesaikan sekaligus karena sumber daya selalu terbatas. Data prevalensi, insidensi, angka kematian, disabilitas, beban penyakit, dan distribusi risiko dapat membantu menentukan masalah mana yang harus ditangani lebih dahulu. Misalnya, daerah dengan angka kematian ibu yang tinggi membutuhkan intervensi berbeda dari daerah dengan dominasi penyakit tidak menular. Penggunaan statistik membuat proses penentuan prioritas menjadi lebih terbuka, terukur, dan dapat dipertanggungjawabkan kepada masyarakat (Gordis, 2019; Kestenbaum, 2019; WHO, 2023).

Dalam evaluasi program kesehatan, biostatistika membantu menilai apakah suatu intervensi benar-benar memberikan perubahan. Evaluasi tidak cukup hanya melihat bahwa program telah dilaksanakan atau anggaran telah digunakan. Evaluasi perlu membandingkan indikator sebelum dan sesudah program, antara wilayah intervensi dan pembanding, atau antara kelompok sasaran dan non-sasaran. Analisis statistik membantu menentukan apakah perubahan yang terlihat cukup bermakna atau hanya merupakan variasi biasa. Dengan demikian, biostatistika memperkuat akuntabilitas program kesehatan karena hasilnya dapat dinilai berdasarkan indikator yang jelas (Hulley *et al.*, 2022; Portney, 2020; Creswell & Creswell, 2023).

Teori lain yang penting adalah model Donabedian tentang mutu pelayanan kesehatan. Avedis Donabedian menjelaskan bahwa mutu layanan dapat dinilai melalui tiga komponen utama, yaitu struktur, proses, dan outcome. Struktur mencakup sumber daya, fasilitas, tenaga, alat, dan sistem pendukung. Proses mencakup cara pelayanan diberikan, kepatuhan terhadap standar, komunikasi, dan keselamatan pasien. Outcome mencakup perubahan status kesehatan, kepuasan pasien, komplikasi, kematian, atau capaian program; semua indikator tersebut membutuhkan data dan analisis statistik agar mutu layanan dapat dievaluasi secara objektif (Donabedian, 1988; Joshi *et al.*, 2014; McLaughlin & Olson, 2023).

Dalam epidemiologi, biostatistika sangat penting untuk menilai hubungan antara paparan dan penyakit. Austin Bradford Hill memperkenalkan pertimbangan kausal yang sering digunakan untuk menilai apakah suatu paparan dapat dianggap berhubungan secara kausal dengan suatu penyakit. Pertimbangan tersebut mencakup kekuatan asosiasi, konsistensi, spesifisitas, temporalitas, gradien biologis, plausibilitas, koherensi, eksperimen, dan analogi. Meskipun kriteria Hill tidak boleh digunakan secara kaku, kerangka ini membantu peneliti menilai bukti secara lebih hati-hati. Biostatistika mendukung penilaian tersebut melalui ukuran asosiasi, interval kepercayaan, analisis multivariat, dan pengendalian faktor perancu (Hill, 1965; Rothman *et al.*, 2021; Celentano & Szklo, 2019).

Pengambilan keputusan kesehatan masyarakat juga membutuhkan komunikasi risiko yang jelas. Hasil analisis statistik sering kali sulit dipahami oleh masyarakat jika hanya disampaikan dalam bentuk angka teknis. Peneliti perlu menerjemahkan hasil seperti odds ratio, relative risk, p-value, dan confidence interval ke dalam bahasa yang mudah dipahami. Biostatistika yang baik tidak berhenti pada perhitungan, tetapi berlanjut pada penafsiran yang jujur dan

tidak menyesatkan. Dengan cara ini, biostatistika membantu memperkuat kepercayaan publik terhadap kebijakan kesehatan dan mencegah penyalahgunaan data dalam komunikasi program (Gigerenzer, 2015; Spiegelhalter, 2019; Motulsky, 2018).

1.3 Ruang Lingkup Biostatistika dalam Analisis Data Kesehatan

Ruang lingkup biostatistika dimulai dari pemahaman terhadap data kesehatan. Data kesehatan dapat berbentuk data individu, rumah tangga, komunitas, fasilitas kesehatan, lingkungan, atau populasi. Data tersebut dapat diperoleh melalui survei, eksperimen, observasi, rekam medis, registri penyakit, sistem surveilans, dan basis data administratif. Setiap sumber data memiliki kekuatan dan keterbatasan. Karena itu, analisis biostatistika harus dimulai dari penilaian terhadap asal data, cara pengumpulan data, kelengkapan data, dan potensi bias yang mungkin muncul (Sullivan, 2022; Hulley *et al.*, 2022; Polit & Beck, 2021).

Ruang lingkup berikutnya adalah pengelompokan jenis data dan skala pengukuran. Stanley Smith Stevens memperkenalkan empat skala pengukuran, yaitu nominal, ordinal, interval, dan rasio. Skala nominal digunakan untuk kategori tanpa urutan, seperti jenis kelamin atau jenis penyakit. Skala ordinal digunakan untuk kategori yang memiliki tingkatan, seperti tingkat pendidikan atau derajat nyeri. Skala interval dan rasio digunakan untuk data numerik, seperti suhu, usia, kadar hemoglobin, tekanan darah, berat badan, dan tinggi badan; pemahaman skala ini penting karena menentukan metode analisis yang tepat (Stevens, 1946; Agresti, 2018; Daniel & Cross, 2018).

Biostatistika juga mencakup statistik deskriptif. Statistik deskriptif bertujuan menggambarkan data secara ringkas, jelas, dan informatif. Ukuran yang sering digunakan meliputi frekuensi, persentase, rasio, proporsi, rata-rata, median, modus, rentang, varians, simpangan baku, dan kuartil. Statistik deskriptif membantu pembaca memahami karakteristik responden, distribusi variabel, dan kecenderungan umum data. Dalam laporan penelitian kesehatan, statistik deskriptif menjadi bagian penting karena memberikan gambaran awal sebelum analisis yang lebih kompleks dilakukan (Rosner, 2016; Dawson & Trapp, 2020; Petrie & Sabin, 2020).

Ruang lingkup biostatistika juga mencakup statistik inferensial. Statistik inferensial digunakan untuk menarik kesimpulan tentang populasi berdasarkan data sampel. Proses ini melibatkan konsep estimasi parameter, interval kepercayaan, pengujian hipotesis, p-value, power, dan ukuran efek. Statistik inferensial penting karena sebagian besar penelitian kesehatan tidak mungkin mengukur seluruh populasi. Dengan teknik inferensial, peneliti dapat memperkirakan apakah temuan pada sampel cukup kuat untuk menjelaskan pola yang mungkin terjadi pada populasi yang lebih luas (Sullivan, 2022; Motulsky, 2018; Peacock & Peacock, 2020).

Biostatistika juga mencakup analisis hubungan dan prediksi. Analisis hubungan digunakan untuk mengetahui apakah dua atau lebih variabel memiliki keterkaitan. Korelasi digunakan untuk melihat arah dan kekuatan hubungan antara variabel numerik, sedangkan regresi digunakan untuk menjelaskan atau memprediksi perubahan suatu variabel berdasarkan variabel lain. Dalam kesehatan masyarakat, regresi sering digunakan untuk menganalisis faktor risiko penyakit, kepatuhan pengobatan, mutu pelayanan, perilaku kesehatan, dan determinan sosial kesehatan. Analisis ini membantu peneliti memahami hubungan yang lebih

kompleks, terutama ketika terdapat banyak faktor yang bekerja secara bersamaan (Vittinghoff *et al.*, 2012; Gelman *et al.*, 2020; Hernán & Robins, 2020).

Analisis biostatistika juga mencakup perbandingan antar kelompok. Uji statistik seperti uji-t, Mann-Whitney, chi-square, ANOVA, Kruskal-Wallis, dan uji lanjut digunakan untuk menilai perbedaan antar kelompok. Pemilihan uji tergantung pada tujuan penelitian, jumlah kelompok, jenis data, distribusi data, dan terpenuhinya asumsi analisis. Dalam penelitian kesehatan, perbandingan antar kelompok banyak digunakan untuk menilai efektivitas intervensi, perbedaan status kesehatan, atau perbedaan paparan risiko. Pemahaman ruang lingkup ini penting agar peneliti tidak sekadar mengikuti rumus, tetapi mampu memilih metode yang sesuai dengan pertanyaan penelitian (Daniel & Cross, 2018; Field, 2018; Sullivan, 2022).

Ruang lingkup modern biostatistika semakin luas karena berkembangnya perangkat lunak statistik dan analisis data digital. SPSS, R, Excel, Stata, SAS, dan Python membantu peneliti mengelola data dalam ukuran besar dengan proses yang lebih efisien. Namun, perangkat lunak hanya alat bantu dan tidak dapat menggantikan pemahaman metodologis. Kesalahan dalam mendefinisikan variabel, memilih uji, membaca output, atau menafsirkan hasil tetap dapat terjadi meskipun analisis dilakukan dengan perangkat lunak canggih. Karena itu, ruang lingkup biostatistika mencakup kemampuan teknis, pemahaman teori, etika data, dan keterampilan interpretasi hasil penelitian (Field, 2018; Wickham *et al.*, 2023; Harris *et al.*, 2021).

1.4 Kedudukan Buku Ini dalam Pembelajaran Biostatistika untuk Kesehatan dan Penelitian

Buku ini disusun untuk membantu mahasiswa, dosen, peneliti pemula, dan praktisi kesehatan memahami biostatistika secara bertahap. Banyak pembelajar kesehatan menganggap statistik sebagai materi yang sulit karena terlalu banyak rumus dan istilah teknis. Buku ini menempatkan biostatistika sebagai alat berpikir ilmiah, bukan sekadar kumpulan perhitungan. Dengan pendekatan tersebut, pembaca diharapkan memahami alasan suatu metode digunakan, bukan hanya cara menjalankan analisis. Kedudukan buku ini adalah sebagai jembatan antara konsep statistik, data kesehatan, dan praktik penelitian kesehatan masyarakat (Daniel & Cross, 2018; Sullivan, 2022; Dawson & Trapp, 2020).

Secara pedagogis, buku ini sejalan dengan pendekatan outcome-based education atau pendidikan berbasis capaian pembelajaran. William G. Spady menjelaskan bahwa pendidikan berbasis capaian menekankan hasil belajar yang jelas, terukur, dan relevan dengan kompetensi yang dibutuhkan peserta didik. Dalam konteks biostatistika, capaian pembelajaran tidak hanya berupa kemampuan menghitung, tetapi juga kemampuan memahami data, memilih uji, menggunakan perangkat lunak, dan menafsirkan hasil. Pembelajaran biostatistika yang baik harus menghubungkan konsep, latihan, kasus, dan penerapan nyata dalam penelitian kesehatan. Karena itu, setiap bab dalam buku ini dirancang untuk bergerak dari pemahaman dasar menuju kemampuan analisis yang lebih aplikatif (Spady, 1994; Biggs *et al.*, 2022; Creswell & Creswell, 2023).

Buku ini juga menggunakan prinsip constructive alignment yang dikembangkan oleh John Biggs. Prinsip ini menjelaskan bahwa tujuan pembelajaran, aktivitas belajar, dan asesmen harus saling selaras. Dalam buku ini, capaian

pembelajaran diarahkan agar pembaca mampu memahami konsep, menerapkan metode, menganalisis data, dan mengevaluasi hasil. Aktivitas belajar dapat dilakukan melalui pembacaan konsep, latihan soal, interpretasi tabel, pembacaan output software, dan studi kasus kesehatan. Dengan pendekatan ini, biostatistika tidak diajarkan sebagai materi yang terpisah dari penelitian, tetapi sebagai bagian dari proses ilmiah yang utuh (Biggs *et al.*, 2022; Anderson *et al.*, 2001; Portney, 2020).

Kedudukan buku ini juga dapat dilihat dari susunan bab yang berjenjang. Bab awal membahas konsep dasar, jenis data, pengumpulan data, manajemen data, dan penyajian data. Bab tengah membahas ukuran statistik, distribusi probabilitas, statistik inferensial, dan uji hipotesis. Bab lanjutan membahas korelasi, regresi, ANOVA, uji lanjut, penerapan epidemiologi, dan penggunaan software statistik. Susunan ini dibuat agar pembaca tidak langsung masuk ke analisis kompleks tanpa memahami dasar data dan logika statistik. Struktur bertahap seperti ini penting karena kesalahan pada tahap awal, seperti salah mengukur variabel atau salah mengelola data, akan memengaruhi seluruh hasil analisis (Hulley *et al.*, 2022; Polit & Beck, 2021; Wickham *et al.*, 2023).

Buku ini juga menempatkan biostatistika sebagai bagian dari kompetensi penelitian kesehatan. Penelitian kesehatan tidak hanya membutuhkan pemahaman masalah, tetapi juga kemampuan merumuskan pertanyaan penelitian, memilih desain, menentukan sampel, mengukur variabel, menganalisis data, dan menyusun kesimpulan. Dalam setiap tahap tersebut, biostatistika memiliki peran penting. Misalnya, penentuan ukuran sampel membutuhkan pemahaman tentang variasi, tingkat kepercayaan, power, dan efek yang diharapkan. Analisis data juga membutuhkan pemahaman tentang asumsi uji, jenis variabel, dan

interpretasi hasil secara kontekstual (Creswell & Creswell, 2023; Hulley *et al.*, 2022; Portney, 2020).

Buku ini memiliki kedudukan praktis karena menghubungkan teori biostatistika dengan masalah kesehatan masyarakat. Contoh penerapan dapat mencakup analisis kejadian diare, stunting, hipertensi, diabetes, penyakit menular, mutu pelayanan kesehatan, kesehatan lingkungan, keselamatan kerja, dan perilaku kesehatan. Penerapan ini penting agar pembaca melihat hubungan langsung antara angka statistik dan masalah kesehatan nyata. Dalam pendidikan kesehatan masyarakat, pembelajaran yang berbasis kasus membantu mahasiswa memahami bahwa data bukan sekadar angka, tetapi representasi dari kondisi populasi. Dengan demikian, buku ini diharapkan memperkuat kemampuan pembaca dalam menggunakan data untuk menjawab masalah kesehatan secara ilmiah dan bertanggung jawab (Detels *et al.*, 2021; Friis & Sellers, 2021; Brownson *et al.*, 2018).

Pada akhirnya, kedudukan buku ini bukan untuk menggantikan buku statistik lanjutan, tetapi untuk menjadi dasar yang kuat bagi pembaca kesehatan dan penelitian. Pembaca yang telah memahami isi buku ini akan lebih siap mempelajari analisis statistik yang lebih kompleks. Buku ini juga dapat menjadi pegangan awal untuk membaca artikel ilmiah, menyusun proposal penelitian, menganalisis data skripsi atau tesis, dan mengevaluasi laporan program kesehatan. Kekuatan utama buku ini terletak pada penyajian yang sederhana, bertahap, dan terkait langsung dengan konteks kesehatan. Dengan cara ini, biostatistika dapat dipahami sebagai ilmu yang membantu peneliti berpikir lebih jernih, mengambil keputusan lebih tepat, dan menyampaikan bukti ilmiah secara lebih bertanggung jawab (Motulsky, 2018; Spiegelhalter, 2019; Sullivan, 2022).

1.5 Capaian Pembelajaran Dasar Biostatistika, Jenis Data, dan Skala Pengukuran

Capaian pembelajaran pada bagian dasar biostatistika diarahkan agar pembaca memahami bahwa biostatistika bukan sekadar teknik berhitung, melainkan cara berpikir ilmiah untuk membaca data kesehatan. Pembaca perlu mampu menjelaskan pengertian biostatistika, fungsi biostatistika, serta perbedaannya dengan statistik umum. Kemampuan dasar ini penting karena penelitian kesehatan selalu berhadapan dengan data yang berasal dari manusia, lingkungan, pelayanan kesehatan, dan populasi. Pembaca juga diharapkan memahami bahwa setiap angka dalam penelitian memiliki konteks, sumber, dan batas penafsiran. Dengan pemahaman tersebut, pembaca dapat melihat biostatistika sebagai alat untuk membangun bukti ilmiah, bukan sebagai prosedur mekanis yang terpisah dari masalah kesehatan (Daniel & Cross, 2018; Dawson & Trapp, 2020; Sullivan, 2022).

Landasan capaian pembelajaran pada bagian ini dapat dikaitkan dengan teori outcome-based education yang dikembangkan oleh William G. Spady. Teori ini menekankan bahwa pembelajaran harus dimulai dari hasil akhir yang ingin dicapai oleh peserta didik. Dalam konteks biostatistika, hasil belajar tidak cukup hanya berupa kemampuan menghafal definisi, tetapi harus mencakup kemampuan menjelaskan konsep, mengidentifikasi data, memilih metode, dan menafsirkan hasil. Indikator pembelajaran dapat dirumuskan dalam bentuk kemampuan membedakan populasi dan sampel, parameter dan statistik, variabel bebas dan variabel terikat, serta data primer dan data sekunder. Pendekatan ini membuat proses belajar lebih terarah karena pembaca mengetahui kompetensi apa yang harus dikuasai sebelum masuk ke analisis statistik yang lebih lanjut (Spady, 1994; Biggs *et al.*, 2022; Creswell & Creswell, 2023).

Capaian pembelajaran dasar juga dapat dijelaskan melalui taksonomi tujuan belajar yang direvisi oleh Anderson dan Krathwohl. Taksonomi ini membagi kemampuan kognitif ke dalam beberapa tingkatan, yaitu mengingat, memahami, menerapkan, menganalisis, mengevaluasi, dan mencipta. Pada tahap awal pembelajaran biostatistika, pembaca perlu mengingat istilah dasar seperti data, variabel, populasi, sampel, parameter, statistik, distribusi, dan probabilitas. Setelah itu, pembaca perlu memahami hubungan antaristilah tersebut dalam konteks penelitian kesehatan. Pada tahap lebih tinggi, pembaca perlu mampu menerapkan konsep tersebut untuk memilih metode analisis yang sesuai dengan pertanyaan penelitian (Anderson *et al.*, 2001; Biggs *et al.*, 2022; Portney, 2020).

Bagian penting dari capaian pembelajaran ini adalah kemampuan memahami jenis data. Data dalam penelitian kesehatan dapat dibedakan menjadi data kualitatif dan kuantitatif. Data kualitatif menggambarkan kategori, seperti jenis kelamin, status merokok, jenis penyakit, atau tingkat pendidikan. Data kuantitatif menggambarkan nilai numerik, seperti umur, tekanan darah, kadar hemoglobin, indeks massa tubuh, atau lama rawat inap. Pembaca harus mampu mengenali jenis data karena kesalahan mengenali data akan berdampak pada kesalahan memilih tabel, grafik, ukuran statistik, dan uji hipotesis (Agresti, 2018; Rosner, 2016; Sullivan, 2022).

Capaian pembelajaran berikutnya adalah kemampuan memahami skala pengukuran. Stanley Smith Stevens memperkenalkan empat skala pengukuran, yaitu nominal, ordinal, interval, dan rasio. Skala nominal digunakan untuk data kategori tanpa tingkatan, seperti golongan darah atau status penyakit. Skala ordinal digunakan untuk data kategori bertingkat, seperti tingkat nyeri, tingkat kepuasan, atau derajat keparahan penyakit. Skala interval dan rasio digunakan untuk data numerik, tetapi skala rasio memiliki

titik nol absolut, seperti berat badan, tinggi badan, umur, dan kadar gula darah. Pemahaman skala pengukuran menjadi indikator penting karena skala menentukan cara penyajian data dan jenis analisis yang dapat digunakan (Stevens, 1946; Daniel & Cross, 2018; Peacock & Peacock, 2020).

Pada akhir bagian ini, pembaca diharapkan mampu menunjukkan beberapa indikator penguasaan konsep dasar biostatistika. Indikator tersebut meliputi kemampuan menjelaskan fungsi biostatistika dalam penelitian kesehatan, membedakan populasi dan sampel, mengidentifikasi jenis variabel, menentukan skala pengukuran, serta memilih bentuk penyajian awal yang sesuai. Pembaca juga diharapkan mampu menghindari kesalahan umum, seperti memperlakukan data ordinal sebagai data rasio tanpa pertimbangan metodologis. Capaian ini menjadi dasar bagi bab-bab berikutnya karena semua analisis statistik bergantung pada pemahaman data sejak awal. Jika konsep dasar ini kuat, pembaca akan lebih mudah memahami manajemen data, statistik deskriptif, distribusi probabilitas, inferensi, uji hipotesis, regresi, ANOVA, dan penggunaan perangkat lunak statistik (Motulsky, 2018; Field, 2018; Wickham *et al.*, 2023).

1.6 Capaian Pembelajaran Pengumpulan, Manajemen, dan Penyajian Data Kesehatan

Capaian pembelajaran pada bagian pengumpulan data diarahkan agar pembaca memahami bahwa kualitas analisis sangat bergantung pada kualitas data yang dikumpulkan. Data yang dianalisis dengan metode statistik canggih tetap dapat menghasilkan kesimpulan yang salah jika proses pengumpulan datanya lemah. Pembaca perlu mampu menjelaskan sumber data kesehatan, seperti survei, wawancara, observasi, rekam medis, registri penyakit, sistem surveilans, data laboratorium, dan data administratif.

Pembaca juga perlu memahami perbedaan antara data primer dan data sekunder dalam penelitian kesehatan. Dengan kemampuan ini, pembaca dapat menilai apakah data yang digunakan sesuai dengan tujuan penelitian dan cukup kuat untuk menjawab pertanyaan penelitian (Hulley *et al.*, 2022; Polit & Beck, 2021; Creswell & Creswell, 2023).

Landasan teoritis dalam pengumpulan data dapat dikaitkan dengan konsep validitas dan reliabilitas. Validitas menunjukkan sejauh mana instrumen benar-benar mengukur apa yang seharusnya diukur. Reliabilitas menunjukkan sejauh mana hasil pengukuran konsisten jika dilakukan berulang dalam kondisi yang sebanding. Indikator capaian pembelajaran pada aspek ini mencakup kemampuan menjelaskan validitas isi, validitas konstruk, validitas kriteria, reliabilitas antarpemilai, dan reliabilitas internal. Dalam penelitian kesehatan, konsep ini penting karena kesalahan pengukuran dapat menyebabkan bias dan menurunkan kekuatan kesimpulan penelitian (Portney, 2020; Polit & Beck, 2021; Hulley *et al.*, 2022).

Capaian pembelajaran berikutnya adalah kemampuan memahami prinsip manajemen data kesehatan. Manajemen data mencakup proses perancangan format data, pemberian kode variabel, entri data, pemeriksaan kelengkapan, pembersihan data, penyimpanan data, dan dokumentasi. Pembaca perlu memahami bahwa data kesehatan sering mengandung nilai hilang, data ganda, kesalahan input, nilai ekstrem, atau kategori yang tidak konsisten. Kesalahan tersebut harus diperiksa sebelum analisis dilakukan. Dengan demikian, indikator penguasaan manajemen data meliputi kemampuan membuat kode variabel, menyusun data dictionary, melakukan pengecekan data hilang, dan menyiapkan data untuk analisis statistik (Harris *et al.*, 2021; Wickham *et al.*, 2023; Sullivan, 2022).

Prinsip manajemen data modern juga dapat dijelaskan melalui konsep FAIR yang diperkenalkan oleh Wilkinson dan kolega. FAIR merupakan singkatan dari Findable, Accessible, Interoperable, and Reusable. Prinsip ini menekankan bahwa data penelitian harus mudah ditemukan, dapat diakses sesuai etika, dapat digunakan lintas sistem, dan dapat digunakan kembali dengan dokumentasi yang memadai. Dalam konteks kesehatan, penerapan FAIR perlu tetap memperhatikan kerahasiaan identitas responden dan persetujuan etik. Indikator capaian pembelajaran pada bagian ini mencakup kemampuan menjelaskan pentingnya metadata, keamanan data, anonimisasi, dan dokumentasi prosedur analisis (Wilkinson *et al.*, 2016; Harris *et al.*, 2021; WHO, 2020).

Capaian pembelajaran penyajian data diarahkan agar pembaca mampu menyampaikan informasi statistik secara sederhana dan tepat. Penyajian data dapat dilakukan melalui tabel, grafik, diagram, dan ringkasan naratif. Tabel cocok digunakan untuk menyajikan angka secara rinci, sedangkan grafik membantu pembaca melihat pola, perbandingan, tren, dan distribusi. Pembaca perlu memahami bahwa bentuk penyajian harus sesuai dengan jenis data dan tujuan komunikasi. Misalnya, diagram batang lebih cocok untuk data kategori, sedangkan histogram lebih tepat untuk data numerik yang ingin dilihat distribusinya (Field, 2018; Petrie & Sabin, 2020; Wickham *et al.*, 2023).

Pada akhir bagian ini, pembaca diharapkan mampu menghubungkan pengumpulan, manajemen, dan penyajian data sebagai satu rangkaian kerja ilmiah. Data yang baik harus dikumpulkan dengan instrumen yang tepat, dikelola secara rapi, dan disajikan secara jujur. Pembaca perlu mampu menjelaskan bahwa tabel dan grafik tidak boleh dibuat hanya untuk memperindah laporan, tetapi harus membantu pembaca memahami hasil penelitian. Indikator capaian pembelajaran meliputi kemampuan memilih sumber data, menilai kualitas instrumen, membersihkan data, menyusun

tabel distribusi, membuat grafik yang sesuai, dan menulis interpretasi singkat. Capaian ini menjadi dasar penting sebelum pembaca mempelajari ukuran pemusatan, ukuran penyebaran, distribusi probabilitas, dan uji statistik pada bab selanjutnya (Daniel & Cross, 2018; Dawson & Trapp, 2020; Sullivan, 2022).

1.7 Capaian Pembelajaran Ukuran Statistik Deskriptif dan Distribusi Probabilitas

Capaian pembelajaran pada bagian statistik deskriptif diarahkan agar pembaca mampu merangkum data kesehatan secara ringkas, jelas, dan bermakna. Statistik deskriptif membantu peneliti mengenali karakteristik data sebelum melakukan analisis lanjutan. Pembaca perlu memahami bahwa deskripsi data bukan tahap sederhana yang dapat diabaikan. Kesalahan membaca statistik deskriptif dapat menyebabkan salah tafsir terhadap kondisi populasi atau kelompok penelitian. Oleh karena itu, pembaca harus mampu memilih ukuran statistik yang sesuai dengan jenis data, distribusi data, dan tujuan analisis (Rosner, 2016; Daniel & Cross, 2018; Sullivan, 2022).

Ukuran pemusatan data menjadi capaian penting dalam statistik deskriptif. Ukuran ini mencakup rata-rata, median, dan modus. Rata-rata berguna untuk data numerik yang distribusinya relatif simetris, sedangkan median lebih tepat jika data memiliki nilai ekstrem atau distribusi menceng. Modus digunakan untuk menunjukkan nilai atau kategori yang paling sering muncul. Indikator penguasaan pada bagian ini meliputi kemampuan menghitung, memilih, dan menafsirkan ukuran pemusatan sesuai karakteristik data kesehatan (Dawson & Trapp, 2020; Peacock & Peacock, 2020; Petrie & Sabin, 2020).

Ukuran penyebaran data juga menjadi bagian penting dari capaian pembelajaran. Ukuran ini mencakup rentang, varians, simpangan baku, kuartil, rentang interkuartil, dan koefisien variasi. Penyebaran data menunjukkan sejauh mana nilai-nilai dalam suatu variabel tersebar dari pusat datanya. Dua kelompok dapat memiliki rata-rata yang sama, tetapi tingkat variasinya berbeda. Dalam kesehatan masyarakat, informasi variasi penting karena dapat menunjukkan ketimpangan risiko, perbedaan respons individu, atau heterogenitas kondisi populasi (Rosner, 2016; Motulsky, 2018; Sullivan, 2022).

Pembelajaran statistik deskriptif juga perlu dikaitkan dengan konsep distribusi data. Distribusi data menggambarkan pola sebaran nilai dalam suatu variabel. Data dapat berdistribusi normal, menceng ke kanan, menceng ke kiri, memiliki puncak tunggal, atau memiliki lebih dari satu puncak. Pemahaman distribusi penting karena banyak metode statistik memiliki asumsi tentang bentuk distribusi data. Indikator capaian pembelajaran pada bagian ini mencakup kemampuan membaca histogram, boxplot, nilai skewness, nilai kurtosis, serta mengenali nilai ekstrem yang dapat memengaruhi hasil analisis (Field, 2018; Daniel & Cross, 2018; Peacock & Peacock, 2020).

Capaian pembelajaran berikutnya adalah pemahaman dasar tentang probabilitas. Probabilitas menjelaskan peluang terjadinya suatu peristiwa dalam kondisi tertentu. Dalam kesehatan, probabilitas digunakan untuk memahami risiko penyakit, peluang kesembuhan, sensitivitas dan spesifisitas tes diagnostik, serta kemungkinan kejadian efek samping. Pembaca perlu memahami bahwa probabilitas bukan kepastian, melainkan ukuran peluang berdasarkan data atau model tertentu. Pemahaman ini penting agar pembaca tidak menafsirkan angka risiko secara berlebihan atau terlalu sederhana (Kestenbaum, 2019; Rothman *et al.*, 2021; Gordis, 2019).

Distribusi probabilitas juga menjadi dasar penting untuk memahami analisis statistik lanjutan. Distribusi binomial digunakan untuk data dengan dua kemungkinan hasil, seperti sakit dan tidak sakit. Distribusi Poisson sering digunakan untuk menghitung jumlah kejadian dalam waktu atau ruang tertentu, seperti jumlah kasus penyakit dalam suatu wilayah. Distribusi normal banyak digunakan dalam analisis kesehatan karena banyak variabel biologis mendekati pola sebaran normal, meskipun tetap perlu diuji secara empiris. Indikator capaian pembelajaran pada bagian ini meliputi kemampuan membedakan distribusi binomial, Poisson, dan normal, serta memahami penggunaannya dalam konteks data kesehatan (Rosner, 2016; Daniel & Cross, 2018; Sullivan, 2022).

Pada akhir bagian ini, pembaca diharapkan mampu menjelaskan hubungan antara statistik deskriptif dan distribusi probabilitas. Statistik deskriptif membantu menggambarkan data yang diperoleh, sedangkan distribusi probabilitas membantu memahami pola peluang yang mendasari data tersebut. Keduanya menjadi jembatan menuju statistik inferensial. Pembaca yang memahami ukuran pemusatan, ukuran penyebaran, dan distribusi data akan lebih siap memahami interval kepercayaan, p-value, dan uji hipotesis. Dengan demikian, capaian pembelajaran pada bagian ini menjadi fondasi bagi analisis yang lebih kompleks dalam penelitian kesehatan dan epidemiologi (Motulsky, 2018; Peacock & Peacock, 2020; Sullivan, 2022).

1.8 Capaian Pembelajaran Statistik Inferensial dan Uji Hipotesis

Capaian pembelajaran statistik inferensial diarahkan agar pembaca mampu menarik kesimpulan tentang populasi berdasarkan data sampel. Dalam penelitian kesehatan, peneliti jarang dapat mengukur seluruh populasi karena

keterbatasan waktu, biaya, tenaga, dan akses. Oleh karena itu, sampel digunakan untuk memperkirakan keadaan populasi. Statistik inferensial membantu peneliti menilai seberapa kuat hasil sampel dapat digunakan untuk menjelaskan populasi yang lebih luas. Pembaca perlu memahami bahwa inferensi selalu mengandung ketidakpastian, sehingga hasil penelitian harus ditafsirkan dengan hati-hati (Daniel & Cross, 2018; Rosner, 2016; Sullivan, 2022).

Landasan teoritis statistik inferensial dapat dikaitkan dengan teori estimasi dan pengambilan sampel. Estimasi digunakan untuk memperkirakan parameter populasi, seperti rata-rata, proporsi, risiko, atau perbedaan antar kelompok. Estimasi dapat disajikan dalam bentuk estimasi titik dan estimasi interval. Interval kepercayaan memberikan rentang nilai yang masuk akal bagi parameter populasi berdasarkan data sampel. Indikator capaian pembelajaran pada bagian ini mencakup kemampuan menjelaskan makna estimasi titik, interval kepercayaan, tingkat kepercayaan, dan kesalahan baku (standard error) (Peacock & Peacock, 2020; Petrie & Sabin, 2020; Motulsky, 2018).

Uji hipotesis merupakan bagian penting dari statistik inferensial. Ronald A. Fisher memperkenalkan penggunaan nilai signifikansi dan p-value sebagai alat untuk menilai kekuatan bukti terhadap hipotesis nol. Dalam pendekatan Fisher, p-value menunjukkan seberapa tidak biasa data yang diperoleh jika hipotesis nol benar. Namun, p-value tidak menunjukkan besar pengaruh, tidak membuktikan kebenaran hipotesis, dan tidak menggantikan pertimbangan ilmiah. Karena itu, pembaca perlu memahami p-value sebagai salah satu bagian dari interpretasi, bukan sebagai satu-satunya dasar keputusan penelitian (Fisher, 1935/1990; Motulsky, 2018; Wasserstein & Lazar, 2016).

Jerzy Neyman dan Egon Pearson mengembangkan kerangka pengujian hipotesis yang menekankan keputusan statistik, kesalahan tipe I, kesalahan tipe II, dan kekuatan uji.

Kesalahan tipe I terjadi ketika peneliti menolak hipotesis nol yang sebenarnya benar. Kesalahan tipe II terjadi ketika peneliti gagal menolak hipotesis nol yang sebenarnya salah. Power atau kekuatan uji menunjukkan peluang penelitian menemukan perbedaan atau hubungan yang benar-benar ada. Indikator capaian pembelajaran pada bagian ini mencakup kemampuan merumuskan hipotesis nol dan alternatif, menentukan tingkat signifikansi, menjelaskan kesalahan tipe I dan tipe II, serta memahami makna power dalam rancangan penelitian (Neyman & Pearson, 1933; Hulley *et al.*, 2022; Portney, 2020).

Capaian pembelajaran uji hipotesis juga mencakup kemampuan memilih uji statistik yang sesuai. Pemilihan uji tidak boleh hanya mengikuti kebiasaan, tetapi harus didasarkan pada tujuan analisis, jenis data, skala pengukuran, jumlah kelompok, distribusi data, dan hubungan antarvariabel. Uji parametrik digunakan jika asumsi tertentu terpenuhi, seperti data numerik dan distribusi yang sesuai. Uji nonparametrik digunakan ketika data tidak memenuhi asumsi parametrik atau ketika skala data lebih terbatas. Indikator penguasaan pada bagian ini meliputi kemampuan memilih uji-t, chi-square, Mann-Whitney, Wilcoxon, ANOVA, Kruskal-Wallis, korelasi, dan regresi sesuai kebutuhan analisis (Field, 2018; Daniel & Cross, 2018; Sullivan, 2022).

Interpretasi hasil uji hipotesis menjadi capaian yang sangat penting dalam penelitian kesehatan. Pembaca perlu mampu membaca hasil analisis secara utuh, termasuk p-value, interval kepercayaan, ukuran efek, arah hubungan, dan relevansi klinis atau kesehatan masyarakat. Hasil yang signifikan secara statistik belum tentu penting secara praktis. Sebaliknya, hasil yang tidak signifikan dapat tetap memiliki makna jika ukuran sampel kecil, variabilitas besar, atau terdapat keterbatasan desain penelitian. Oleh karena itu, capaian pembelajaran pada bagian ini menekankan bahwa keputusan ilmiah harus mempertimbangkan statistik, desain

penelitian, konteks masalah, dan kualitas data (Rothman *et al.*, 2021; Hernán & Robins, 2020; Spiegelhalter, 2019).

Pada akhir bagian ini, pembaca diharapkan mampu memahami statistik inferensial sebagai alat untuk membuat kesimpulan berbasis bukti. Pembaca perlu mampu menjelaskan logika inferensi, merumuskan hipotesis, memilih uji statistik, membaca output, dan menyusun interpretasi yang tidak berlebihan. Capaian ini menjadi dasar untuk mempelajari uji parametrik, uji nonparametrik, korelasi, regresi, ANOVA, dan analisis epidemiologi pada bab-bab berikutnya. Penguasaan statistik inferensial juga membantu pembaca membaca artikel ilmiah secara kritis. Dengan kemampuan ini, pembaca dapat menilai apakah kesimpulan penelitian kesehatan didukung oleh data, metode, dan interpretasi yang memadai (Hulley *et al.*, 2022; Creswell & Creswell, 2023; Sullivan, 2022).

1.9 Capaian Pembelajaran Analisis Hubungan, Prediksi, ANOVA, dan Uji Lanjut

Capaian pembelajaran pada bagian analisis hubungan diarahkan agar pembaca mampu memahami keterkaitan antarvariabel dalam data kesehatan. Banyak pertanyaan penelitian kesehatan tidak hanya menanyakan “berapa besar masalah”, tetapi juga “faktor apa yang berhubungan dengan masalah tersebut”. Analisis hubungan membantu peneliti menilai apakah perubahan pada satu variabel berkaitan dengan perubahan pada variabel lain. Contohnya, peneliti dapat menilai hubungan antara usia dan tekanan darah, tingkat pendidikan dan perilaku merokok, atau kepadatan hunian dan kejadian infeksi saluran pernapasan. Pembaca diharapkan mampu membedakan hubungan statistik, hubungan praktis, dan hubungan kausal agar tidak menarik kesimpulan secara berlebihan dari data yang tersedia (Daniel & Cross, 2018; Sullivan, 2022; Rothman *et al.*, 2021).

Landasan awal analisis hubungan dapat dikaitkan dengan perkembangan teori korelasi dan regresi yang dipelopori oleh Francis Galton dan Karl Pearson. Galton memperkenalkan gagasan regresi dalam kajian variasi biologis, sedangkan Pearson mengembangkan koefisien korelasi untuk mengukur kekuatan dan arah hubungan linear antarvariabel. Dalam pembelajaran biostatistika, konsep ini penting karena membantu pembaca memahami bahwa hubungan antarvariabel dapat diukur secara kuantitatif. Indikator capaian pembelajaran pada bagian ini mencakup kemampuan menjelaskan arah hubungan positif dan negatif, kekuatan hubungan, serta keterbatasan korelasi. Pembaca juga perlu memahami bahwa korelasi tidak membuktikan sebab-akibat, terutama jika desain penelitian tidak mendukung kesimpulan kausal (Stanton, 2001; Agresti, 2018; Motulsky, 2018).

Capaian pembelajaran analisis prediksi diarahkan agar pembaca mampu memahami fungsi regresi dalam penelitian kesehatan. Regresi digunakan untuk menjelaskan atau memperkirakan nilai suatu variabel berdasarkan satu atau lebih variabel lain. Regresi linear digunakan ketika variabel terikat berbentuk numerik, seperti tekanan darah atau kadar kolesterol. Regresi logistik digunakan ketika variabel terikat berbentuk dikotomi, seperti sakit dan tidak sakit, sembuh dan tidak sembuh, atau patuh dan tidak patuh. Indikator capaian pembelajaran mencakup kemampuan membedakan regresi linear dan logistik, membaca koefisien regresi, memahami nilai p , interval kepercayaan, serta menafsirkan hasil dalam konteks masalah kesehatan (Gelman *et al.*, 2020; Field, 2018; Sullivan, 2022).

Analisis prediksi juga menuntut pembaca memahami konsep variabel perancu, interaksi, dan penyesuaian model. Dalam kesehatan masyarakat, hubungan antara paparan dan luaran sering dipengaruhi oleh banyak faktor lain. Misalnya, hubungan antara aktivitas fisik dan hipertensi dapat

dipengaruhi oleh umur, indeks massa tubuh, pola makan, dan status merokok. Regresi membantu peneliti menilai hubungan utama setelah mempertimbangkan faktor lain dalam model. Namun, pembaca harus memahami bahwa model statistik yang baik tidak hanya ditentukan oleh hasil signifikan, tetapi juga oleh dasar teori, desain penelitian, kualitas data, dan kelayakan interpretasi (Hernán & Robins, 2020; Rothman *et al.*, 2021; Hulley *et al.*, 2022).

Capaian pembelajaran pada bagian ANOVA diarahkan agar pembaca mampu membandingkan rata-rata lebih dari dua kelompok secara tepat. Analisis varians atau ANOVA dikembangkan oleh Ronald A. Fisher sebagai bagian dari rancangan eksperimen untuk menilai variasi antar kelompok dan variasi dalam kelompok. Dalam penelitian kesehatan, ANOVA dapat digunakan untuk membandingkan rata-rata tekanan darah pada tiga kelompok umur, kadar hemoglobin berdasarkan status gizi, atau skor pengetahuan setelah beberapa model edukasi. Pembaca perlu memahami bahwa ANOVA tidak hanya melihat perbedaan rata-rata, tetapi juga menilai apakah variasi antar kelompok lebih besar daripada variasi di dalam kelompok. Indikator capaian pembelajaran mencakup kemampuan menjelaskan tujuan ANOVA, asumsi normalitas, homogenitas varians, independensi data, serta interpretasi nilai F dan p-value (Fisher, 1935/1990; Daniel & Cross, 2018; Field, 2018).

Uji lanjut atau post hoc test menjadi bagian penting setelah ANOVA menunjukkan hasil yang bermakna. ANOVA dapat menjawab apakah terdapat perbedaan rata-rata di antara beberapa kelompok, tetapi tidak langsung menunjukkan kelompok mana yang berbeda. John Tukey mengembangkan pendekatan *honestly significant difference* untuk membandingkan pasangan kelompok dengan pengendalian kesalahan ganda. Dalam praktik penelitian kesehatan, uji lanjut membantu peneliti menafsirkan hasil dengan lebih rinci dan tidak hanya berhenti pada pernyataan

“ada perbedaan”. Indikator capaian pembelajaran pada bagian ini mencakup kemampuan menjelaskan kapan uji lanjut diperlukan, memilih jenis uji lanjut yang sesuai, dan menafsirkan hasil perbandingan pasangan kelompok secara hati-hati (Tukey, 1949; Field, 2018; Peacock & Peacock, 2020).

Pada akhir bagian ini, pembaca diharapkan mampu menghubungkan analisis hubungan, prediksi, ANOVA, dan uji lanjut sebagai alat untuk menjawab pertanyaan penelitian yang lebih kompleks. Korelasi membantu melihat hubungan awal, regresi membantu menjelaskan dan memprediksi hubungan antarvariabel, sedangkan ANOVA membantu membandingkan lebih dari dua kelompok. Uji lanjut memperjelas letak perbedaan setelah analisis utama menunjukkan hasil yang bermakna. Pembaca juga harus mampu menyadari bahwa setiap metode memiliki syarat, batasan, dan konteks penggunaan. Capaian ini menjadi dasar penting untuk membaca artikel ilmiah, menyusun hasil penelitian, serta menggunakan perangkat lunak statistik secara lebih kritis (Motulsky, 2018; Sullivan, 2022; Wickham *et al.*, 2023).

1.10 Capaian Pembelajaran Penerapan Biostatistika dalam Epidemiologi, Kesehatan Masyarakat, dan Software Statistik

Capaian pembelajaran pada bagian penerapan biostatistika diarahkan agar pembaca mampu menggunakan konsep statistik untuk menjawab masalah epidemiologi dan kesehatan masyarakat. Epidemiologi membutuhkan biostatistika untuk mengukur frekuensi penyakit, menilai hubungan paparan dan luaran, serta mengevaluasi dampak intervensi. Dalam konteks kesehatan masyarakat, biostatistika membantu menjelaskan besarnya masalah, pola distribusi, kelompok rentan, dan faktor determinan. Pembaca

diharapkan mampu menghubungkan ukuran statistik dengan masalah nyata seperti penyakit menular, penyakit tidak menular, gizi, kesehatan lingkungan, kesehatan kerja, dan mutu pelayanan kesehatan. Dengan kemampuan tersebut, pembaca dapat memahami bahwa biostatistika merupakan alat utama dalam membangun bukti untuk tindakan kesehatan masyarakat (Friis & Sellers, 2021; Detels *et al.*, 2021; Gordis, 2019).

Capaian pembelajaran epidemiologi mencakup kemampuan memahami ukuran frekuensi penyakit. Ukuran tersebut meliputi insidensi, prevalensi, angka kematian, case fatality rate, dan berbagai ukuran beban penyakit. Insidensi menunjukkan kejadian baru dalam periode tertentu, sedangkan prevalensi menunjukkan jumlah kasus yang ada pada satu waktu atau periode. Perbedaan ini penting karena strategi program untuk penyakit akut dapat berbeda dari strategi untuk penyakit kronis. Indikator pembelajaran pada bagian ini mencakup kemampuan memilih ukuran frekuensi yang sesuai dengan jenis masalah kesehatan dan tujuan analisis (Kestenbaum, 2019; Rothman *et al.*, 2021; Celentano & Szklo, 2019).

Capaian pembelajaran berikutnya adalah kemampuan memahami ukuran asosiasi dalam epidemiologi. Ukuran asosiasi digunakan untuk menilai kekuatan hubungan antara paparan dan kejadian penyakit. Ukuran yang sering digunakan meliputi risk ratio, odds ratio, prevalence ratio, attributable risk, dan population attributable fraction. Pemilihan ukuran asosiasi bergantung pada desain penelitian, jenis data, dan tujuan analisis. Pembaca harus mampu menafsirkan ukuran asosiasi secara tepat, termasuk memahami bahwa nilai asosiasi yang tinggi tetap perlu dinilai bersama interval kepercayaan, potensi bias, perancu, dan kualitas desain penelitian (Rothman *et al.*, 2021; Hernán & Robins, 2020; Szklo & Nieto, 2019).

Penerapan biostatistika dalam kesehatan masyarakat juga terkait dengan pengambilan keputusan berbasis bukti. Pendekatan evidence-based public health menekankan bahwa program dan kebijakan seharusnya disusun berdasarkan bukti ilmiah, data lokal, kebutuhan populasi, sumber daya, dan nilai masyarakat. Biostatistika membantu menilai apakah suatu intervensi efektif, apakah capaian indikator membaik, dan apakah perubahan yang terjadi cukup kuat untuk dijadikan dasar keputusan. Dalam evaluasi program, pembaca perlu memahami indikator input, proses, output, outcome, dan dampak. Indikator capaian pembelajaran mencakup kemampuan membaca hasil evaluasi program, menilai efektivitas intervensi, dan menjelaskan hasil kepada pemangku kepentingan secara sederhana (Brownson *et al.*, 2018; Detels *et al.*, 2021; WHO, 2023).

Capaian pembelajaran pada bagian software statistik diarahkan agar pembaca mampu menggunakan perangkat lunak sebagai alat bantu analisis, bukan sebagai pengganti pemahaman konsep. SPSS banyak digunakan karena tampilannya mudah dipahami dan cocok untuk pembelajar pemula. R memiliki keunggulan pada fleksibilitas, reproduksibilitas, visualisasi data, dan analisis lanjutan. Excel berguna untuk manajemen data awal, perhitungan sederhana, dan penyajian data, tetapi penggunaannya untuk analisis statistik perlu dilakukan dengan hati-hati. Pembaca diharapkan mampu memahami fungsi dasar masing-masing software, memilih alat yang sesuai, dan tetap memeriksa kebenaran input, proses, serta output analisis (Field, 2018; Harris *et al.*, 2021; Wickham *et al.*, 2023).

Penerapan software statistik juga harus mengikuti prinsip reproduksibilitas dan dokumentasi analisis. Reproduksi berarti hasil analisis dapat diperiksa ulang oleh peneliti lain dengan data dan prosedur yang sama. Dalam penelitian kesehatan, prinsip ini penting karena keputusan

berbasis data harus dapat diaudit dan dipertanggungjawabkan. Pembaca perlu membiasakan diri menyimpan kode variabel, sintaks, catatan pembersihan data, versi data, dan keputusan analisis. Indikator capaian pembelajaran pada bagian ini mencakup kemampuan membuat dokumentasi analisis, menyimpan output secara rapi, dan menjelaskan langkah analisis dari data mentah sampai hasil akhir (Harris *et al.*, 2021; Wickham *et al.*, 2023; Wilkinson *et al.*, 2016).

Pada akhir bagian ini, pembaca diharapkan mampu menerapkan biostatistika dalam penelitian dan praktik kesehatan masyarakat secara bertanggung jawab. Penerapan tersebut mencakup kemampuan menghitung ukuran epidemiologi, memilih uji statistik, menggunakan software, membaca output, dan menyusun interpretasi hasil. Pembaca juga perlu memahami aspek etika, terutama ketika menggunakan data individu, data rekam medis, atau data surveilans. Data kesehatan mengandung informasi sensitif sehingga analisis harus menjaga kerahasiaan, keamanan, dan kepentingan masyarakat. Dengan capaian ini, pembaca dapat menggunakan biostatistika sebagai alat untuk memperkuat penelitian, pelayanan, kebijakan, dan program kesehatan masyarakat (WHO, 2020; Hulley *et al.*, 2022; Creswell & Creswell, 2023).

1.11 Integrasi Capaian Pembelajaran Bab 1 dengan Struktur Keseluruhan Buku

Bab 1 berfungsi sebagai pintu masuk untuk memahami keseluruhan isi buku Biostatistika: Analisis Data untuk Kesehatan dan Penelitian. Capaian pembelajaran dalam bab ini tidak berdiri sendiri, tetapi menjadi dasar untuk mempelajari bab-bab berikutnya secara bertahap. Pembaca diarahkan untuk memahami biostatistika sebagai ilmu yang menghubungkan data, metode analisis, dan keputusan

kesehatan. Bab ini juga menegaskan bahwa pembelajaran biostatistika harus dimulai dari pemahaman masalah, bukan langsung dari rumus atau software. Dengan demikian, Bab 1 memberi arah konseptual agar pembaca memahami mengapa setiap bab dalam buku ini penting untuk dikuasai (Daniel & Cross, 2018; Sullivan, 2022; Biggs *et al.*, 2022).

Integrasi capaian pembelajaran Bab 1 dengan Bab 2 dan Bab 3 terlihat pada pemahaman dasar biostatistika, jenis data, dan skala pengukuran. Bab 2 membangun landasan konseptual tentang fungsi biostatistika dalam ilmu kesehatan, sedangkan Bab 3 memperkuat kemampuan mengenali jenis data dan skala pengukuran. Capaian pada bagian ini penting karena semua analisis statistik bergantung pada karakteristik data. Pembaca yang keliru memahami jenis data akan mudah salah memilih metode analisis. Oleh karena itu, capaian Bab 1 menjadi pengantar yang mengarahkan pembaca untuk membaca data secara lebih cermat sejak awal (Agresti, 2018; Rosner, 2016; Peacock & Peacock, 2020).

Integrasi berikutnya terlihat pada Bab 4 dan Bab 5 yang membahas pengumpulan, manajemen, dan penyajian data kesehatan. Bab 1 menegaskan bahwa kualitas analisis statistik sangat ditentukan oleh kualitas data. Bab 4 kemudian memperluas pembahasan tentang sumber data, instrumen, validitas, reliabilitas, pengkodean, pembersihan data, dan keamanan data. Bab 5 melanjutkan pembahasan tersebut dengan keterampilan menyajikan data melalui tabel, grafik, dan diagram. Dengan alur ini, pembaca memahami bahwa data harus dikumpulkan dengan benar, dikelola secara rapi, dan disajikan secara jujur sebelum dianalisis lebih lanjut (Polit & Beck, 2021; Hulley *et al.*, 2022; Wickham *et al.*, 2023).

Bab 6 dan Bab 7 terhubung dengan capaian pembelajaran tentang statistik deskriptif dan distribusi probabilitas. Bab 6 membantu pembaca memahami ukuran pemusatan dan penyebaran data, sedangkan Bab 7

memperkenalkan konsep probabilitas dan distribusi data. Kedua bab ini menjadi fondasi sebelum pembaca mempelajari statistik inferensial. Tanpa pemahaman tentang rata-rata, median, simpangan baku, distribusi normal, binomial, dan Poisson, pembaca akan kesulitan memahami interval kepercayaan dan uji hipotesis. Karena itu, Bab 1 menempatkan statistik deskriptif dan probabilitas sebagai jembatan antara data mentah dan penarikan kesimpulan ilmiah (Motulsky, 2018; Petrie & Sabin, 2020; Kestenbaum, 2019).

Bab 8 sampai Bab 11 terhubung dengan capaian pembelajaran tentang inferensi, uji hipotesis, korelasi, regresi, ANOVA, dan uji lanjut. Bab 8 memperkenalkan logika statistik inferensial, sedangkan Bab 9 membahas uji hipotesis parametrik dan nonparametrik. Bab 10 kemudian memperluas kemampuan pembaca dalam menganalisis hubungan dan prediksi melalui korelasi dan regresi. Bab 11 melengkapi kemampuan tersebut dengan analisis perbedaan lebih dari dua kelompok melalui ANOVA dan uji lanjut. Integrasi ini menunjukkan bahwa pembelajaran statistik harus bergerak dari konsep dasar menuju kemampuan analisis yang lebih kompleks secara berurutan (Field, 2018; Gelman *et al.*, 2020; Daniel & Cross, 2018).

Bab 12 dan Bab 13 menjadi tahap penerapan dari seluruh capaian pembelajaran dalam buku ini. Bab 12 menghubungkan biostatistika dengan epidemiologi dan kesehatan masyarakat, termasuk ukuran frekuensi penyakit, ukuran asosiasi, dan evaluasi program. Bab 13 memperkuat keterampilan teknis melalui penggunaan SPSS, R, dan Excel dalam analisis data kesehatan. Kedua bab ini menunjukkan bahwa biostatistika tidak selesai pada perhitungan manual, tetapi harus diterapkan dalam penelitian dan pengambilan keputusan kesehatan. Dengan demikian, integrasi capaian pembelajaran Bab 1 dengan seluruh struktur buku membentuk alur belajar yang utuh, mulai dari memahami

data sampai menggunakan hasil analisis untuk mendukung tindakan kesehatan masyarakat yang berbasis bukti (Friis & Sellers, 2021; Brownson *et al.*, 2018; Harris *et al.*, 2021).

Pada akhirnya, Bab 1 memberi fondasi untuk membangun literasi biostatistika yang praktis, kritis, dan etis. Pembaca diharapkan tidak hanya mampu menjalankan uji statistik, tetapi juga memahami alasan pemilihan metode, batasan data, dan makna hasil analisis. Integrasi capaian pembelajaran dalam buku ini mendorong pembaca untuk berpikir sistematis sejak tahap perumusan masalah sampai interpretasi hasil. Keterampilan tersebut penting bagi mahasiswa, dosen, peneliti, tenaga kesehatan, dan pengelola program yang bekerja dengan data kesehatan. Dengan penguasaan bertahap dari Bab 1 sampai Bab 13, pembaca dapat menggunakan biostatistika sebagai dasar pengambilan keputusan ilmiah dalam penelitian dan praktik kesehatan masyarakat (Spiegelhalter, 2019; WHO, 2020; Creswell & Creswell, 2023).

DAFTAR PUSTAKA

- Agresti, A. (2018). *Statistical methods for the social sciences* (5th ed.). Pearson.
- Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Rath, J., & Wittrock, M. C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
- Biggs, J., Tang, C., & Kennedy, G. (2022). *Teaching for quality learning at university* (5th ed.). Open University Press.
- Brownson, R. C., Baker, E. A., Deshpande, A. D., & Gillespie, K. N. (2018). *Evidence-based public health* (3rd ed.). Oxford University Press.
- Celentano, D. D., & Szklo, M. (2019). *Gordis epidemiology* (6th ed.). Elsevier.
- Creswell, J. W., & Creswell, J. D. (2023). *Research design: Qualitative, quantitative, and mixed methods approaches* (6th ed.). SAGE Publications.
- Daniel, W. W., & Cross, C. L. (2018). *Biostatistics: A foundation for analysis in the health sciences* (11th ed.). Wiley.
- Dawson, B., & Trapp, R. G. (2020). *Basic & clinical biostatistics* (5th ed.). McGraw-Hill Education.
- Detels, R., Karim, Q. A., Baum, F., Li, L., & Leyland, A. H. (Eds.). (2021). *Oxford textbook of global public health* (7th ed.). Oxford University Press.
- Donabedian, A. (1988). The quality of care: How can it be assessed? *JAMA*, 260(12), 1743–1748. <https://doi.org/10.1001/jama.1988.03410120089003>

- Field, A. (2018). *Discovering statistics using IBM SPSS Statistics* (5th ed.). SAGE Publications.
- Fisher, R. A. (1990). *The design of experiments* (8th ed.). Hafner Press. (Original work published 1935)
- Friis, R. H., & Sellers, T. A. (2021). *Epidemiology for public health practice* (6th ed.). Jones & Bartlett Learning.
- Gelman, A., Hill, J., & Vehtari, A. (2020). *Regression and other stories*. Cambridge University Press.
- Gigerenzer, G. (2015). *Risk savvy: How to make good decisions*. Penguin Books.
- Gordis, L. (2019). *Epidemiology* (6th ed.). Elsevier.
- Harris, J. K., Beatty, K. E., Barbero, C., Howard, A. F., Cheskin, R. A., Shapiro, R. M., & Mays, G. P. (2021). *Reproducible medical research with R*. CRC Press.
- Hernán, M. A., & Robins, J. M. (2020). *Causal inference: What if*. Chapman & Hall/CRC.
- Hill, A. B. (1965). The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine*, 58(5), 295–300.
- Hulley, S. B., Cummings, S. R., Browner, W. S., Grady, D. G., & Newman, T. B. (2022). *Designing clinical research* (5th ed.). Wolters Kluwer.
- Joshi, M. S., Ransom, E. R., Nash, D. B., & Ransom, S. B. (Eds.). (2014). *The healthcare quality book: Vision, strategy, and tools* (3rd ed.). Health Administration Press.
- Kestenbaum, B. (2019). *Epidemiology and biostatistics: An introduction to clinical research* (2nd ed.). Springer.
- McLaughlin, C. P., & Olson, J. R. (2023). *Healthcare operations management* (4th ed.). AUPHA Press.

- Motulsky, H. (2018). *Intuitive biostatistics: A nonmathematical guide to statistical thinking* (4th ed.). Oxford University Press.
- Neyman, J., & Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A*, 231, 289–337.
- Peacock, J. L., & Peacock, P. J. (2020). *Oxford handbook of medical statistics* (2nd ed.). Oxford University Press.
- Petrie, A., & Sabin, C. (2020). *Medical statistics at a glance* (4th ed.). Wiley-Blackwell.
- Polit, D. F., & Beck, C. T. (2021). *Nursing research: Generating and assessing evidence for nursing practice* (11th ed.). Wolters Kluwer.
- Portney, L. G. (2020). *Foundations of clinical research: Applications to evidence-based practice* (4th ed.). F. A. Davis.
- Rosner, B. (2016). *Fundamentals of biostatistics* (8th ed.). Cengage Learning.
- Rothman, K. J., Lash, T. L., VanderWeele, T. J., & Haneuse, S. (2021). *Modern epidemiology* (4th ed.). Wolters Kluwer.
- Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn't. *BMJ*, 312(7023), 71–72. <https://doi.org/10.1136/bmj.312.7023.71>
- Spady, W. G. (1994). *Outcome-based education: Critical issues and answers*. American Association of School Administrators.
- Spiegelhalter, D. (2019). *The art of statistics: Learning from data*. Pelican.

- Stanton, J. M. (2001). Galton, Pearson, and the peas: A brief history of linear regression for statistics instructors. *Journal of Statistics Education*, 9(3). <https://doi.org/10.1080/10691898.2001.11910537>
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103(2684), 677–680. <https://doi.org/10.1126/science.103.2684.677>
- Sullivan, L. M. (2022). *Essentials of biostatistics in public health* (4th ed.). Jones & Bartlett Learning.
- Szklo, M., & Nieto, F. J. (2019). *Epidemiology: Beyond the basics* (4th ed.). Jones & Bartlett Learning.
- Tukey, J. W. (1949). Comparing individual means in the analysis of variance. *Biometrics*, 5(2), 99–114. <https://doi.org/10.2307/3001913>
- Vittinghoff, E., Glidden, D. V., Shiboski, S. C., & McCulloch, C. E. (2012). *Regression methods in biostatistics: Linear, logistic, survival, and repeated measures models* (2nd ed.). Springer.
- Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on p-values: Context, process, and purpose. *The American Statistician*, 70(2), 129–133. <https://doi.org/10.1080/00031305.2016.1154108>
- Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). *R for data science: Import, tidy, transform, visualize, and model data* (2nd ed.). O'Reilly Media.

- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3, Article 160018. <https://doi.org/10.1038/sdata.2016.18>
- World Health Organization. (2020). *SCORE for health data technical package: Global report on health data systems and capacity, 2020*. World Health Organization.
- World Health Organization. (2023). *World health statistics 2023: Monitoring health for the SDGs*. World Health Organization.

BAB 2

KONSEP DASAR BIOSTATISTIKA

Oleh Rahmawati

Biostatistika merupakan cabang ilmu statistika yang secara khusus dikembangkan untuk menjawab permasalahan di bidang kesehatan, kedokteran, biologi, dan ilmu-ilmu hayati lainnya. Perkembangan ilmu kesehatan modern yang berbasis bukti (*evidence-based health sciences*) menuntut setiap keputusan klinis, kebijakan kesehatan, dan kesimpulan ilmiah didasarkan pada analisis data yang valid dan dapat dipertanggungjawabkan secara ilmiah. Dalam konteks ini, biostatistika menjadi fondasi utama yang tidak terpisahkan dari seluruh proses penelitian kesehatan (Dahlan, 2014; Rosner, 2016).

Berbagai ahli telah mengemukakan definisi biostatistika dengan penekanan yang saling melengkapi. Menurut Rosner (2016), biostatistika adalah penerapan metode statistik untuk merancang penelitian, menganalisis data, dan menarik kesimpulan dalam bidang biologi dan kesehatan. Definisi ini menekankan peran biostatistika sebagai alat metodologis dalam keseluruhan proses penelitian. Gordis (2014) mendefinisikan biostatistika sebagai seperangkat metode kuantitatif yang digunakan untuk menggambarkan distribusi penyakit, mengevaluasi hubungan sebab-akibat, serta menilai efektivitas intervensi kesehatan. Definisi ini menempatkan biostatistika sebagai komponen kunci dalam epidemiologi dan kesehatan masyarakat.

Menurut Pagano dan Gauvreau (2018), biostatistika adalah ilmu yang mempelajari cara pengumpulan, pengorganisasian, analisis, dan interpretasi data yang berasal dari sistem biologis. Penekanan utama dalam definisi ini adalah pada proses analisis dan interpretasi data biologis secara sistematis. Dahlan (2014) menyatakan bahwa biostatistika merupakan ilmu yang memanfaatkan data numerik untuk mendukung pengambilan keputusan klinis dan penelitian kesehatan secara objektif. Definisi ini menegaskan fungsi praktis biostatistika dalam praktik kedokteran dan kesehatan.

Atkins dan de Paula (2014) memandang biostatistika sebagai bagian dari statistika terapan yang berperan dalam menjelaskan fenomena biologis yang bersifat kompleks dan bervariasi. Pendekatan ini menekankan pentingnya pemodelan statistik dalam memahami sistem kehidupan. Bland (2015) menjelaskan biostatistika sebagai alat untuk mengkuantifikasi ketidakpastian dalam penelitian medis dan kesehatan. Definisi ini menyoroti peran biostatistika dalam pengujian hipotesis dan estimasi parameter.

Zar (2010) mendefinisikan biostatistika sebagai penerapan metode statistik untuk menganalisis data yang berasal dari sistem biologis dan ekologi, dengan tujuan menarik inferensi yang valid mengenai proses kehidupan. Kirkwood dan Sterne (2003) menyatakan bahwa biostatistika adalah penggunaan metode statistik untuk merancang studi medis, menganalisis data kesehatan, serta menafsirkan hasil penelitian secara kritis dan objektif.

Daniel dan Cross (2018) mendefinisikan biostatistika sebagai cabang statistika terapan yang berfokus pada pengolahan dan analisis data dalam penelitian biomedis dan kesehatan masyarakat. Altman (1991) memandang biostatistika sebagai alat ilmiah untuk memastikan bahwa kesimpulan penelitian medis didasarkan pada bukti kuantitatif yang sah dan bebas dari bias metodologis.

Dari berbagai definisi tersebut, dapat disimpulkan bahwa biostatistika merupakan disiplin ilmu yang berfokus pada penerapan prinsip dan metode statistika untuk mendukung penelitian, pengambilan keputusan, dan pengembangan ilmu di bidang kesehatan dan biologi. Biostatistika tidak hanya berfungsi sebagai alat analisis data, tetapi juga sebagai kerangka berpikir ilmiah dalam menghadapi ketidakpastian dan variasi data kesehatan.

Biostatistika tidak hanya berperan pada tahap analisis data, tetapi juga sejak tahap awal penelitian, yaitu perumusan masalah, penentuan desain penelitian, pemilihan variabel, hingga penentuan besar sampel. Kesalahan pemahaman konsep dasar biostatistika pada tahap awal penelitian dapat menyebabkan hasil penelitian menjadi bias, tidak presisi, dan sulit digeneralisasikan. Oleh karena itu, penguasaan konsep dasar biostatistika merupakan kompetensi wajib bagi mahasiswa dan peneliti di bidang kesehatan (Pagano & Gauvreau, 2018).

Dalam praktik kesehatan masyarakat dan klinis, biostatistika digunakan untuk menggambarkan masalah kesehatan, mengidentifikasi faktor risiko penyakit, serta mengevaluasi efektivitas intervensi kesehatan. Indikator kesehatan seperti angka kesakitan, angka kematian, prevalens, dan insidens penyakit seluruhnya dihitung dan dianalisis menggunakan prinsip biostatistika. Tanpa pendekatan statistik yang tepat, data kesehatan hanya akan menjadi kumpulan angka tanpa makna yang jelas (Gordis, 2014).

Selain itu, kemajuan teknologi informasi dan sistem pencatatan kesehatan elektronik (*electronic health records*) telah menghasilkan data kesehatan dalam jumlah besar. Fenomena ini menuntut tenaga kesehatan dan peneliti untuk memiliki kemampuan literasi statistik agar mampu membaca, menganalisis, dan menginterpretasikan data secara kritis. Biostatistika berperan penting dalam menjembatani data

mentah menjadi informasi dan pengetahuan yang bermanfaat bagi peningkatan derajat kesehatan masyarakat (Rosner, 2016).

Bab ini disusun untuk memberikan pemahaman yang komprehensif mengenai konsep dasar biostatistika sebagai landasan bagi pembahasan metode statistik lanjutan pada bab-bab berikutnya. Materi dalam bab ini meliputi pengertian dan ruang lingkup biostatistika, jenis dan skala data, konsep populasi dan sampel, variabel penelitian, serta prinsip dasar penyajian dan peringkasan data. Dengan penguasaan materi pada bab ini, diharapkan mahasiswa mampu memahami dan menerapkan analisis data secara tepat dalam penelitian kesehatan dan praktik profesional.

2.1 Konsep Biostatistika: Pengertian dan Ruang Lingkup

2.1.1 Konsep Dasar Biostatistika

Biostatistika merupakan cabang dari statistika terapan yang berfokus pada penerapan metode kuantitatif untuk memahami fenomena biologis dan kesehatan yang bersifat kompleks, bervariasi, dan mengandung ketidakpastian. Konsep utama dalam biostatistika adalah bahwa variasi merupakan karakteristik alamiah dari sistem biologis, sehingga data kesehatan tidak pernah bersifat identik antar individu, waktu, maupun tempat. Oleh karena itu, diperlukan pendekatan statistik untuk mengelola variasi tersebut agar dapat ditarik kesimpulan ilmiah yang sahih (Rosner, 2016; Bland, 2015).

Secara konseptual, biostatistika tidak hanya berfungsi sebagai alat hitung, tetapi juga sebagai kerangka berpikir ilmiah (*statistical thinking*). Kerangka

ini mencakup kemampuan untuk merumuskan masalah penelitian secara kuantitatif, mengidentifikasi sumber variasi dan bias, serta menilai tingkat ketidakpastian dari suatu hasil penelitian. Dengan demikian, biostatistika berperan penting dalam menjamin objektivitas dan keandalan kesimpulan penelitian kesehatan (Altman, 1991).

2.1.2 Biostatistika sebagai Ilmu dan Alat Analisis

Sebagai ilmu, biostatistika memiliki landasan teori yang kuat, meliputi teori probabilitas, distribusi peluang, inferensi statistik, dan pemodelan data. Landasan teoritis ini memungkinkan peneliti untuk melakukan estimasi parameter populasi, pengujian hipotesis, serta prediksi berdasarkan data sampel. Sebagai alat analisis, biostatistika digunakan secara praktis untuk mengolah data hasil pengukuran kesehatan, baik pada penelitian eksperimental maupun observasional (Pagano & Gauvreau, 2018).

Pendekatan biostatistika memungkinkan peneliti membedakan antara variasi yang terjadi secara kebetulan (*random variation*) dan variasi yang mencerminkan pengaruh faktor tertentu. Perbedaan ini sangat penting dalam penelitian kesehatan, khususnya dalam menilai hubungan sebab-akibat dan efektivitas suatu intervensi medis (Gordis, 2014).

2.1.3 Ruang Lingkup Biostatistika dalam Penelitian Kesehatan

Ruang lingkup biostatistika mencakup seluruh tahapan penelitian kesehatan, mulai dari perencanaan hingga interpretasi hasil. Pada tahap perencanaan penelitian, biostatistika berperan dalam pemilihan desain penelitian yang sesuai, penentuan variabel penelitian, serta perhitungan besar sampel yang

memadai untuk mencapai kekuatan uji (*power*) yang optimal. Kesalahan pada tahap ini dapat menyebabkan hasil penelitian menjadi tidak valid atau tidak dapat digeneralisasikan (Kirkwood & Sterne, 2003).

Pada tahap pengumpulan data, prinsip biostatistika digunakan untuk memastikan bahwa proses pengukuran dilakukan secara sistematis, konsisten, dan bebas dari bias. Selanjutnya, pada tahap analisis data, biostatistika digunakan untuk melakukan statistik deskriptif guna menggambarkan karakteristik data, serta statistik inferensial untuk menarik kesimpulan mengenai populasi berdasarkan data sampel (Daniel & Cross, 2018).

2.1.4 Ruang Lingkup Biostatistika dalam Praktik Kesehatan

Ruang lingkup biostatistika dalam penelitian kesehatan dapat dipahami sebagai rangkaian tahapan sistematis yang saling berkaitan, mulai dari perumusan masalah hingga interpretasi hasil penelitian. Setiap tahapan memerlukan penerapan prinsip biostatistika yang tepat agar hasil penelitian valid, reliabel, dan dapat digeneralisasikan.

Dalam praktik kesehatan, biostatistika digunakan untuk mendukung pengambilan keputusan klinis dan kebijakan kesehatan berbasis bukti. Analisis biostatistika memungkinkan tenaga kesehatan untuk menilai efektivitas terapi, memantau tren penyakit, serta mengevaluasi keberhasilan program kesehatan masyarakat. Indikator kesehatan seperti angka kesakitan, prevalens, insidens, dan mortalitas dihitung dan dianalisis menggunakan prinsip biostatistika (Gordis, 2014).

Selain itu, perkembangan sistem informasi kesehatan dan *big data* kesehatan menuntut pemahaman konsep biostatistika yang lebih mendalam. Biostatistika berperan dalam mengolah data kesehatan berskala besar agar dapat diinterpretasikan secara tepat dan digunakan sebagai dasar perencanaan pelayanan kesehatan yang efektif dan efisien (Rosner, 2016).

1. Perumusan Masalah dan Tujuan Penelitian

Tahap awal penelitian kesehatan adalah perumusan masalah dan tujuan penelitian. Pada tahap ini, biostatistika berperan dalam membantu peneliti merumuskan pertanyaan penelitian yang dapat diuji secara kuantitatif, serta menentukan variabel yang relevan untuk menjawab permasalahan tersebut. Rumusan masalah yang jelas dan terukur menjadi dasar bagi seluruh tahapan penelitian berikutnya (Altman, 1991).

2. Penentuan Desain Penelitian

Biostatistika digunakan untuk memilih desain penelitian yang paling sesuai dengan tujuan penelitian, seperti desain deskriptif, analitik observasional (cross-sectional, case-control, cohort), maupun eksperimental (uji klinis). Pemilihan desain yang tepat akan menentukan jenis data yang dikumpulkan, metode analisis yang digunakan, serta kekuatan kesimpulan penelitian (Gordis, 2014).

3. Penentuan Populasi dan Sampel

Pada tahap ini, prinsip biostatistika digunakan untuk mendefinisikan populasi target dan menentukan teknik pengambilan sampel yang representatif. Selain itu, dilakukan perhitungan besar sampel untuk memastikan penelitian

memiliki kekuatan uji (*statistical power*) yang memadai dalam mendeteksi perbedaan atau hubungan yang bermakna secara statistik (Kirkwood & Sterne, 2003).

4. Penentuan Variabel dan Skala Pengukuran

Biostatistika berperan dalam mengklasifikasikan variabel penelitian (independen, dependen, perancu) serta menentukan skala pengukuran data (nominal, ordinal, interval, rasio). Ketepatan dalam tahap ini sangat penting karena akan memengaruhi pemilihan metode analisis statistik pada tahap selanjutnya (Pagano & Gauvreau, 2018).

5. Pengumpulan Data

Pada tahap pengumpulan data, prinsip biostatistika digunakan untuk memastikan bahwa data dikumpulkan secara sistematis, konsisten, dan minim bias. Instrumen pengukuran harus memiliki validitas dan reliabilitas yang baik agar data yang dihasilkan mencerminkan kondisi sebenarnya (Daniel & Cross, 2018).

6. Pengolahan dan Penyajian Data

Data yang telah dikumpulkan selanjutnya diolah melalui proses editing, coding, dan tabulasi. Biostatistika digunakan untuk menyajikan data dalam bentuk tabel dan grafik sehingga memudahkan pemahaman pola dan karakteristik data secara deskriptif (Zar, 2010).

7. Analisis Data Statistik

Tahap analisis data merupakan inti dari penerapan biostatistika. Analisis dapat berupa statistik deskriptif untuk menggambarkan data,

serta statistik inferensial untuk menguji hipotesis dan menarik kesimpulan mengenai populasi berdasarkan data sampel. Pemilihan uji statistik harus disesuaikan dengan tujuan penelitian, jenis data, dan asumsi statistik yang berlaku (Bland, 2015).

8. Interpretasi dan Penarikan Kesimpulan

Hasil analisis statistik harus diinterpretasikan secara ilmiah dengan mempertimbangkan konteks kesehatan, makna klinis, dan keterbatasan penelitian. Biostatistika membantu peneliti membedakan antara signifikansi statistik dan signifikansi klinis sehingga kesimpulan yang diambil tidak bersifat menyesatkan (Rosner, 2016).

9. Pelaporan dan Diseminasi Hasil Penelitian

Tahap akhir ruang lingkup biostatistika adalah pelaporan hasil penelitian dalam bentuk laporan ilmiah, skripsi, tesis, disertasi, atau publikasi jurnal. Penyajian hasil analisis statistik harus dilakukan secara transparan, akurat, dan sesuai kaidah ilmiah agar dapat dipahami dan dimanfaatkan oleh pembaca serta pemangku kepentingan (Dahlan, 2014).

2.1.5 Keterkaitan Biostatistika dengan Ilmu Lain

Biostatistika memiliki keterkaitan erat dengan berbagai disiplin ilmu, seperti epidemiologi, kesehatan masyarakat, kedokteran klinis, farmasi, dan ilmu gizi. Dalam epidemiologi, biostatistika digunakan untuk menganalisis pola distribusi penyakit dan faktor risikonya. Dalam kedokteran klinis, biostatistika mendukung evaluasi uji klinis dan penelitian diagnostik. Keterpaduan biostatistika dengan ilmu-ilmu tersebut menjadikan biostatistika sebagai komponen

fundamental dalam pengembangan ilmu dan praktik kesehatan modern (Dahlan, 2014).

Biostatistika merupakan disiplin ilmu yang bersifat interdisipliner, karena penerapannya tidak dapat dipisahkan dari berbagai bidang ilmu yang berkaitan dengan kesehatan, biologi, dan penelitian ilmiah. Biostatistika berfungsi sebagai alat analisis, pengambilan keputusan, dan validasi ilmiah, sehingga menjadi jembatan antara data empiris dan pengetahuan ilmiah yang dapat dipertanggungjawabkan. Keterkaitan ini menjadikan biostatistika sebagai komponen esensial dalam pengembangan ilmu pengetahuan modern, khususnya di bidang kesehatan dan ilmu hayati.

Biostatistika tidak berdiri sebagai ilmu yang terpisah, melainkan berkembang melalui interaksi dinamis dengan berbagai disiplin ilmu. Keterkaitan ini menjadikan biostatistika berfungsi sebagai bahasa kuantitatif universal yang memungkinkan berbagai bidang ilmu memahami fenomena empiris secara objektif, terukur, dan dapat diuji secara ilmiah. Dalam konteks ilmu kesehatan dan penelitian, biostatistika berperan sebagai penghubung antara teori, data, dan pengambilan keputusan berbasis bukti.

1. Biostatistika dan Ilmu Kedokteran

Dalam ilmu kedokteran, biostatistika berperan penting dalam penelitian klinis, uji efektivitas obat, diagnosis penyakit, serta evaluasi terapi. Konsep seperti probabilitas, risiko relatif, odds ratio, dan interval kepercayaan digunakan untuk menilai hubungan antara faktor risiko dan kejadian penyakit. Tanpa biostatistika, hasil penelitian klinis tidak dapat diinterpretasikan secara objektif dan berpotensi menyesatkan praktik klinik (Gordis, 2014).

2. Biostatistika dan Kesehatan Masyarakat

Biostatistika memiliki hubungan yang sangat erat dengan kesehatan masyarakat, terutama dalam epidemiologi, surveilans penyakit, perencanaan program kesehatan, dan evaluasi kebijakan kesehatan. Analisis data morbiditas, mortalitas, prevalensi, dan insidensi penyakit sangat bergantung pada metode biostatistika. Melalui pendekatan statistik, tren kesehatan masyarakat dapat diidentifikasi dan digunakan sebagai dasar pengambilan keputusan berbasis bukti (*evidence-based public health*) (Kirkwood & Sterne, 2003).

Biostatistika berperan dalam ilmu kesehatan lingkungan, terutama dalam penilaian risiko lingkungan, analisis paparan bahan berbahaya, dan evaluasi dampak lingkungan terhadap kesehatan manusia. Data kualitas air, udara, dan tanah dianalisis secara statistik untuk menentukan hubungan sebab-akibat antara faktor lingkungan dan gangguan kesehatan masyarakat (Ott & Longnecker, 2015).

Dalam ilmu sosial kesehatan dan perilaku, biostatistika digunakan untuk menganalisis faktor sosial, ekonomi, dan perilaku yang memengaruhi derajat kesehatan masyarakat. Variabel-variabel sosial yang bersifat kompleks memerlukan pendekatan statistik untuk mengidentifikasi pola, hubungan, dan determinan kesehatan secara kuantitatif. Dengan demikian, biostatistika mendukung pengembangan kebijakan kesehatan yang berbasis data sosial empiris (Gordis, 2014).

Secara keseluruhan, keterkaitan biostatistika dengan berbagai ilmu menunjukkan bahwa biostatistika berfungsi sebagai fondasi analitik

lintas disiplin. Integrasi biostatistika dengan ilmu lain tidak hanya meningkatkan kualitas penelitian, tetapi juga memastikan bahwa keputusan di bidang kesehatan didasarkan pada data yang valid, reliabel, dan dapat dipertanggungjawabkan secara ilmiah.

3. Biostatistika dan Epidemiologi

Epidemiologi dan biostatistika merupakan dua bidang yang saling melengkapi. Epidemiologi berfokus pada distribusi dan determinan masalah kesehatan, sedangkan biostatistika menyediakan metode kuantitatif untuk menganalisis dan menginterpretasikan data epidemiologis. Desain studi seperti kohort, kasus-kontrol, dan uji klinis memerlukan analisis statistik yang tepat agar kesimpulan yang dihasilkan valid secara ilmiah (Rosner, 2016).

4. Biostatistika dan Ilmu Biologi

Dalam ilmu biologi, biostatistika digunakan untuk menganalisis variasi biologis, genetika, biologi molekuler, dan bioteknologi. Fenomena biologis yang bersifat kompleks dan bervariasi antarindividu memerlukan pendekatan statistik untuk memahami pola, hubungan, dan mekanisme yang mendasarinya. Biostatistika memungkinkan peneliti biologi untuk menguji hipotesis dan menarik kesimpulan yang dapat digeneralisasikan (Zar, 2010).

5. Biostatistika dan Metodologi Penelitian

Biostatistika merupakan bagian integral dari metodologi penelitian, khususnya dalam penentuan desain penelitian, ukuran sampel, teknik analisis data, dan interpretasi hasil. Kesalahan dalam penerapan statistik dapat menyebabkan bias dan

kesimpulan yang tidak valid. Oleh karena itu, pemahaman konsep biostatistika menjadi syarat utama dalam menghasilkan penelitian yang berkualitas dan dapat dipercaya (Dahlan, 2014).

Secara keseluruhan, keterkaitan biostatistika dengan berbagai ilmu lain menunjukkan bahwa biostatistika bukan hanya alat bantu analisis, melainkan fondasi ilmiah dalam penelitian dan praktik kesehatan. Integrasi biostatistika dengan disiplin ilmu lain mendukung pengambilan keputusan yang rasional, objektif, dan berbasis data.

2.2 Peran Biostatistika dalam Bidang Kesehatan

Biostatistika merupakan cabang ilmu statistika yang diaplikasikan dalam bidang kesehatan dan ilmu hayati. Perannya sangat fundamental karena hampir seluruh keputusan dalam kesehatan modern—baik klinis, epidemiologis, maupun kebijakan kesehatan—bergantung pada analisis data yang sistematis, objektif, dan terukur. Biostatistika menyediakan metode untuk merancang penelitian, mengumpulkan data, menganalisis data, menginterpretasikan hasil, serta menarik kesimpulan berbasis bukti (Rosner, 2016; Pagano & Gauvreau, 2018).

Dalam konteks kesehatan, biostatistika bukan sekadar alat hitung, melainkan instrumen ilmiah yang memastikan bahwa intervensi, diagnosis, maupun kebijakan yang diambil benar-benar didukung oleh data yang valid dan reliabel.

2.2.1 Peran dalam Penelitian Klinis

Dalam penelitian klinis, biostatistika berperan sejak tahap perencanaan hingga interpretasi hasil. Peran tersebut meliputi:

1. **Penentuan desain penelitian** (*cross-sectional, case-control, cohort, randomized controlled trial*).
2. **Penentuan besar sampel (*sample size calculation*)** untuk memastikan kekuatan uji (*power*) memadai.
3. **Randomisasi dan kontrol bias** dalam uji klinis.
4. **Analisis efektivitas dan keamanan terapi** menggunakan uji statistik yang sesuai.

Tanpa pendekatan statistik yang tepat, hasil penelitian klinis dapat menghasilkan kesimpulan yang bias atau menyesatkan (Friedman *et al.*, 2015). Oleh karena itu, dalam uji klinis terkontrol acak (RCT), biostatistika memastikan bahwa perbedaan hasil antar kelompok benar-benar disebabkan oleh intervensi, bukan oleh faktor kebetulan atau variabel perancu.

Dalam penelitian klinis, biostatistika berperan dalam menata desain penelitian, mengendalikan bias, menentukan ukuran sampel, dan mengevaluasi efektivitas intervensi (Piantadosi, 2017). Metode statistik modern seperti analisis Bayesian digunakan untuk pengambilan keputusan adaptif dalam uji klinis fase lanjut (Berry *et al.*, 2019).

Contoh Aplikasi:

1. Menentukan **ukuran sampel yang optimal** melalui perhitungan power dan efek yang diharapkan (Bacchetti, 2019).
2. Penggunaan **metode Bayesian** untuk uji adaptif dalam terapi kanker (Schnall & Simon, 2018).

2.2.2 Peran dalam Epidemiologi

Epidemiologi sangat bergantung pada biostatistika dalam mengukur distribusi dan determinan penyakit dalam populasi. Beberapa indikator epidemiologis yang dianalisis menggunakan metode statistik antara lain:

1. Insidensi
2. Prevalensi
3. Mortality rate
4. Case fatality rate
5. Relative risk (RR)
6. Odds ratio (OR)

Melalui analisis statistik, hubungan antara faktor risiko dan kejadian penyakit dapat diuji secara kuantitatif (Gordis, 2014). Misalnya, dalam studi kohort, biostatistika digunakan untuk menghitung risiko relatif guna mengetahui seberapa besar paparan meningkatkan kemungkinan terjadinya penyakit.

Selain itu, biostatistika juga digunakan dalam surveilans penyakit dan pemodelan wabah, seperti pada pandemi COVID-19, di mana analisis tren dan prediksi epidemiologis sangat krusial untuk pengambilan kebijakan kesehatan masyarakat (Kirkwood & Sterne, 2003).

Epidemiologi menggunakan biostatistika untuk mengukur kejadian, pola, dan faktor risiko penyakit dalam populasi. Analisis hubungan antara paparan dan hasil kesehatan menggunakan statistik modern termasuk regresi multivariat, survival analysis, dan model risiko kompetitif (Rothman *et al.*, 2018; Borgan *et al.*, 2019).

Contoh Aplikasi:

1. Estimasi **relative risk** dan **odds ratio** dalam studi kohort dan kasus-kontrol.
2. Pemodelan distribusi penyakit melalui *spatial statistics* dan *disease mapping* (Lawson, 2020).

2.2.3 Peran dalam Kesehatan Masyarakat

Biostatistika merupakan disiplin ilmu statistika yang diterapkan dalam ilmu kesehatan dan kedokteran untuk menilai, menganalisis, dan mengambil keputusan berbasis data kesehatan. Perannya semakin krusial seiring berkembangnya volume data kesehatan, kompleksitas desain penelitian, dan kebutuhan bukti ilmiah (Suresh & Chandrasekaran, 2018). Statistik bukan hanya alat perhitungan, tetapi kerangka ilmiah untuk mendukung validitas, reliabilitas, dan generalisasi hasil penelitian kesehatan. Dalam bidang kesehatan masyarakat, biostatistika berfungsi sebagai dasar dalam:

1. Perencanaan program kesehatan
2. Evaluasi intervensi kesehatan
3. Monitoring dan evaluasi program
4. Analisis determinan sosial kesehatan

Program imunisasi, pencegahan stunting, pengendalian penyakit menular, hingga edukasi kesehatan berbasis komunitas memerlukan analisis statistik untuk menilai efektivitasnya. Data pre-test dan post-test dalam intervensi edukasi, misalnya, dianalisis menggunakan uji t berpasangan atau uji Wilcoxon untuk menentukan signifikansi peningkatan pengetahuan (Daniel & Cross, 2018). Keputusan berbasis bukti (evidence-based public health) tidak dapat dilepaskan dari analisis biostatistik yang kuat.

Dalam intervensi kesehatan masyarakat, biostatistika digunakan untuk merencanakan, menganalisis, dan mengevaluasi program secara kuantitatif. Analisis statistik pada data surveilans membantu mengidentifikasi tren penyakit dan mengevaluasi dampak kebijakan kesehatan (Zhao & Chen, 2019).

Contoh Aplikasi:

1. Analisis *time-series* untuk tren penyakit menular.
2. Uji efektivitas program pencegahan stunting menggunakan model regresi multilevel.

2.2.4 Peran dalam Diagnostik dan Laboratorium Klinik

Dalam diagnostik medis dan laboratorium klinik, biostatistika digunakan untuk:

1. Menentukan nilai rujukan (*reference value*)
2. Menghitung sensitivitas dan spesifisitas
3. Menghitung nilai prediktif positif (PPV) dan negatif (NPV)
4. Analisis kurva ROC (*Receiver Operating Characteristic*)

Evaluasi validitas suatu metode pemeriksaan laboratorium, seperti pemeriksaan protein urin, glukosa darah, atau elektrolit serum, membutuhkan analisis statistik untuk memastikan akurasi dan presisi metode tersebut (Altman & Bland, 1994). Tanpa analisis statistik, hasil laboratorium tidak dapat ditafsirkan secara ilmiah dan berpotensi menimbulkan kesalahan diagnosis.

Analisis statistik digunakan untuk menilai validitas dan akurasi tes diagnostik termasuk sensitivitas, spesifisitas, nilai prediktif, dan kurva ROC. Metode statistik membantu memastikan bahwa alat

diagnostik memiliki performa yang optimal (Zweig & Campbell, 1993; Mandrekar, 2017).

Contoh Aplikasi:

1. Evaluasi tes cepat COVID-19 melalui kurva ROC (*Receiver Operating Characteristic*).
2. Penggunaan statistik nonparametrik untuk perbandingan antar metode laboratorium.

2.2.5 Peran dalam Pengambilan Kebijakan Kesehatan

Pengambilan kebijakan kesehatan memerlukan data yang sahih dan terukur. Biostatistika membantu dalam:

1. Analisis tren penyakit
2. Evaluasi beban penyakit (*burden of disease*)
3. Analisis cost-effectiveness
4. Penentuan prioritas intervensi kesehatan

Organisasi kesehatan seperti WHO dan Kementerian Kesehatan menggunakan analisis statistik populasi untuk menentukan strategi nasional pengendalian penyakit (WHO, 2022). Data tanpa analisis statistik tidak memiliki kekuatan untuk dijadikan dasar kebijakan.

Statistik populasi dan evaluasi risiko kesehatan menjadi dasar perumusan kebijakan. Analisis data besar (*big data*) membentuk landasan keputusan berbasis bukti dalam strategi kesehatan nasional maupun global (Lee *et al.*, 2021).

Contoh Aplikasi:

1. Pemodelan prediktif epidemik untuk skenario kebijakan kesehatan.

2. Evaluasi *cost-effectiveness* intervensi kesehatan masyarakat.

DAFTAR PUSTAKA

- Dahlan, M. S. (2014). *Statistik untuk Kedokteran dan Kesehatan*. Jakarta: Salemba Medika.
- Gordis, L. (2014). *Epidemiology* (5th ed.). Philadelphia: Elsevier.
- Pagano, M., & Gauvreau, K. (2018). *Principles of Biostatistics* (2nd ed.). Boston: Cengage Learning.
- Rosner, B. (2016). *Fundamentals of Biostatistics* (8th ed.). Boston: Cengage Learning.
- Altman, D. G. (1991). *Practical Statistics for Medical Research*. London: Chapman & Hall.
- Bland, M. (2015). *An Introduction to Medical Statistics* (4th ed.). Oxford: Oxford University Press.
- Daniel, W. W., & Cross, C. L. (2018). *Biostatistics: A Foundation for Analysis in the Health Sciences* (11th ed.). Hoboken: Wiley.
- Kirkwood, B. R., & Sterne, J. A. C. (2003). *Essential Medical Statistics* (2nd ed.). Oxford: Blackwell Science.
- Zar, J. H. (2010). *Biostatistical Analysis* (5th ed.). Upper Saddle River: Pearson.
- Chow, S. C., & Liu, J. P. (2014). *Design and analysis of bioavailability and bioequivalence studies* (3rd ed.). Boca Raton: CRC Press.
- Gibson, R. S. (2005). *Principles of nutritional assessment* (2nd ed.). New York: Oxford University Press.
- Gordis, L. (2014). *Epidemiology* (5th ed.). Philadelphia: Elsevier Saunders.

- Ott, W. R., & Longnecker, M. (2015). *An introduction to statistical methods and data analysis*. Boston: Cengage Learning.
- Polit, D. F., & Beck, C. T. (2017). *Nursing research: Generating and assessing evidence for nursing practice* (10th ed.). Philadelphia: Wolters Kluwer.
- Altman, D. G., & Bland, J. M. (1994). Diagnostic tests 1: Sensitivity and specificity. *BMJ*, 308(6943), 1552.
- Daniel, W. W., & Cross, C. L. (2018). *Biostatistics: A foundation for analysis in the health sciences* (11th ed.). Wiley.
- Friedman, L. M., Furberg, C. D., & DeMets, D. L. (2015). *Fundamentals of clinical trials* (4th ed.). Springer.
- Gordis, L. (2014). *Epidemiology* (5th ed.). Elsevier Saunders.
- Kirkwood, B. R., & Sterne, J. A. C. (2003). *Essential medical statistics* (2nd ed.). Blackwell Science.
- Pagano, M., & Gauvreau, K. (2018). *Principles of biostatistics* (2nd ed.). CRC Press.
- Rosner, B. (2016). *Fundamentals of biostatistics* (8th ed.). Cengage Learning.
- Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B., & Richardson, W. S. (2000). Evidence based medicine: What it is and what it isn't. *BMJ*, 312(7023), 71–72.
- World Health Organization. (2022). *World health statistics 2022*. WHO Press.
- Suresh, K. P., & Chandrasekaran, V. (2018). Sample size estimation and power analysis for clinical research studies. *Journal of Human Reproductive Sciences*, 11(4), 263–268.
- Berry, S. M., Carlin, B. P., Lee, J. J., & Muller, P. (2019). *Bayesian adaptive methods for clinical trials*. Chapman & Hall/CRC.

- Bacchetti, P. (2019). *Current sample size practices in clinical research and the role of the biostatistician*. *Journal of Biopharmaceutical Statistics*, 29(1), 5–12.
- Borgan, Ø., Keogh, R. H., & Tsiatis, A. (2019). *Statistical methods in epidemiologic studies*. Springer.
- Lawson, A. B. (2020). *Bayesian disease mapping: Hierarchical modeling in spatial epidemiology* (3rd ed.). CRC Press.
- Zhao, Y., & Chen, Q. (2019). *Statistical methods in public health interventions*. *International Journal of Public Health*, 64, 343–352.
- Liu, Y., et al. (2020). *Multilevel regression analysis of public health program outcomes*. *Public Health Reports*, 135(3), 312–322.
- Lee, K. H., et al. (2021). *Big data analytics in public health policy formulation*. *Health Policy and Technology*, 10(1), 123–130.
- Neumann, P. J., et al. (2020). *Cost-effectiveness analysis in health policy decision making*. *Medical Decision Making*, 40(7), 815–826.

BAB 3

PENGUMPULAN DAN MANAJEMEN DATA KESEHATAN

Oleh Asruria Sani Fajriah

Integrasi data dalam perawatan kesehatan secara signifikan meningkatkan efisiensi dan efektivitas dengan memanfaatkan analisis data besar, meningkatkan hasil klinis, dan mendorong pengambilan keputusan yang lebih baik. Transformasi ini terbukti melalui berbagai inisiatif dan sistem yang memanfaatkan data untuk merampingkan proses dan mempersonalisasi perawatan.

Analisis data besar memungkinkan pemantauan data kesehatan secara terus menerus, yang dapat mengurangi biaya sekaligus meningkatkan kualitas layanan (Mancini ASL Bari, 2014). Sistem HealthCare Carolinas mencontohkan hal ini dengan membangun infrastruktur data terpusat yang memprediksi hasil dan menginformasikan tindakan klinis, yang mengarah pada peningkatan perawatan pasien (Dulin, Lovin and Wright, 2014).

Program seperti NPS MedicineWise di Australia memanfaatkan data kesehatan untuk meningkatkan hasil di bidang-bidang seperti perawatan kanker dan manajemen diabetes (Teede *et al.*, 2019). Model Health Information Exchange (HIE) memfasilitasi berbagi data antar organisasi, meningkatkan kualitas perawatan dan efisiensi operasional (Salomi *et al.*, 2020).

Terlepas dari potensi manfaatnya, tantangan seperti rintangan teknologi dan kebutuhan akan perubahan budaya dalam organisasi dapat menghambat penggunaan data yang efektif (Dulin, Lovin and Wright, 2014). Namun, mengatasi tantangan ini dapat membuka peluang lebih lanjut untuk meningkatkan pemberian layanan kesehatan.

Data yang tidak akurat berdampak signifikan pada layanan kesehatan dengan mengorbankan keselamatan pasien, meningkatkan biaya, dan menghambat pemberian layanan kesehatan yang efektif. Kualitas perawatan kesehatan secara langsung terkait dengan keakuratan data, karena dokumentasi dan komunikasi yang buruk dapat menyebabkan kesalahan dan efek samping (Clark, 2015). Masalah ini diperburuk oleh kompleksitas yang diperkenalkan oleh catatan kesehatan elektronik (EHR) dan model reformasi pembayaran, yang memerlukan data yang tepat untuk implementasi yang efektif (Clark, 2015).

Data yang tidak akurat dapat menyebabkan kesalahan diagnosis dan perawatan yang tidak tepat, membahayakan keselamatan pasien. Studi kasus menunjukkan bahwa kesalahan kecil dalam data, seperti pengukuran tekanan darah, dapat menyebabkan kesalahan perhitungan yang signifikan dalam efektivitas pengobatan (Delva and Posthumus, 2022).

Kualitas data yang buruk juga dapat mengakibatkan peningkatan biaya operasional karena inefisiensi dan kebutuhan akan tindakan korektif. Organisasi dapat menghadapi hukuman finansial jika mereka gagal memenuhi standar pelaporan kualitas karena data yang tidak akurat (Clark, 2015).

Salah identifikasi pasien dan sampel dapat menyebabkan konsekuensi yang parah, termasuk perawatan yang salah dan perawatan yang tertunda (Lippi *et al.*, 2017). Pandemi COVID-19 menyoroti kebutuhan kritis akan data

yang akurat, karena ketidakakuratan menghambat respons kesehatan masyarakat yang efektif (Arvisais-Anhalt *et al.*, 2021).

3.1 Peran Pengumpulan dan Manajemen Data dalam Biostatistika

3.1.1 Data Sebagai Fondasi Analisis Statistik

Data memainkan peran penting dalam pengembangan analisis statistik yang akurat, berfungsi sebagai dasar untuk pengambilan keputusan berdasarkan informasi di berbagai bidang. Integrasi data ke dalam metodologi statistik memungkinkan ekstraksi wawasan yang penting untuk memahami fenomena kompleks dan memandu tindakan strategis.

Dalam ilmu data, statistik mengubah data mentah menjadi wawasan yang dapat ditindaklanjuti, memfasilitasi pengambilan keputusan yang efektif (Gupta and V., 2020). Alat statistik digunakan untuk manajemen data, analisis, dan visualisasi, memastikan bahwa data dapat ditafsirkan dan berguna untuk keputusan strategis.

Analisis data juga penting dalam mencapai Tujuan Pembangunan Berkelanjutan (SDGs) dengan menyediakan indikator kemajuan yang andal (Al-Guraibawi and Kamel, 2023). Mengatasi masalah kualitas data, seperti outlier, sangat penting untuk memastikan keakuratan analisis yang menginformasikan strategi pengembangan.

3.1.2 Hubungan Pengumpulan Data, Manajemen Data, dan Biostatistika

Biostatistik memainkan peran penting dalam penelitian kesehatan dengan meningkatkan pengumpulan data, manajemen, dan analisis yang penting untuk pengambilan keputusan berdasarkan informasi. Tujuannya yaitu untuk menyediakan kerangka kerja untuk merancang studi, memilih metode statistik yang tepat, dan menafsirkan hasil, sehingga memastikan keandalan dan validitas temuan terkait kesehatan.

Biostatistik membantu dalam desain studi penelitian, memastikan bahwa metode pengumpulan data kuat dan sesuai untuk pertanyaan penelitian. Ini membantu dalam menentukan ukuran sampel dan kriteria seleksi, yang sangat penting untuk validitas hasil penelitian (Manja and Lakshminrusimha, 2014).

Metode biostatistik memfasilitasi organisasi dan pengelolaan kumpulan data besar, memungkinkan penanganan dan analisis data yang efisien. Teknik statistik yang canggih memungkinkan peneliti untuk menganalisis data yang kompleks, yang mengarah ke kesimpulan yang lebih akurat tentang intervensi dan hasil kesehatan (Bangdiwala, 1999).

Biostatistik sangat penting untuk menafsirkan temuan penelitian, membimbing tenaga kesehatan dan tenaga medis serta para pembuat kebijakan dalam membuat keputusan berbasis bukti (Awe *et al.*, 2022). Biostatistik juga berperan dalam mengevaluasi tes diagnostik dan kemanjuran pengobatan, yang sangat penting untuk inisiatif kesehatan masyarakat (Manja and Lakshminrusimha, 2014). Kurangnya pemanfaatan biostatistik di wilayah tertentu, terutama di negara berkembang, menimbulkan tantangan. Kurangnya

profesional terlatih dapat menghambat kualitas penelitian dan perencanaan program kesehatan (Awe *et al.*, 2022).

3.1.3 Konsep "*Garbage In, Garbage Out*" Dalam Penelitian Kesehatan

Konsep "Garbage In, Garbage Out" (GIGO) dalam penelitian kesehatan menekankan pentingnya kualitas data dalam menghasilkan hasil yang akurat. Entri dan manajemen data yang buruk dapat menyebabkan kesimpulan yang salah, merusak integritas penelitian kesehatan. Prinsip ini sangat relevan dalam berbagai konteks, termasuk kesehatan masyarakat, analisis data media sosial, dan meta-analisis, di mana kualitas data input secara langsung mempengaruhi validitas temuan penelitian.

Data berkualitas tinggi harus akurat, valid, lengkap, dan tersedia untuk memastikan hasil kesehatan yang andal (Kilkenny and Robinson, 2018). Sebuah studi di India mengungkapkan bahwa hanya 5,5% makalah penelitian kesehatan masyarakat melaporkan jaminan kualitas selama entri data, menyoroti kesenjangan yang signifikan dalam standar penelitian (Faizi, Kumar and Kazmi, 2018).

Penggunaan media sosial sebagai sumber data memerlukan kerangka penilaian kualitas yang ketat untuk menghindari kesimpulan yang bias, karena metode pengumpulan data yang tidak tepat dapat mendistorsi temuan (Kim, Huang and Emery, 2016). Meta-analisis, sering dipandang sebagai sumber bukti tinggi, dapat menghasilkan hasil yang menyesatkan jika studi yang disertakan berkualitas buruk atau jika entri data cacat (Weinmann, 2009).

Studi yang lebih kecil, yang lebih rentan terhadap kesalahan acak, dapat mengubah hasil meta-analisis, menekankan perlunya pemilihan dan validasi data input yang cermat (Weinmann, 2009). Prinsip GIGO menggarisbawahi perlunya langkah-langkah kualitas data yang ketat, Namun di sisi lain mengungkapkan bahwa meningkatnya ketergantungan pada analisis data besar dapat menyebabkan rasa puas diri mengenai integritas data, hal ini berpotensi memperburuk masalah GIGO dalam penelitian kesehatan (Raghavan, 2015).

3.1.4 Dampak Kualitas Data Terhadap Hasil Analisis

Kualitas data dapat mempengaruhi hasil analisis statistik, karena ketidakakuratan dapat menyebabkan kesimpulan yang menyesatkan dan pengambilan keputusan yang buruk. Data berkualitas tinggi meningkatkan keandalan hasil statistik, sementara data berkualitas rendah dapat menimbulkan bias dan kesalahan. Strategi untuk meningkatkan kualitas data meliputi audit sistematis, kepatuhan terhadap metrik kualitas, dan pelatihan untuk penanganan data. Bagian berikut menguraikan aspek-aspek ini.

Studi menunjukkan bahwa audit data dapat mengungkapkan tingkat kesalahan yang tinggi, seperti perbedaan 17,1% dalam variabel kunci, yang secara langsung mempengaruhi kesimpulan statistik. Peningkatan kualitas data pasca-audit menyebabkan perkiraan kematian dan kejadian penyakit yang lebih tinggi, menunjukkan peran penting integritas data dalam penelitian kesehatan (Giganti *et al.*, 2019).

Menerapkan daftar periksa berdasarkan dimensi seperti kelengkapan, akurasi, dan konsistensi dapat secara signifikan mengurangi cacat pada kumpulan data. Misalnya, sebuah proyek menunjukkan peningkatan kumpulan data dengan nol cacat dari 13,3% menjadi

81% setelah intervensi. Memberikan pelatihan untuk penanganan data dan konsultasi dengan para ahli dapat meningkatkan metode penataan dan penangkapan data, memastikan kualitas data yang lebih baik (Shaheen *et al.*, 2019).

3.2 Metode Pengumpulan Data Kesehatan

3.2.1 Wawancara

Mengumpulkan data kesehatan melalui wawancara dapat meningkatkan pemahaman tentang pengalaman pasien dan hasil kesehatan. Ada berbagai metode, masing-masing dengan keuntungan dan tantangan yang berbeda. Berikut ini metode wawancara utama yang digunakan dalam pengumpulan data kesehatan, menekankan aplikasi dan implikasinya.

A. Jenis Wawancara

1. Wawancara Terstruktur: Ini melibatkan serangkaian pertanyaan tetap, memastikan konsistensi di seluruh wawancara. Mereka berguna untuk pengumpulan data kuantitatif tetapi dapat membatasi kedalaman respons (Stuckey, 2013).
2. Wawancara Semi-Terstruktur: Menggabungkan pertanyaan yang telah ditentukan dengan fleksibilitas untuk mengeksplorasi topik secara lebih rinci, wawancara ini memungkinkan data kualitatif yang lebih kaya dan umumnya digunakan dalam penelitian kesehatan (Bolderston, 2012; Stuckey, 2013).
3. Wawancara Naratif: Ini berfokus pada cerita pribadi, memungkinkan peserta untuk berbagi pengalaman mereka secara organik, yang dapat

mengungkapkan wawasan tentang perilaku dan sikap kesehatan (Stuckey, 2013).

B. Teknik Pengumpulan Data

1. Wawancara Tatap Muka: Ini memfasilitasi pembuatan laporan dan pemahaman bernuansa tetapi dapat intensif sumber daya (Bolderston, 2012).
2. Wawancara Jarak Jauh: Dilakukan melalui telepon atau platform online, metode ini meningkatkan aksesibilitas tetapi mungkin tidak memiliki sentuhan pribadi dari interaksi tatap langsung (Bolderston, 2012).
3. Kelompok Fokus: Wawancara kelompok dapat menghasilkan perspektif yang beragam tetapi dapat menyebabkan suara dominan membayangi peserta yang lebih tenang (Bolderston, 2012).

C. Pertimbangan Etis

1. Persetujuan Terinformasi: Memastikan peserta memahami tujuan dan penggunaan data mereka sangat penting untuk kepatuhan etis (Bolderston, 2012).
2. Kerahasiaan: Melindungi identitas peserta dan integritas data sangat penting, terutama dalam konteks kesehatan yang sensitif (Bolderston, 2012).

3.2.2 Kuesioner

Mengumpulkan data kesehatan melalui kuesioner melibatkan berbagai metode yang dapat mempengaruhi kualitas dan keandalan data yang diperoleh. Pendekatan yang berbeda, seperti kuesioner yang dikelola sendiri, wawancara, dan metode campuran, digunakan untuk melayani populasi dan konteks yang beragam.

Kuesioner sering digunakan untuk survei skala besar, memungkinkan responden untuk menyelesaikannya sesuai keinginan mereka. Kuesioner dapat didistribusikan melalui surat atau platform online, seperti yang terlihat dalam studi di seluruh negara anggota Uni Eropa (Unim *et al.*, 2022).

Jaminan Kualitas dalam Pengumpulan Data:

1. Pra-Pengujian dan Validasi: Memvalidasi kuesioner melalui pra-penelitian memastikan bahwa pertanyaan jelas dan relevan. Langkah ini sangat penting untuk menjaga integritas data (Tomášková *et al.*, 2003).
2. Pemantauan Tingkat Respons: Melacak tingkat respons dan menggunakan strategi untuk menindaklanjuti dengan non-responden dapat meningkatkan kelengkapan data. Misalnya, sebuah penelitian mencapai tingkat pengulangan 60,3%, menunjukkan keandalan yang kuat (Tomášková *et al.*, 2003).
3. Mitigasi Bias: Mode administrasi yang berbeda dapat menimbulkan bias. Kesadaran akan bias ini sangat penting bagi peneliti untuk memastikan validitas temuan mereka (Bowling, 2005).

3.2.3 Observasi

Pengamatan/ Observasi adalah metode penting untuk mengumpulkan data kesehatan, menawarkan wawasan yang mungkin diabaikan oleh metode lain. Observasi melibatkan pengamatan dan dokumentasi secara sistematis terhadap perilaku dan interaksi dalam pengaturan alami, yang dapat mengungkapkan faktor kontekstual yang mempengaruhi hasil kesehatan. Pendekatan ini sangat berguna dalam penelitian perawatan kesehatan, di mana memahami lingkungan

dan perilaku sangat penting untuk desain intervensi yang efektif.

Jenis Metode Pengamatan

1. Pengamatan Kualitatif: Melibatkan catatan lapangan terperinci yang menangkap konteks, peserta, dan kegiatan yang terkait dengan fenomena kesehatan (Weston, Krein and Harrod, 2021).
2. Pengamatan Spasial: Berfokus pada lingkungan fisik dan dampaknya terhadap perilaku kesehatan, dikategorikan ke dalam berbagai pendekatan metodologis (Lipson-Smith and McLaughlan, 2022).
3. Pengamatan Terstruktur: Menggunakan kriteria yang telah ditentukan untuk memantau domain perawatan kesehatan tertentu, seperti kesalahan dan kinerja tim, memastikan pengumpulan data yang sistematis (Carthey, 2003).

Keunggulan Observasi

1. Wawasan Kontekstual yang Kaya: Pengamatan dapat mengungkap perilaku dan faktor lingkungan yang mungkin tidak diartikulasikan peserta dalam wawancara atau survei (Harvey, 2018).
2. Melengkapi Metode Lainnya: Ini dapat meningkatkan data dari survei dan wawancara, memberikan pemahaman yang lebih komprehensif tentang masalah kesehatan (Rajan, Basch and Ethan, 2013).

Tantangan dan Keterbatasan

1. Sumber Daya Intensif: Studi observasional dapat memakan waktu dan membutuhkan pengamat terlatih untuk memastikan keandalan (Lipson-Smith and McLaughlan, 2022).
2. Masalah Generalisasi: Temuan mungkin tidak selalu berlaku untuk konteks yang berbeda karena sifat

spesifik dari pengaturan yang diamati (Lipson-Smith and McLaughlan, 2022).

3.2.4 Telaah Dokumen

Pengumpulan data kesehatan melalui telaah dokumen menggunakan berbagai metodologi untuk mengumpulkan dan menganalisis informasi data. Proses ini sangat penting untuk menginformasikan pembuatan kebijakan, memantau sistem, dan meningkatkan layanan kesehatan.

Jenis dan Sumber Dokumen

1. Catatan Administratif: Ini termasuk catatan pasien, log perawatan, dan laporan layanan kesehatan, yang memberikan data komprehensif tentang interaksi dan hasil pasien (Nurcahyati, 2022; Unim *et al.*, 2022).
2. Survei Pemeriksaan Kesehatan (HES): Survei standar yang mengumpulkan pengukuran kesehatan objektif dan sampel biologis, menawarkan data yang andal untuk pemantauan kesehatan dan pengembangan kebijakan (Tolonen *et al.*, 2018).
3. Survei Wawancara Kesehatan: Survei ini mengumpulkan informasi kesehatan yang dilaporkan sendiri, melengkapi data dari sumber administrasi (Unim *et al.*, 2022).

Kerangka Metodologis

1. Tujuh Fase Analisis Dokument: Kerangka kerja ini mencakup mengidentifikasi tujuan, memilih dokumen, merancang matriks ekstraksi, pengujian percontohan, pengumpulan dan analisis data, menilai kredibilitas, dan menangani pertimbangan etis (Moilanen *et al.*, 2022).

2. Penilaian Mutu: Banyak proyek menggunakan standar internasional untuk penilaian kualitas, meskipun kepatuhan bervariasi secara signifikan antar wilayah (Unim *et al.*, 2022).

Aplikasi dalam Kebijakan dan Praktik

1. Menginformasikan Pengembangan Kebijakan: Data dari survei dan catatan kesehatan telah berperan penting dalam mengidentifikasi masalah kesehatan, menetapkan target, dan mengevaluasi kebijakan kesehatan (Oyebode and Mindell, 2014; Tolonen *et al.*, 2018).
2. Pelaporan Rutin: Laporan rutin yang dihasilkan dari catatan medis mendukung evaluasi layanan kesehatan yang sedang berlangsung dan alokasi sumber daya (Nurchayati, 2022).

3.3 Instrumen Pengumpulan Data Kesehatan

Instrumen pengumpulan data kesehatan adalah alat penting yang dirancang untuk mengumpulkan, menganalisis, dan menafsirkan informasi terkait kesehatan secara efektif. Instrumen ini bervariasi dalam struktur dan tujuan, sering disesuaikan dengan populasi atau kondisi kesehatan tertentu.

Instrumen sering dikembangkan melalui proses metodologis yang ketat, termasuk tinjauan literatur dan validasi ahli. Misalnya, sebuah studi tentang pengumpulan data untuk perawatan lansia menggunakan Teori Kebutuhan Manusia Dasar, mencapai indeks validasi konten 0,80 atau lebih tinggi (Porto and Nóbrega, 2008). Studi lain yang berfokus pada pasien onko-hematologi menciptakan instrumen yang divalidasi yang menilai berbagai kebutuhan, mencapai indeks validitas konten 95% dan skor reliabilitas 0,93 (Rodrigues *et al.*, 2021).

Pengumpulan data menggunakan seluler saat ini semakin banyak digunakan untuk memfasilitasi pengumpulan data dalam pengaturan perawatan kesehatan. Platform Care4Value mencontohkan hal ini dengan memungkinkan profesional kesehatan memantau data pasien secara efisien, sehingga mengoptimalkan alokasi sumber daya (Ferreira, Maximiano and Reis, 2019). Selain itu, kerangka kerja seperti QuestionSys memungkinkan pakar domain untuk membuat dan memodifikasi aplikasi pengumpulan data seluler dengan mudah (Schobel *et al.*, 2017).

Hal yang perlu diperhatikan oleh para peneliti bahwa instrumen pengumpulan data perlu dilakukan uji coba instrumen untuk mengembangkan dan memvalidasi alat pengukuran yang dapat secara akurat menilai konstruksi atau hasil tertentu. Proses ini biasanya melibatkan beberapa tahap, termasuk desain, pengujian, dan evaluasi, memastikan bahwa instrumen dapat diandalkan dan valid untuk tujuan penggunaannya.

Tahapan Pengembangan

1. Desain Awal: Instrumen dirancang berdasarkan kerangka teoritis dan kebutuhan empiris, seperti yang terlihat dalam pengembangan kuesioner penyesuaian diri untuk remaja (Khairun and Hakim, 2022).
2. Pengujian: Instrumen menjalani pengujian percobaan untuk menilai fungsionalitas dan efektivitasnya. Misalnya, tes penalaran ilmiah untuk siswa biologi divalidasi melalui uji coba dengan siswa aktual (Yanto, Subali and Suyanto, 2019).
3. Validasi dan Keandalan: Pasca uji coba, instrumen dievaluasi untuk validitas dan keandalannya, memastikan mereka mengukur apa yang mereka inginkan. Keandalan instrumen penalaran ilmiah ditemukan 0,74, menunjukkan konsistensi yang baik (Yanto, Subali and Suyanto, 2019).

Pertimbangan Etis

1. **Transparansi:** Peneliti memiliki kewajiban etis untuk melaporkan hasil uji coba secara komprehensif, berkontribusi pada domain publik dan mencegah duplikasi upaya penelitian yang tidak perlu (Salvi, 2005).
2. **Masukan Pemangku Kepentingan:** Melibatkan pemangku kepentingan, seperti pakar dan peserta, selama proses pengembangan meningkatkan relevansi dan penerapan instrumen (Saifudin and Fitrianawati, 2021).

Digitalisasi instrumen pengumpulan data seperti Google Forms, KoboToolbox, dan RedCap telah secara signifikan mengubah metodologi penelitian, terutama dalam konteks kesehatan masyarakat dan kemanusiaan. Platform ini meningkatkan manajemen data, merampingkan alur kerja, dan memfasilitasi analisis real-time, menjadikannya sangat berharga bagi para peneliti. Di bawah ini adalah aspek kunci dari fungsi dan aplikasi mereka.

Kemampuan Pengumpulan Data yang Ditingkatkan

1. **KoboToolbox dan ODK:** Platform ini mendukung pengumpulan data off-grid, memungkinkan peneliti mengumpulkan data di lingkungan yang menantang. Mereka telah digunakan secara efektif selama keadaan darurat kesehatan, seperti epidemi Ebola, untuk mengelola kumpulan data besar dari ribuan peserta (Marks *et al.*, 2020).
2. **RedCap:** Platform ini menawarkan template standar untuk penelitian tuberkulosis, mempromosikan harmonisasi data lintas studi. Desain modularnya memungkinkan peneliti untuk menyesuaikan pengumpulan data sesuai dengan kebutuhan studi spesifik (Allie *et al.*, 2021).

Open Source dan Interoperabilitas

Integrasi dengan Perangkat Lunak Statistik: Alat seperti ODK dan KoboToolbox dapat diintegrasikan dengan perangkat lunak statistik (misalnya, R) untuk analisis data lanjutan, meningkatkan proses penelitian keseluruhan (Marks *et al.*, 2020).

Aplikasi dan Tantangan Praktis

1. Survei Lapangan: Alat digital telah menggantikan survei berbasis kertas tradisional, meningkatkan akurasi dan efisiensi dalam pengumpulan data untuk proyek kemanusiaan. Misalnya, KoboCollect telah berhasil diimplementasikan dalam proyek listrik off-grid, menunjukkan manfaat praktis dari digitalisasi (Dauenhauer *et al.*, 2018).
2. Pertimbangan Implementasi: Implementasi efektif sistem pengambilan data elektronik memerlukan perencanaan yang cermat, termasuk desain alur kerja dan pelatihan staf, untuk memastikan manajemen data yang berhasil dalam studi klinis (Eade *et al.*, 2021).

3.4 Manajemen Data Kesehatan

3.4.1 Siklus Manajemen Data

Siklus manajemen data mencakup beberapa tahapan penting yang memastikan penanganan data yang efektif sepanjang masa hidupnya. Siklus ini mencakup pencatatan data, organisasi, penyimpanan, pemeliharaan, dan pengarsipan. Masing-masing memainkan peran penting dalam memaksimalkan utilitas dan umur panjang data. Di bawah ini adalah aspek kunci dari setiap tahap.

Pencatatan Data

1. Melibatkan pengumpulan data yang sistematis, memastikan bahwa semua informasi yang relevan ditangkap secara akurat.
2. Penting untuk membangun fondasi yang andal untuk kegiatan manajemen data berikutnya (Michener, 2015).

Organisasi Data

1. Data harus disusun sedemikian rupa sehingga memudahkan akses dan analisis, seringkali menggunakan database atau spreadsheet (Tavakoli *et al.*, 2006).
2. Organisasi yang tepat mencakup penggunaan metadata untuk menggambarkan tipe data dan metode pengumpulan, meningkatkan kegunaan (Strasser *et al.*, 2012).

Penyimpanan Data

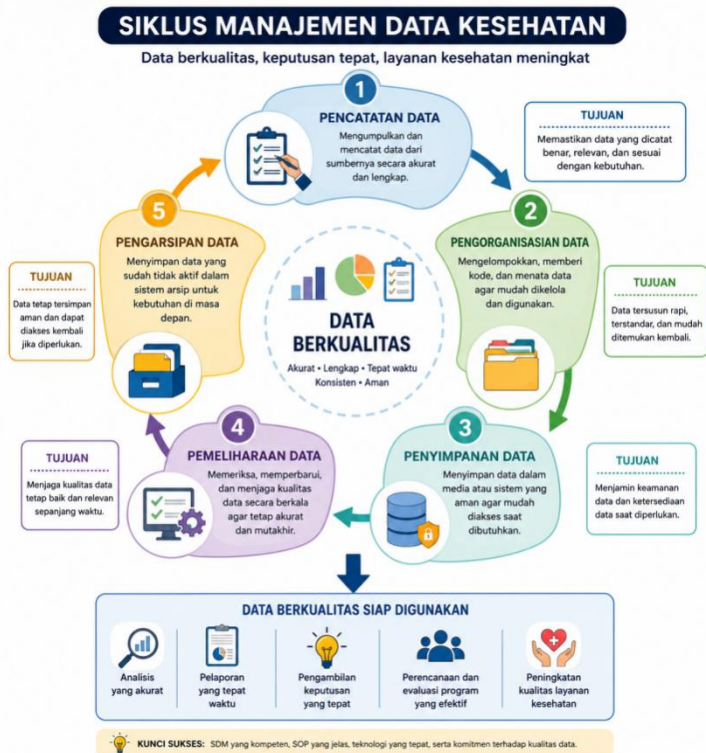
1. Solusi penyimpanan yang efektif, seperti sistem manajemen basis data, sangat penting untuk menjaga integritas dan aksesibilitas data (Maritz, 2003).
2. Pertimbangan meliputi keamanan, pencadangan, dan skalabilitas untuk mengakomodasi kumpulan data yang terus bertumbuh (Rümpel, 2011).

Pemeliharaan Data

1. Pembaruan rutin dan proses jaminan kualitas diperlukan untuk memastikan data tetap akurat dan relevan (Michener, 2015).
2. Ini termasuk memantau data untuk kesalahan dan menerapkan tindakan korektif yang diperlukan (Tavakoli *et al.*, 2006).

Pengarsipan Data

1. Pengarsipan melibatkan pelestarian data jangka panjang untuk penggunaan di masa mendatang, memastikannya tetap dapat diakses dan dapat digunakan dari waktu ke waktu (Rümpel, 2011).
2. Strategi pengarsipan harus mempertimbangkan format data dan potensi teknologi masa depan untuk pengambilan (Maritz, 2003).



Gambar 2. 1 Manajemen Data Kesehatan yang Berkualitas

3.4.2 Strategi Meminimalkan Bias Pengumpulan Data

Strategi yang efektif dapat diterapkan pada berbagai tahap proses penelitian, mulai dari desain hingga pengumpulan dan analisis data. Berikut menguraikan strategi utama untuk mengurangi bias.

1. Refleksivitas dan Demokratisasi

Melibatkan kelompok masyarakat yang kurang terwakili dalam proses pengumpulan data dapat membantu membatasi bias penelitian yang berasal dari sudut pandang peneliti. Peneliti juga harus secara kritis merefleksikan bias mereka sendiri dan bagaimana ini dapat mempengaruhi interpretasi data, terutama ketika mempelajari kelompok yang kurang terwakili (Das, Podder and Semaan, 2022).

2. Desain Metodologis yang Ketat

- a. Sampling Multi-Tahap: Menerapkan teknik pengambilan sampel multi-tahap dapat meningkatkan keterwakilan dan mengurangi bias seleksi dalam studi, terutama dalam pengaturan penelitian multi-situs (Tankumpuan *et al.*, 2023).
- b. Protokol Buta: Melakukan penelitian dengan protokol buta, di mana peneliti tidak mengetahui kelompok perlakuan, dapat secara signifikan mengurangi bias pengamat dan meningkatkan integritas hasil (Holman *et al.*, 2015).

3. Pelatihan dan Pertimbangan Etis

Memberikan pelatihan komprehensif untuk pengumpul data memastikan konsistensi dan mengurangi bias pengukuran, yang dapat timbul dari berbagai tingkat keahlian (Tankumpuan *et al.*, 2023). Peneliti dapat memanfaatkan teknik terjemahan dan memvalidasi konten untuk

relevansi budaya sehingga dapat meminimalkan bias yang terkait dengan bahasa dan konteks (Tankumpuan *et al.*, 2023).

3.4.3 Dokumentasi dan Metadata Data Kesehatan

Metadata memberikan informasi penting tentang proses pengumpulan data, meningkatkan pemahaman dan kegunaan bagi peneliti dan praktisi (McMahon Christiana *et al.*, 2025). Penilaian kualitas metadata secara teratur sangat penting untuk memastikan kelengkapan, akurasi, dan konsistensi, sebagaimana dibuktikan oleh studi tentang repositori data kesehatan federal (Marc *et al.*, 2017).

Kerangka Kerja dan Standar

1. Model Harmonisasi: Model Proses Bisnis Longitudinal Generik (GLBPM) dan standar seperti Data Documentation Initiative (DDI) dan Standard for Data and Metadata eXchange (SDMX) berperan penting dalam menciptakan metadata terstruktur untuk data kesehatan (Kanjala, 2019).
2. Model Data Konsep Kesehatan Nasional (HCDM): Model ini memfasilitasi manajemen metadata standar di berbagai domain kesehatan, mengatasi hambatan semantik dalam pertukaran data (Yang *et al.*, 2022).

DAFTAR PUSTAKA

- Al-Guraibawi, M. and Kamel, H. H. (2023) 'Statistical data processing and its role in achieving sustainable development goals', *AIP Conference Proceedings*. American Institute of Physics Inc., 2776(1). doi: 10.1063/5.0135981/2878068.
- Allie, T. *et al.* (2021) 'TBDBT: A TB DataBase Template for collection of harmonized TB clinical research data in REDCap, facilitating data standardisation for inter-study comparison and meta-analyses', *PLOS ONE*. Public Library of Science, 16(3), p. e0249165. doi: 10.1371/JOURNAL.PONE.0249165.
- Arvisais-Anhalt, S. *et al.* (2021) 'What the Coronavirus Disease 2019 (COVID-19) Pandemic Has Reinforced: The Need for Accurate Data', *Clinical Infectious Diseases*. Oxford Academic, 72(6), pp. 920–923. doi: 10.1093/CID/CIAA1686.
- Awe, O. O. *et al.* (2022) 'Building Biostatistics Capacity in Developing Countries', *Promoting Statistical Practice and Collaboration in Developing Countries*. Chapman and Hall/CRC, pp. 495–514. doi: 10.1201/9781003261148-43/BUILDING-BIOSTATISTICS-CAPACITY-DEVELOPING-COUNTRIES-OLAWALE-AWE-ADENIYI-FRANCIS-FAGBAMIGBE-EGWIM-EVANS-ABDULHAKEEM-ABAYOMI-OLORUKOBA-OLUWAFUNMILOLA-DEBORAH-AWE-MUMINI-IDOWU-ADARABIOYO-OLUMIDE-CHARLES-AYENI-BAYOWA-TENIOLA-BABALO.

- Bangdiwala, S. I. (1999) 'The role of the biostatistician in biomedical research', *Rev Med Chil*, 127(2), pp. 223–8. Available at: <https://pubmed.ncbi.nlm.nih.gov/10436704/> (Accessed: 6 June 2026).
- Bolderston, A. (2012) 'Conducting a Research Interview', *Journal of Medical Imaging and Radiation Sciences*. Elsevier, 43(1), pp. 66–76. doi: 10.1016/J.JMIR.2011.12.002.
- Bowling, A. (2005) 'Mode of questionnaire administration can have serious effects on data quality', *Journal of Public Health*. Oxford Academic, 27(3), pp. 281–291. doi: 10.1093/PUBMED/FDI031.
- Carthey, J. (2003) 'The role of structured observational research in health care', *BMJ Quality & Safety*. BMJ Publishing Group Ltd, 12(suppl 2), pp. ii13–ii16. doi: 10.1136/QHC.12.SUPPL_2.II13.
- Clark, J. S. (2015) 'Assessing and improving EHR Data Quality (Updated)', *Journal of AHIMA / American Health Information Management Association*, 86(5), pp. 58–64.
- Das, D., Podder, A. and Semaan, B. (2022) 'Note: A Sociomaterial Perspective on Trace Data Collection: Strategies for Democratizing and Limiting Bias', *ACM International Conference Proceeding Series*. Association for Computing Machinery, Par F180472, pp. 569–573. doi: 10.1145/3530190.3534835;WGROU:STRING:ACM.

- Dauenhauer, P. *et al.* (2018) 'Improving Shoestring Surveys for Off-Grid Humanitarian Power Projects: Kilowatts for Humanity and KoboCollect', *GHTC 2018 - IEEE Global Humanitarian Technology Conference, Proceedings*. Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/GHTC.2018.8601657.
- Delva, W. and Posthumus, C. (2022) 'The Peril of Imprecise Data in Healthcare – A Simple Example of Blood Pressure Measurements and Cardiovascular Mortality Risk'. OSF. doi: 10.31219/OSF.IO/SGZQB.
- Dulin, M. F., Lovin, C. A. and Wright, J. A. (2014) 'Bringing Big Data to the Forefront of Healthcare Delivery: The Experience of Carolinas HealthCare System - PubMed', *Front Health Serv Manage*, 32(4), pp. 3–14. Available at: <https://pubmed.ncbi.nlm.nih.gov/28199237/> (Accessed: 6 June 2026).
- Eade, D. *et al.* (2021) 'Electronic Data Capture-Study Implementation and Start-up', *Journal of the Society for Clinical Data Management*. Society for Clinical Data Management, 1(4), pp. 1–24. doi: 10.47912/JSCDM.30.
- Faizi, N., Kumar, A. M. and Kazmi, S. (2018) 'Omission of quality assurance during data entry in public health research from India: Is there an elephant in the room?', *Indian journal of public health*. NLM (Medline), 62(2), pp. 150–152. doi: 10.4103/IJPH.IJPH_386_16.
- Ferreira, P. H., Maximiano, M. and Reis, C. I. (2019) 'Care4Value: A clinical instrument data collection platform for long-term health care units', *6th IEEE Portuguese Meeting on Bioengineering, ENBENG 2019 - Proceedings*. Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/ENBENG.2019.8692494.

- Giganti, M. J. *et al.* (2019) 'The impact of data quality and source data verification on epidemiologic inference: a practical application using HIV observational data', *BMC Public Health* 2019 19:1. BioMed Central, 19(1), pp. 1748-. doi: 10.1186/S12889-019-8105-2.
- Gupta, P. and V., N. (2020) 'The Impact and Importance of Statistics in Data Science', *International Journal of Computer Applications*, 176(24), pp. 10–14. doi: 10.5120/ijca2020920215.
- Harvey, S. A. (2018) 'Observe Before You Leap: Why Observation Provides Critical Insights for Formative Research and Intervention Design That You'll Never Get From Focus Groups, Interviews, or KAP Surveys', *Global Health: Science and Practice*. Global Health: Science and Practice, 6(2), pp. 299–316. doi: 10.9745/GHSP-D-17-00328.
- Holman, L. *et al.* (2015) 'Evidence of Experimental Bias in the Life Sciences: Why We Need Blind Data Recording', *PLOS Biology*. Public Library of Science, 13(7), p. e1002190. doi: 10.1371/JOURNAL.PBIO.1002190.
- Kanjala, C. (2019) 'Provenance of "after the fact" harmonised community-based demographic and HIV surveillance data from ALPHA cohorts'.
- Khairun, D. Y. and Hakim, I. Al (2022) 'Pengembangan Instrumen Penyesuaian Diri Remaja', *Indonesian Journal of Educational Counseling*. Universitas Mathla ul Anwar Banten, 6(2), pp. 103–110. doi: 10.30653/001.202262.183.

- Kilkenny, M. F. and Robinson, K. M. (2018) 'Data quality: "Garbage in – garbage out"', *Health Information Management Journal*. SAGE Publications Inc., 47(3), pp. 103–105. doi: 10.1177/1833358318774357;JOURNAL:JOURNAL:HI MD;WGROU:STRING:PUBLICATION.
- Kim, Y., Huang, J. and Emery, S. (2016) 'Garbage in, garbage out: Data collection, quality assessment and reporting standards for social media data use in health research, infodemiology and digital disease detection', *Journal of Medical Internet Research*. JMIR Publications Inc., 18(2), p. e4738. doi: 10.2196/jmir.4738.
- Lippi, G. *et al.* (2017) 'Patient and sample identification: Out of the maze?', *Journal of Medical Biochemistry*. Society of Medical Biochemists of Serbia and Montenegro, 36(2), pp. 107–112. doi: 10.1515/JOMB-2017-0003.
- Lipson-Smith, R. and McLaughlan, R. (2022) 'Mapping Healthcare Spaces: A Systematic Scoping Review of Spatial and Behavioral Observation Methods', *Health Environments Research and Design Journal*. SAGE Publications Inc., 15(3), pp. 351–374. doi: 10.1177/19375867221089702;ISSUE:ISSUE:DOI.
- Mancini ASL Bari, M. (2014) 'Exploiting Big Data for Improving Healthcare Services', *Journal of e-Learning and Knowledge Society*, 10(2), pp. 23–33. doi: 10.20368/1971-8829/929.
- Manja, V. and Lakshminrusimha, S. (2014) 'Principles of Use of Biostatistics in Research', *NeoReviews*. American Academy of Pediatrics, 15(4), pp. e133–e150. doi: 10.1542/NEO.15-4-E133.

- Marc, D. T. *et al.* (2017) 'Assessing Metadata Quality of a Federally Sponsored Health Data Repository', *AMIA Annual Symposium Proceedings*, 2016, p. 864. Available at: <https://pmc.ncbi.nlm.nih.gov/articles/PMC5333273/> (Accessed: 6 June 2026).
- Maritz, S. G. (2003) 'Data management: Managing data as an organisational resource', *Acta Commercii*. AOSIS, 3(1), pp. 75–84. doi: 10.4102/AC.V3I1.44.
- Marks, M. *et al.* (2020) 'Electronic Data Management in Public Health and Humanitarian Crises. Upgrades, Scalability and Impact of ODK'. doi: 10.21203/RS.3.RS-52854/V1.
- McMahon Christiana *et al.* (2025) 'Improving metadata quality assessment in public health and epidemiology'. doi: 10.3233/978-1-61499-512-8-939.
- Michener, W. K. (2015) 'Ten Simple Rules for Creating a Good Data Management Plan', *PLOS Computational Biology*. Public Library of Science, 11(10), p. e1004525. doi: 10.1371/JOURNAL.PCBI.1004525.
- Moilanen, T. *et al.* (2022) 'Developing a Feasible and Credible Method for Analyzing Healthcare Documents as Written Data', *Global Qualitative Nursing Research*. SAGE Publications Inc., 9. doi: 10.1177/23333936221108706;SUBPAGE:STRING:FULL.
- Nurchayati, S. (2022) 'Pemanfaatan Data Rekam Medis Dalam Pelaporan Bulanan Di Puskesmas Kejaksan Cirebon', *Indonesian Journal of Health Information Management*, 2(1).

- Oyebode, O. and Mindell, J. S. (2014) 'A review of the use of health examination data from the Health Survey for England in government policy development and implementation', *Archives of Public Health* 72:1. BioMed Central, 72(1), pp. 24-. doi: 10.1186/2049-3258-72-24.
- Porto, M. L. L. and Nóbrega, M. M. L. (2008) 'Instrument of collection of data for the attendance to the aged one in the of health the family program', *Revista de Enfermagem UFPE on line*. Revista de Enfermagem, UFPE Online, 2(1), pp. 1–8. doi: 10.5205/reuol.399-11159-1-LE.0201200801.
- Raghavan, D. (2015) *Big Data: Not Really the Same as Level 1 Data, Cancer Network*. Available at: <https://www.cancernetwork.com/view/big-data-not-really-same-level-1-data> (Accessed: 6 June 2026).
- Rajan, S., Basch, C. and Ethan, D. (2013) 'Observational Data Collection of Environmental and Behavioral Characteristics: Strengths, Limitations, and Implications for Health Communication', *Journal of Mass Communication and Journalism*. OMICS Publishing Group, 03(05). doi: 10.4172/2165-7912.1000E141.
- Rodrigues, S. M. N. *et al.* (2021) 'Data collection instrument for hematological diseases in na outpatient setting: a validation study', *Revista Brasileira de Enfermagem*. Associação Brasileira de Enfermagem, 74(3), p. e20201034. doi: 10.1590/0034-7167-2020-1034.
- Rümpel, S. (2011) 'The life cycle of research data', *Handbuch Forschungsdatenmanagement*, pp. 25–34. Available at: <http://opus4.kobv.de/opus4-fhpotsdam/frontdoor/index/index/docId/193>.

- Saifudin, M. F. and Fitrianawati, M. (2021) 'Pengembangan Instrumen Tracer Study Berbasis Cluster Sebagai Kontrol Mutu Lulusan Perguruan Tinggi', *Jurnal Penjaminan Mutu*. Institut Hindu Dharma Negeri Denpasar, 7(1), p. 1. doi: 10.25078/JPM.V7I1.1786.
- Salomi, M. J. A. *et al.* (2020) 'Adopting Healthcare Information Exchange among Organizations, Regions, and Hospital Systems toward Quality, Sustainability, and Effectiveness', *Technology and Investment*. Scientific Research Publishing, 11(3), pp. 58–97. doi: 10.4236/TI.2020.113005.
- Salvi, V. S. (2005) 'Research trials: registration, reporting and publication - PubMed', *J Postgrad Med*, 51(2), pp. 83–4. Available at: <https://pubmed.ncbi.nlm.nih.gov/16006696/> (Accessed: 6 June 2026).
- Schobel, J. *et al.* (2017) 'Towards Patterns for Defining and Changing Data Collection Instruments in Mobile Healthcare Scenarios', *Proceedings - IEEE Symposium on Computer-Based Medical Systems*. Institute of Electrical and Electronics Engineers Inc., 2017-June, pp. 101–102. doi: 10.1109/CBMS.2017.61.
- Shaheen, N. A. *et al.* (2019) 'Reducing defects in the datasets of clinical research studies: conformance with data quality metrics', *BMC Medical Research Methodology* 2019 19:1. BioMed Central, 19(1), pp. 98-. doi: 10.1186/S12874-019-0735-7.
- Strasser, C. *et al.* (2012) 'UC Office of the President CDL Staff Publications Title Primer on Data Management: What you always wanted to know Permalink <https://escholarship.org/uc/item/7tf5q7n3> Publication Date Copyright Information'. doi: 10.5060/D2251G48.

- Stuckey, H. L. (2013) 'Three types of interviews: Qualitative research methods in social health', *Journal of Social Health and Diabetes*. Thieme Medical and Scientific Publishers Private Ltd., 01(02), pp. 056–059. doi: 10.4103/2321-0656.115294.
- Tankumpuan, T. *et al.* (2023) "'Snapshot" cohort study design across multi-site research settings: challenges, strategies, and solutions', *European Journal of Cardiovascular Nursing*. Oxford Academic, 22(6), pp. 664–668. doi: 10.1093/EURJCN/ZVAD060.
- Tavakoli, A. *et al.* (2006) 'Data Management Plans: Stages Components and Activities', *Applications and Applied Mathematics*, 1(2), pp. 141–151. Available at: https://www.researchgate.net/publication/242083907_Data_Management_Plans_Stages_Components_and_Activities (Accessed: 6 June 2026).
- Teede, H. J. *et al.* (2019) 'Australian Health Research Alliance: national priorities in data-driven health care improvement', *Medical Journal of Australia*. John Wiley and Sons Inc., 211(11), pp. 494–497.e1. doi: 10.5694/MJA2.50409;WGROU:STRING:PUBLICATION.
- Tolonen, H. *et al.* (2018) 'European health examination surveys – a tool for collecting objective information about the health of the population', *Archives of Public Health* 2018 76:1. BioMed Central, 76(1), pp. 38-. doi: 10.1186/S13690-018-0282-4.
- Tomášková, © H *et al.* (2003) 'METHODICAL APPROACH TO DATA PROCESSING FROM A QUESTIONNAIRE SURVEY', <http://biomed.papers.upol.cz/doi/10.5507/bp.2003.015.html>. Biomedical Papers, 147(1), pp. 101–107. doi: 10.5507/BP.2003.015.

- Unim, B. *et al.* (2022) 'Health data collection methods and procedures across EU member states: findings from the InfAct Joint Action on health information', *Archives of Public Health* 2021 80:1. BioMed Central, 80(1), pp. 17-. doi: 10.1186/S13690-021-00780-4.
- Weinmann, S. (2009) 'Metaanalysen zur psychopharmakotherapie: Garbage in garbage out?', *Psychiatrische Praxis*. © Georg Thieme Verlag KG Stuttgart · New York, 36(6), pp. 255–257. doi: 10.1055/S-0029-1220425/ID/6/BIB.
- Weston, L. E., Krein, S. L. and Harrod, M. (2021) 'Using observation to better understand the healthcare context', *Qualitative Research in Medicine & Healthcare*. Elsevier, 5(3), p. 9821. doi: 10.4081/QRMH.2021.9821.
- Yang, Z. *et al.* (2022) 'Defining health data elements under the HL7 development framework for metadata management', *Journal of Biomedical Semantics* 2022 13:1. BioMed Central, 13(1), pp. 10-. doi: 10.1186/S13326-022-00265-5.
- Yanto, B. E., Subali, B. and Suyanto, S. (2019) 'Measurement Instrument of Scientific Reasoning Test for Biology Education Students', *Distance Education*, 12(1), pp. 1383–1398. doi: 10.29333/IJI.2019.12188A.

BAB 4

PENYAJIAN DATA: TABEL, GRAFIK, DAN DIAGRAM

Oleh Resti Ariani

Penyajian data merupakan tahap penting dalam proses analisis biostatistika karena data yang telah dikumpulkan tidak akan memiliki makna ilmiah apabila tidak disusun, diringkas, dan ditampilkan secara sistematis. Dalam bidang kesehatan, data dapat berasal dari rekam medis, survei epidemiologi, hasil laboratorium, registri penyakit, penelitian klinis, pemantauan status gizi, surveilans penyakit menular, maupun evaluasi program kesehatan. Data tersebut sering kali berjumlah besar, beragam, dan kompleks sehingga diperlukan teknik penyajian yang mampu membantu pembaca memahami pola, kecenderungan, perbedaan, hubungan, serta kemungkinan masalah dalam data (WHO, 2021).

Penyajian data tidak hanya berfungsi memperindah laporan penelitian. Lebih dari itu, penyajian data merupakan instrumen komunikasi ilmiah. Tabel, grafik, dan diagram membantu peneliti menyampaikan informasi statistik secara ringkas, akurat, dan mudah ditafsirkan. Dalam laporan kesehatan masyarakat, visualisasi data juga berperan penting dalam mendukung pengambilan keputusan, misalnya menentukan prioritas intervensi, mengenali peningkatan kasus penyakit tertentu, membandingkan capaian

antarwilayah, atau mengevaluasi efektivitas suatu program. WHO (2021) menekankan bahwa visualisasi data yang baik sangat diperlukan dalam laporan kesehatan karena dapat membantu menyampaikan pesan kebijakan secara lebih meyakinkan dan mudah dipahami oleh pengambil keputusan.

Dalam konteks biostatistika, penyajian data yang baik harus memenuhi prinsip ketepatan, kejelasan, kesederhanaan, konsistensi, dan kejujuran ilmiah. Data tidak boleh ditampilkan secara menyesatkan, misalnya dengan memotong skala grafik tanpa penjelasan, menggunakan persentase tanpa menyebutkan jumlah dasar, menampilkan diagram yang tidak sesuai dengan jenis data, atau menyembunyikan informasi penting seperti data hilang (Bhakta *et al.*, 2025).

4.1 Tujuan Penyajian Data

Secara umum, penyajian data dalam biostatistika memiliki beberapa tujuan utama:

Pertama, penyajian data bertujuan untuk meringkas informasi agar data mentah yang semula panjang dan tidak teratur dapat dibaca secara lebih mudah. Misalnya, data umur 500 pasien tidak perlu ditampilkan satu per satu, tetapi dapat diringkas dalam bentuk rata-rata, median, rentang, simpangan baku, atau distribusi frekuensi menurut kelompok umur (Vickers *et al.*, 2020)

Kedua, penyajian data membantu menunjukkan pola atau kecenderungan. Dalam epidemiologi, grafik garis dapat menunjukkan tren kasus demam berdarah dari bulan ke bulan. Apabila terdapat kenaikan tajam pada musim hujan, pola tersebut dapat menjadi dasar bagi petugas kesehatan untuk meningkatkan kegiatan pencegahan (CDC, 2024).

Ketiga, penyajian data berguna untuk membandingkan kelompok. Misalnya, peneliti dapat membandingkan proporsi hipertensi antara laki-laki dan perempuan, rerata kadar hemoglobin pada kelompok intervensi dan kontrol, atau cakupan imunisasi antarwilayah.

Keempat, penyajian data dapat membantu menunjukkan hubungan antarvariabel. Misalnya, diagram pencar untuk menggambarkan hubungan antara indeks massa tubuh dan tekanan darah sistolik.

Kelima, penyajian data berfungsi sebagai alat komunikasi hasil penelitian dalam artikel ilmiah, laporan skripsi, tesis, disertasi, laporan rumah sakit, maupun *policy brief* kesehatan.

4.2 Prinsip Umum Penyajian Data

Penyajian data yang baik mengikuti prinsip-prinsip berikut (Krause *et al.*, 2024):

Prinsip	Penjelasan
Kejelasan	Setiap tabel/grafik harus mudah dipahami tanpa penjelasan panjang. Judul harus memuat variabel, populasi, lokasi, dan periode.
Ketepatan	Jenis penyajian harus sesuai jenis data. Data kategorik → diagram batang/tabel. Data numerik → histogram/boxplot/diagram pencar.
Kesederhanaan	Tampilkan informasi penting, bukan semua informasi. Data rinci bisa diletakkan di lampiran.

Prinsip	Penjelasan
Konsistensi	Format angka, satuan, simbol, warna, dan istilah harus seragam di seluruh dokumen.
Kejujuran	Mencerminkan keadaan data sebenarnya. Jelaskan jumlah sampel, data hilang, dan metode pengelompokan.
Aksesibilitas	Dapat dibaca oleh berbagai latar belakang, termasuk yang memiliki keterbatasan membedakan warna.

4.2.1 Penyajian Data dalam Bentuk Tabel

Tabel adalah bentuk penyajian data yang menggunakan susunan baris dan kolom. Tabel sangat berguna ketika peneliti ingin menyajikan angka secara rinci dan memungkinkan pembaca membandingkan nilai secara tepat (Divecha *et al.*, 2023). Dalam penelitian kesehatan, tabel sering digunakan untuk menampilkan karakteristik responden, distribusi penyakit, hasil pemeriksaan laboratorium, hubungan antarvariabel, serta hasil analisis statistik.

Unsur-Unsur Tabel

Tabel yang baik umumnya memiliki unsur berikut.

Unsur Tabel	Penjelasan
Nomor tabel	Memudahkan rujukan dalam teks, misalnya Tabel 1, Tabel 2, dan seterusnya.
Judul tabel	Menjelaskan isi tabel secara ringkas tetapi lengkap.

Unsur Tabel	Penjelasan
Kepala kolom	Menunjukkan nama variabel atau kategori data.
Badan tabel	Berisi angka, kategori, atau hasil perhitungan.
Satuan	Menjelaskan ukuran data, misalnya %, mg/dL, tahun, atau mmHg.
Catatan kaki	Menjelaskan singkatan, sumber data, uji statistik, atau keterangan khusus.
Sumber data	Dicantumkan apabila data berasal dari sumber sekunder.

Judul tabel harus ditempatkan di atas tabel. Judul tidak perlu terlalu panjang, tetapi harus memuat informasi yang cukup. Contoh judul yang baik adalah:

Tabel 4.1 Distribusi Responden Menurut Jenis Kelamin dan Kelompok Umur pada Penelitian Hipertensi di Puskesmas X Tahun 2026

Judul tersebut menjelaskan siapa respondennya, variabel yang ditampilkan, konteks penelitian, tempat, dan tahun.

1) Tabel Distribusi Frekuensi

Tabel distribusi frekuensi digunakan untuk menyajikan jumlah kasus atau responden dalam setiap kategori. Tabel ini lazim digunakan untuk data kategorik seperti jenis kelamin, tingkat pendidikan, status gizi, jenis penyakit, status merokok, atau status imunisasi. **Contoh:**

Tabel 4. 1 Distribusi Responden Menurut Status Merokok pada Penelitian Hipertensi di Puskesmas X Tahun 2026

Status Merokok	Frekuensi (n)	Persentase (%)
Tidak merokok	120	60,0
Perokok ringan	45	22,5
Perokok sedang	25	12,5
Perokok berat	10	5,0
Total	200	100,0

Tabel di atas menunjukkan bahwa mayoritas responden tidak merokok, yaitu 120 orang atau 60,0%. Dalam penulisan hasil, peneliti tidak perlu mengulang seluruh angka dalam tabel. Cukup jelaskan temuan yang paling penting, misalnya kategori terbesar, kategori terkecil, atau pola yang relevan dengan tujuan penelitian.

2) Tabel Distribusi Frekuensi Berkelompok

Pada data numerik yang memiliki banyak variasi nilai, seperti umur, tekanan darah, berat badan, tinggi badan, atau kadar glukosa darah, data dapat dikelompokkan ke dalam interval tertentu. Pengelompokan ini membuat data lebih ringkas dan mudah dibaca.

Contoh:

Tabel 4. 2 Distribusi Responden Berdasarkan Kelompok Umur

Kelompok Umur (tahun)	Frekuensi (n)	Persentase (%)
18-29	35	17,5
30-39	48	24,0
40-49	57	28,5
50-59	40	20,0
≥60	20	10,0
Total	200	100,0

Dalam membuat kelompok interval, peneliti harus memperhatikan dasar ilmiah dan konteks kesehatan. Pengelompokan umur dapat mengikuti kategori epidemiologis, kategori klinis, atau pedoman program kesehatan. Pengelompokan tidak boleh dibuat sembarangan karena dapat memengaruhi interpretasi.

3) Tabel Silang

Tabel silang digunakan untuk menyajikan hubungan antara dua variabel kategorik. Dalam biostatistika, tabel silang sering digunakan sebelum melakukan uji chi-square, uji Fisher, atau analisis asosiasi lainnya.

Contoh:**Tabel 4. 3 Hubungan Status Merokok dengan Kejadian Hipertensi**

Status Merokok	Hipertensi n (%)	Tidak Hipertensi n (%)	Total
Merokok	42 (56,0)	33 (44,0)	75
Tidak merokok	45 (36,0)	80 (64,0)	125
Total	87 (43,5)	113 (56,5)	200

Tabel tersebut menunjukkan bahwa proporsi hipertensi lebih tinggi pada kelompok perokok dibandingkan kelompok tidak merokok. Namun, tabel silang hanya memberikan gambaran awal. Untuk menyimpulkan apakah perbedaan tersebut bermakna secara statistik, diperlukan uji inferensial yang sesuai.

4) Tabel Ringkasan Statistik

Data numerik sering disajikan menggunakan ukuran pemusatan dan ukuran penyebaran. Pemilihan ukuran ringkasan bergantung pada distribusi data. Jika data berdistribusi normal, rerata dan simpangan baku dapat digunakan. Jika data tidak normal atau mengandung pencilan, median dan rentang interkuartil lebih tepat digunakan (Vickers *et al.*, 2020). **Contoh:**

Tabel 4. 4 Karakteristik Numerik Responden Penelitian

Variabel	Rerata ± SD	Median (IQR)	Minimum–Maksimum
Umur (tahun)	45,6 ± 12,4	46 (36–55)	18–72
Tekanan darah sistolik (mmHg)	132,8 ± 18,6	130 (120–145)	95–190
IMT (kg/m ²)	25,1 ± 4,2	24,7 (22,1–27,8)	17,5–38,2

Tabel seperti ini sangat berguna untuk menggambarkan karakteristik numerik responden. Namun, peneliti perlu berhati-hati agar tidak menampilkan terlalu banyak ukuran statistik yang tidak diperlukan. Jika tujuan hanya mendeskripsikan data, cukup gunakan ukuran yang paling sesuai dengan distribusi dan kebutuhan analisis.

5) Tabel Hasil Analisis Statistik

Tabel hasil analisis statistik digunakan untuk menyajikan hasil uji hubungan, perbedaan, regresi, atau model statistik lainnya. Dalam penelitian kesehatan, tabel hasil analisis sebaiknya tidak hanya menampilkan nilai p, tetapi juga ukuran efek seperti odds ratio, risk ratio, mean difference, koefisien regresi, dan interval kepercayaan. **Contoh:**

Tabel 4. 5 Hasil Analisis Faktor Risiko Kejadian Hipertensi

Variabel	OR	95% CI	Nilai p
Merokok	2,25	1,25–4,05	0,007
Obesitas	3,10	1,70–5,65	<0,001
Aktivitas fisik rendah	1,85	1,02–3,36	0,043

Tabel tersebut lebih informatif dibandingkan tabel yang hanya memuat nilai p. Nilai p menunjukkan bukti statistik terhadap hipotesis nol, sedangkan ukuran efek dan interval kepercayaan membantu pembaca menilai besar dan ketepatan estimasi. Dalam pelaporan uji klinis, CONSORT 2025 juga menekankan pentingnya pelaporan yang lengkap, termasuk alur peserta dan informasi hasil yang transparan.

4.3 Penyajian Data dalam Bentuk Grafik

Grafik adalah bentuk penyajian data secara visual. Grafik membantu pembaca memahami pola, tren, perbandingan, dan distribusi data dengan lebih cepat dibandingkan tabel (Divecha *et al.*, 2023). Namun, grafik yang baik harus dipilih berdasarkan jenis data dan tujuan komunikasi. CDC melalui COVE (*CDC Open-Source Visualization Editor*) memberikan panduan berbagai jenis visualisasi data kesehatan, termasuk kapan grafik tertentu sebaiknya digunakan dan kapan sebaiknya dihindari (Divecha *et al.*, 2023; CDC, 2024). Beberapa jenis grafik antara lain:

a) **Diagram Batang**

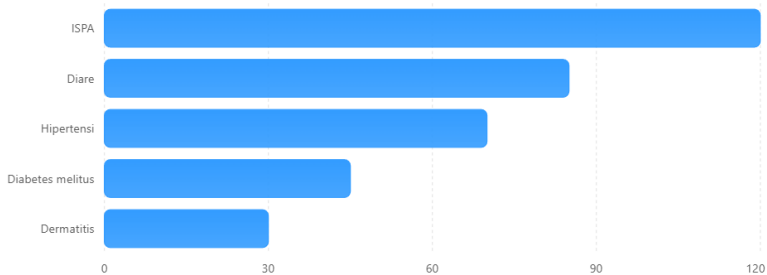
Diagram batang digunakan untuk membandingkan nilai antar kategori. Jenis grafik ini cocok untuk data kategorik, misalnya jumlah kasus menurut jenis penyakit, proporsi status gizi, cakupan imunisasi antarwilayah, atau jumlah pasien menurut jenis kelamin.

Contoh penggunaan

Jenis Penyakit	Jumlah Kasus
ISPA	120
Diare	85
Hipertensi	70
Diabetes melitus	45
Dermatitis	30

Diagram batang dapat dibuat secara vertikal atau horizontal. Diagram batang horizontal lebih baik jika nama kategori panjang atau jumlah kategori cukup banyak. CDC (2024) menyebutkan bahwa diagram batang berguna untuk membandingkan data antar kategori dan menunjukkan perubahan besar, tetapi kurang tepat untuk menunjukkan perubahan kecil secara bertahap.

Contoh: jumlah kasus penyakit di puskesmas



Dalam penulisan hasil, grafik tersebut dapat dijelaskan sebagai berikut: “Berdasarkan diagram batang, ISPA merupakan penyakit dengan jumlah kasus tertinggi, yaitu 120 kasus, sedangkan dermatitis merupakan penyakit dengan jumlah kasus terendah, yaitu 30 kasus.”

b) Histogram

Histogram digunakan untuk menggambarkan distribusi data numerik kontinu, seperti umur, kadar kolesterol, kadar hemoglobin, tekanan darah, atau lama rawat inap. Berbeda dengan diagram batang, batang pada histogram saling menempel karena menunjukkan interval nilai yang berurutan.

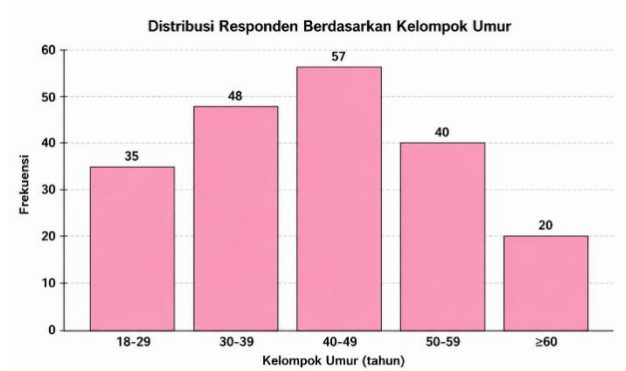
Histogram membantu peneliti mengenali bentuk distribusi data, apakah simetris, menceng ke kanan, menceng ke kiri, atau memiliki pencilan. Informasi ini penting sebelum memilih analisis statistik. Misalnya, data kadar trigliserida sering kali menceng ke kanan sehingga median dan IQR mungkin lebih tepat dibandingkan rerata dan simpangan baku. Rekomendasi penggunaan histogram:

- Sebaran umur responden.
- Melihat distribusi tekanan darah sistolik.
- Mengidentifikasi pencilan pada kadar glukosa darah.

- d. Menilai apakah data mendekati distribusi normal.

Contoh Data: Distribusi umur 200 responden

Kelompok Umur	Frekuensi
18-29	35
30-39	48
40-49	57
50-59	40
≥60	20



Keterangan: Distribusi umur responden right-skewes dengan kelompok terbanyak pada usia 40-49 tahun (57 orang)

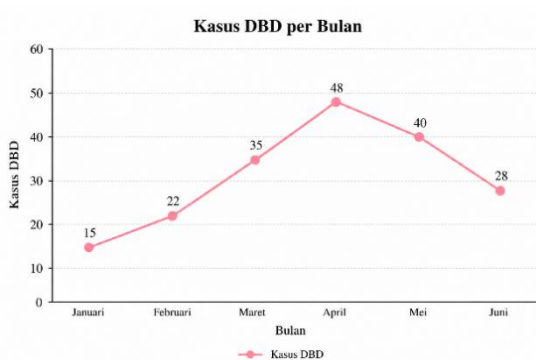
c) Grafik Garis (*Line Chart*)

Grafik garis digunakan untuk menunjukkan perubahan data menurut urutan waktu. Dalam kesehatan masyarakat, grafik garis sering digunakan untuk menggambarkan tren kasus penyakit, angka kematian, cakupan imunisasi,

jumlah kunjungan puskesmas, atau indikator surveilans dari waktu ke waktu.

Contoh:

Bulan	Kasus DBD
Januari	15
Februari	22
Maret	35
April	48
Mei	40
Juni	28



Keterangan: Kasus DBD meningkat dari Januari (15 kasus) hingga puncaknya di April (48 kasus), kemudian menurun pada Mei (40 kasus) dan Juni (28 kasus). Pola ini mengindikasikan adanya variasi musiman penyakit DBD.

Grafik garis dapat menunjukkan apakah kasus mengalami peningkatan, penurunan, fluktuasi musiman, atau lonjakan yang memerlukan tindakan

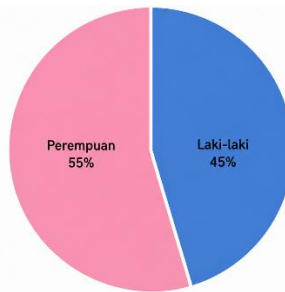
cepat (CDC, 2024). Dalam surveilans epidemiologi, grafik garis sangat berguna karena waktu merupakan dimensi utama dalam memahami penyebaran penyakit.

d) **Diagram Lingkaran (*Pie Chart*)**

Diagram lingkaran digunakan untuk menunjukkan komposisi bagian terhadap keseluruhan. Misalnya, proporsi jenis pembiayaan pasien, proporsi status gizi balita, atau distribusi responden menurut jenis kelamin. Namun, diagram lingkaran sebaiknya digunakan secara terbatas. Jika kategori terlalu banyak atau perbedaan antarbagian terlalu kecil, diagram lingkaran menjadi sulit dibaca.

Contoh penggunaan yang tepat:

Jenis Kelamin	Persentase
Laki-laki	45%
Perempuan	55%



Keterangan: Proporsi responden perempuan (55%) lebih tinggi dibandingkan laki-laki (45%).

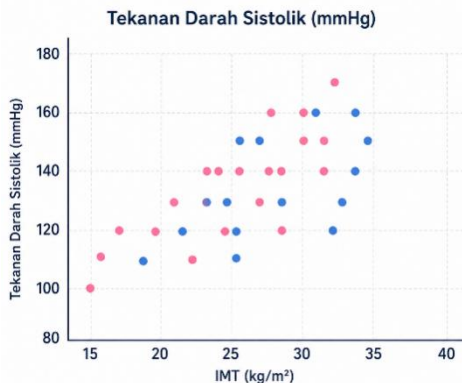
Diagram lingkaran dapat digunakan karena hanya terdapat dua kategori dan keduanya membentuk satu total yang jelas. Sebaliknya, jika terdapat delapan kategori penyakit dengan

persentase yang berdekatan, diagram batang lebih dianjurkan karena lebih mudah dibandingkan (Divecha *et al.*, 2023).

e) Diagram Pencar

Diagram pencar digunakan untuk menggambarkan hubungan antara dua variabel numerik. Setiap titik mewakili satu individu atau satu unit pengamatan. Diagram ini sering digunakan dalam analisis korelasi dan regresi. Contoh hubungan yang dapat ditampilkan dengan diagram pencar:

- Indeks massa tubuh dan tekanan darah sistolik.
- Umur dan kadar kolesterol total.
- Lama merokok dan kapasitas paru.
- Lama rawat inap dan biaya perawatan.



Keterangan: Terdapat kecenderungan positif antara IMT dan tekanan darah sistolik, artinya semakin tinggi IMT, cenderung semakin tinggi tekanan darah sistolik. Dua titik di pojok kanan atas menunjukkan kemungkinan pencilan (outlier).

Apabila titik-titik cenderung naik dari kiri bawah ke kanan atas, hubungan yang terlihat bersifat positif. Apabila titik-titik menurun dari kiri atas ke kanan bawah, hubungan yang terlihat

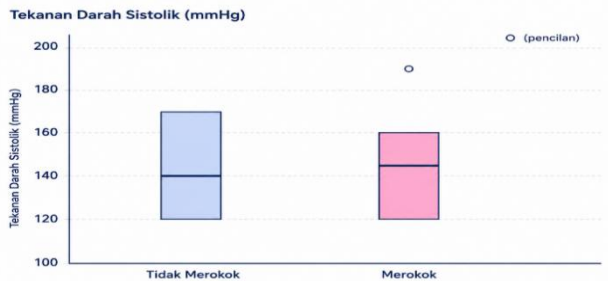
bersifat negatif. Namun, diagram pencar tidak membuktikan hubungan sebab-akibat. Analisis statistik dan pertimbangan desain penelitian tetap diperlukan (Hopewell *et al.*, 2025).

f) Boxplot

Boxplot atau diagram kotak-garis digunakan untuk menampilkan distribusi data numerik berdasarkan median, kuartil, rentang, dan pencilan. Grafik ini sangat berguna untuk membandingkan distribusi antar kelompok.

Contoh penggunaan:

- Membandingkan kadar hemoglobin antara kelompok anemia dan tidak anemia.
- Membandingkan lama rawat inap antara pasien dengan dan tanpa komplikasi.
- Membandingkan tekanan darah sistolik antar kelompok umur.
- Membandingkan skor kualitas hidup sebelum dan sesudah intervensi.



Gambar 4. 1 Boxplot Perbandingan Tekanan Darah Sistolik Berdasarkan Status Merokok

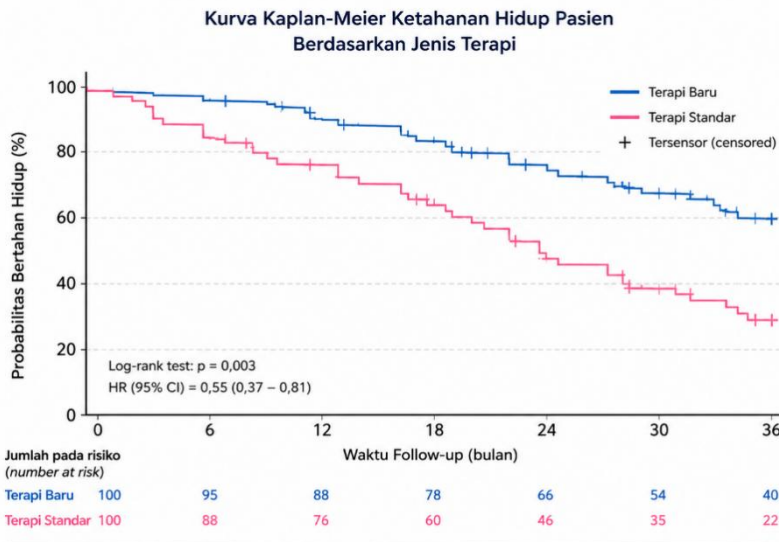
Keterangan: Kelompok perokok memiliki median tekanan darah yang lebih tinggi (sekitar 145 mmHg) dibandingkan bukan perokok (sekitar 125 mmHg).

Sebaran data kelompok perokok juga lebih lebar (variasi lebih besar).

Boxplot sangat berguna ketika data tidak berdistribusi normal atau mengandung pencilan. Dalam penelitian klinis, boxplot sering lebih informatif daripada hanya menampilkan rerata karena memberikan gambaran tentang penyebaran dan variasi data (Divecha *et al.*, 2023).

g) Grafik Kaplan-Meier

Dalam penelitian klinis dan epidemiologi, grafik Kaplan-Meier digunakan untuk menyajikan data survival atau waktu sampai terjadinya suatu peristiwa. Peristiwa tersebut dapat berupa kematian, kekambuhan penyakit, kesembuhan, kegagalan terapi, atau kejadian komplikasi. **Contoh:**



Keterangan: Kurva Kaplan-Meier menunjukkan probabilitas bertahan hidup kumulatif selama 36

bulan. Kelompok terapi baru memiliki ketahanan hidup yang lebih baik dibandingkan terapi standar.

Grafik Kaplan-Meier banyak digunakan dalam penelitian kanker, penyakit jantung, penyakit infeksi, dan uji klinis. Grafik ini menunjukkan probabilitas bertahan dari suatu peristiwa selama periode tertentu. Dalam laporan penelitian, grafik Kaplan-Meier sebaiknya disertai jumlah subjek yang masih berisiko pada beberapa titik waktu agar pembaca dapat menilai kestabilan estimasi (Hopewell *et al.*, 2025).

4.4 Penyajian Data dalam Bentuk Diagram

Istilah diagram sering digunakan secara luas untuk menggambarkan hubungan, proses, alur, struktur, atau klasifikasi. Dalam buku biostatistika untuk kesehatan, diagram penting karena tidak semua informasi dapat disampaikan secara optimal melalui angka. Beberapa informasi lebih tepat disajikan sebagai alur, bagan, atau skema.

a) Diagram Alir Penelitian

Diagram alir digunakan untuk menunjukkan tahapan penelitian, mulai dari populasi awal, proses seleksi sampel, kriteria inklusi dan eksklusi, jumlah responden yang dianalisis, hingga jumlah responden yang keluar dari penelitian. Dalam uji klinis acak, diagram alir CONSORT merupakan standar penting untuk menunjukkan perjalanan peserta penelitian dari tahap pendaftaran hingga analisis. CONSORT 2025 secara eksplisit mencakup checklist dan diagram untuk mendokumentasikan alur peserta dalam uji acak.

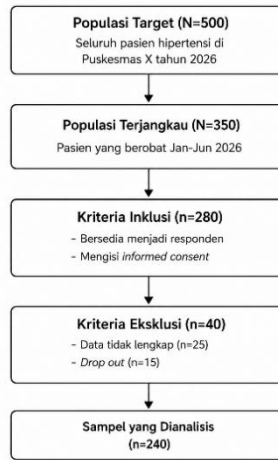
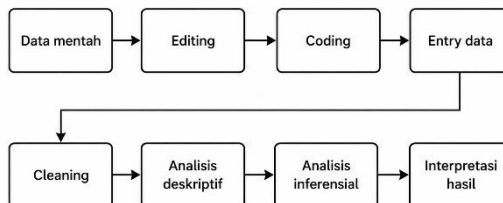


Diagram ini sangat dianjurkan untuk skripsi, tesis, disertasi, artikel observasional, dan uji klinis karena memudahkan pembaca menilai proses pemilihan sampel.

b) Diagram Alir Analisis Data

Diagram alir analisis data menggambarkan proses pengolahan data, mulai dari pengumpulan data, pemeriksaan kelengkapan, pembersihan data, pengkodean, analisis deskriptif, analisis bivariat, hingga analisis multivariat.

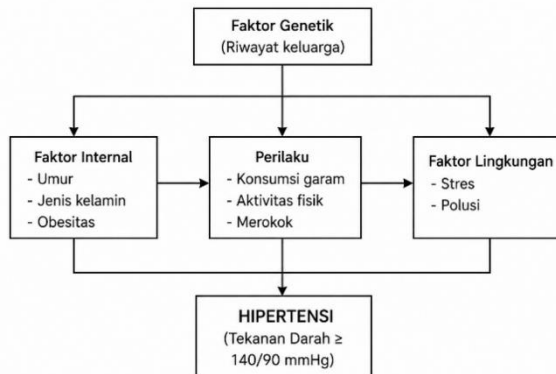
Contoh alur:



c) Diagram Konseptual

Diagram konseptual digunakan untuk menggambarkan hubungan antarvariabel berdasarkan teori. Misalnya, penelitian tentang faktor risiko hipertensi dapat menampilkan variabel umur, jenis kelamin, obesitas, konsumsi garam, aktivitas fisik, stres, dan riwayat keluarga sebagai faktor yang berhubungan dengan hipertensi.

Diagram konseptual membantu pembaca memahami kerangka pikir penelitian. Dalam analisis biostatistika, diagram ini juga dapat membantu menentukan variabel bebas, variabel terikat, variabel perancu, mediator, dan moderator.



d) Diagram Sebab-Akibat

Diagram sebab-akibat, termasuk diagram tulang ikan, dapat digunakan dalam penelitian manajemen kesehatan, keselamatan pasien, mutu pelayanan, atau investigasi kejadian tidak diharapkan. Diagram ini membantu mengelompokkan penyebab masalah ke dalam beberapa kategori, misalnya manusia, metode, lingkungan, alat, kebijakan, dan material.

Contoh penggunaan dalam kesehatan:

Masalah utama: meningkatnya angka infeksi luka operasi.

Kemungkinan penyebab: sterilitas alat, kepatuhan cuci tangan, kondisi ruang operasi, penggunaan antibiotik profilaksis, kondisi pasien, dan teknik operasi.

e) **Diagram Peta atau *Choropleth Map***

Peta tematik digunakan untuk menggambarkan distribusi geografis indikator kesehatan, misalnya angka kejadian DBD antar kecamatan, cakupan imunisasi per kabupaten, atau prevalensi stunting antarprovinsi. Peta sangat berguna ketika lokasi menjadi komponen penting dalam interpretasi data. Namun, penggunaan peta harus hati-hati. Perbedaan warna dapat menimbulkan kesan dramatis jika klasifikasi kategori tidak dijelaskan. Oleh karena itu, legenda, sumber data, satuan, dan metode klasifikasi harus ditampilkan secara jelas.

4.5 Bentuk Penyajian Berdasarkan Jenis Data

Pemilihan tabel, grafik, atau diagram perlu disesuaikan dengan jenis data dan tujuan analisis.

Jenis Data atau Tujuan	Bentuk Penyajian yang Dianjurkan	Contoh dalam Kesehatan
Data kategorik satu variabel	Tabel frekuensi, diagram batang, diagram lingkaran terbatas	Jenis kelamin, status gizi
Data numerik satu variabel	Tabel ringkasan, histogram, boxplot	Umur, tekanan darah
Membandingkan kategori	Diagram batang, tabel silang	Kasus DBD menurut kecamatan

Jenis Data atau Tujuan	Bentuk Penyajian yang Dianjurkan	Contoh dalam Kesehatan
Menunjukkan tren waktu	Grafik garis	Kasus influenza per bulan
Hubungan dua variabel numerik	Diagram pencar	IMT dan tekanan darah
Komposisi bagian	Diagram lingkaran atau stacked bar	Proporsi pembiayaan pasien
Alur penelitian	Diagram alir	Seleksi sampel
Distribusi wilayah	Peta tematik	Prevalensi stunting per kabupaten
Data survival	Kurva Kaplan-Meier	Waktu bebas kekambuhan

Tabel tersebut dapat menjadi pedoman praktis bagi mahasiswa dan peneliti kesehatan dalam memilih bentuk penyajian data yang sesuai. Kesalahan umum yang sering terjadi adalah menggunakan grafik yang tampak menarik tetapi tidak sesuai dengan karakter data.

4.6 Kesalahan Umum dalam Penyajian Data

Beberapa kesalahan yang sering ditemukan dalam laporan penelitian kesehatan antara lain sebagai berikut.

- Judul tabel atau grafik tidak informatif.** Judul seperti “Distribusi Responden” tidak cukup menjelaskan isi tabel. Judul sebaiknya memuat variabel, populasi, tempat, dan waktu jika diperlukan.

- 2) **Persentase ditampilkan tanpa jumlah dasar.** Misalnya, peneliti menulis “70% responden mengalami anemia” tanpa menyebutkan jumlah responden. Persentase harus disertai n agar pembaca mengetahui dasar perhitungan.
- 3) **Grafik terlalu ramai.** Grafik dengan terlalu banyak warna, label, garis, dan kategori dapat mengaburkan pesan utama. Visualisasi yang baik bukan yang paling dekoratif, tetapi yang paling komunikatif.
- 4) **Skala grafik menyesatkan.** Memotong sumbu y tanpa keterangan dapat membuat perbedaan kecil tampak sangat besar. Jika sumbu dipotong, peneliti harus memberi penjelasan yang transparan.
- 5) **Menggunakan diagram lingkaran untuk terlalu banyak kategori.** Diagram lingkaran sulit dibaca jika memiliki banyak bagian atau jika perbedaan antarbagian kecil. Dalam kondisi demikian, diagram batang lebih tepat.
- 6) **Tidak menjelaskan singkatan dan satuan.** Singkatan seperti IMT, Hb, GDP, LDL, OR, RR, dan CI harus dijelaskan, terutama jika pembaca berasal dari latar belakang yang beragam.
- 7) **Mengandalkan nilai p secara berlebihan.** Hasil statistik sebaiknya tidak hanya menampilkan nilai p , tetapi juga ukuran efek dan interval kepercayaan (Bhakta *et al.*, 2025). Hal ini penting karena signifikansi statistik tidak selalu sama dengan signifikansi klinis.
- 8) **Tidak melaporkan data hilang.** Dalam penelitian kesehatan, data hilang dapat memengaruhi interpretasi hasil. Data hilang sebaiknya dijelaskan dalam tabel karakteristik, catatan kaki, atau bagian metode analisis (Dewidar *et al.*, 2025; Bhakta *et al.*, 2025).

Contoh Narasi Penyajian Data dalam Laporan Penelitian

Dalam laporan penelitian, tabel dan grafik tidak boleh dibiarkan “berbicara sendiri” tanpa penjelasan. Penulis perlu memberikan narasi yang menyoroti temuan utama.

Contoh narasi untuk tabel karakteristik responden:

Berdasarkan Tabel 1, sebagian besar responden berada pada kelompok umur 40–49 tahun, yaitu sebanyak 57 orang atau 28,5%. Responden paling sedikit berada pada kelompok umur ≥ 60 tahun, yaitu sebanyak 20 orang atau 10,0%. Temuan ini menunjukkan bahwa responden penelitian didominasi oleh kelompok usia dewasa pertengahan, yang secara epidemiologis mulai memiliki risiko lebih tinggi terhadap penyakit degeneratif seperti hipertensi dan diabetes melitus.

Contoh narasi untuk grafik tren kasus:

Grafik 1 menunjukkan adanya peningkatan kasus DBD dari Januari hingga April, dengan jumlah kasus tertinggi terjadi pada bulan April. Setelah itu, jumlah kasus mulai menurun pada Mei dan Juni. Pola ini dapat mengindikasikan adanya variasi musiman yang perlu diperhatikan dalam perencanaan program pengendalian vektor.

Contoh narasi untuk tabel hasil regresi:

Hasil analisis regresi logistik menunjukkan bahwa obesitas berhubungan dengan kejadian hipertensi dengan OR sebesar 3,10 dan 95% CI 1,70–5,65. Hal ini berarti responden dengan obesitas memiliki peluang lebih tinggi untuk mengalami hipertensi dibandingkan responden yang tidak obesitas. Karena interval kepercayaan tidak melewati angka 1 dan nilai $p < 0,001$, hubungan tersebut bermakna secara statistik.

4.7 Etik dalam Penyajian Data Kesehatan

Data kesehatan sering kali berkaitan dengan informasi individu yang bersifat sensitif. Oleh karena itu, penyajian data harus memperhatikan prinsip kerahasiaan dan perlindungan subjek. Identitas pasien tidak boleh ditampilkan dalam tabel atau grafik. Jika jumlah kasus sangat kecil pada kelompok tertentu, peneliti perlu berhati-hati agar individu tidak dapat dikenali secara tidak langsung.

Selain itu, penyajian data harus menghindari manipulasi visual. Grafik yang dibuat untuk memperkuat argumen tidak boleh mengubah persepsi pembaca secara keliru (Bhakta *et al.*, 2025). Dalam konteks kesehatan masyarakat, visualisasi yang menyesatkan dapat berdampak serius karena dapat memengaruhi kebijakan, alokasi sumber daya, dan persepsi risiko Masyarakat.

Prinsip lain yang perlu diperhatikan adalah keadilan representasi. Data kelompok rentan tidak boleh disembunyikan hanya karena jumlahnya kecil. Namun, penyajiannya harus tetap memperhatikan privasi, konteks sosial, dan potensi stigma. Dalam penelitian kesehatan modern, pelaporan data juga semakin diarahkan untuk memperhatikan inklusivitas dan kesetaraan kesehatan, sebagaimana terlihat dalam pengembangan STROBE-Equity (Dewidar *et al.*, 2025).

DAFTAR PUSTAKA

- Bhakta, S., Xie, J., Pea, F., & Gillespie, S. H. (2025). Data transparency and reproducibility in health research: bridging the gap for early-career researchers. *Frontiers in Antibiotics*, 4. <https://doi.org/10.3389/frabi.2025.1562002>
- CDC (2024) *COVE: Bar Chart and Building Dashboards*. Available at: <https://www.cdc.gov/cove/> (Accessed: 20 June 2026)
- Dewidar, O., Shamseer, L., Melendez-Torres, G. J., Akl, E. A., Ramke, J., Wang, X., Oloyede, O., Young, T., Nicholls, S. G., Marshall, Z., Kennedy, M., Hardy, B.-J., Rizvi, A., Ghogomu, E., Rader, T., Waddington, H. S., Shea, B., Nkangu, M., Ellingwood, H., ... Welch, V. (2025). Improving the reporting on health equity in observational research (STROBE-Equity): extension checklist and elaboration. *BMJ*, 390, e083882. <https://doi.org/10.1136/bmj-2024-083882>
- Divecha, C., Tullu, M., & Karande, S. (2023). Utilizing tables, figures, charts and graphs to enhance the readability of a research paper. *Journal of Postgraduate Medicine*, 69(3), 125–131. https://doi.org/10.4103/jpgm.jpgm_387_23
- Hopewell, S., Chan, A.-W., Collins, G. S., Hróbjartsson, A., Moher, D., Schulz, K. F., Tunn, R., Aggarwal, R., Berkwitz, M., Berlin, J. A., Bhandari, N., Butcher, N. J., Campbell, M. K., Chidebe, R. C. W., Elbourne, D., Farmer, A., Fergusson, D. A., Golub, R. M., Goodman, S. N., ... Boutron, I. (2025). CONSORT 2025 statement: updated guideline for reporting randomised trials. *BMJ*, 389, e081123. <https://doi.org/10.1136/bmj-2024-081123>

- Krause, A., Rennie, N., & Tarran, B. (2024). *Best Practices for Data Visualisation: Styling charts for accessibility*. Royal Statistical Society. <https://royal-statistical-society.github.io/datavisguide/docs/styling.html>
- Vickers, A. J., Assel, M. J., Sjoberg, D. D., Qin, R., Zhao, Z., Koyama, T., Botchway, A., Wang, X., Huo, D., Kattan, M., Zabor, E. C., & Harrell, F. (2020). Guidelines for Reporting of Figures and Tables for Clinical Research in Urology. *Journal of Urology*, 204(1), 121–133. <https://doi.org/10.1097/JU.0000000000001096>

BAB 5

UKURAN PEMUSATAN DAN PENYEBARAN DATA

Oleh Suryani Mansyur

Pada penelitian kesehatan, sering kali peneliti memperoleh banyak data mentah di lapangan yang masih sulit difahami. Hal ini menjadikan statistik deskriptif sebagai salah satu cara analisis yang sederhana yang dibutuhkan dalam meringkas data atau menggambarkan data observasi menjadi angka yang bermakna yang mewakili karakteristik utama data. Ukuran yang biasa digunakan untuk meringkas data tersebut adalah ukuran pemusatan atau yang biasa disebut tendensi sentral (*central tendency*) berupa *mean*, *median*, dan *modus* serta ukuran penyebaran data atau yang biasa disebut variabilitas berupa jangkauan (*range*), keragaman (*variance*), simpangan baku (*standard deviation*), dan koefisien keragaman (*coefficient of variation*).

5.1 Ukuran Pemusatan

Ukuran pemusatan data digunakan untuk mengetahui titik pusat sekumpulan data yang memberikan gambaran mengenai lokasi berkumpulnya data. (Kementrian Pendidikan Tinggi, 2020). Ukuran ini diperlukan untuk menyatakan keterwakilan data dari sekumpulan data yang diamati. Ukuran pemusatan tersebut di antaranya rata-rata (*mean*), nilai tengah (*median*) dan modus.

5.1.1 Mean

Mean adalah nilai rerata dari data (berupa skor atau nilai numerik) yang diperoleh dari pengumpulan data yang besarannya bersifat kuantitas dan tidak bervariasi. Rata-rata (mean) diperoleh dari menghitung jumlah semua data yang dibagi dengan banyaknya data. (Riyanto, 2013)

Pada data tunggal (belum dikelompokkan), rumus rata-rata ($\bar{x}$) adalah

$$\bar{x} = \frac{\sum_{i=1}^n X_i}{n}$$

Keterangan:

$\bar{x}$ = rerata (mean)

$\sum_{i=1}^n X_i$ = nilai data X_1, X_2, X_3 , hingga X_n

n = banyaknya data

Contoh soal:

Hitunglah rata-rata lama rawat (hari) 6 pasien demam tifoid jika masing-masing pasien memiliki lama rawat sebagai berikut:

5 4 6 8 6 7

Jawab:

$$\bar{x} = \frac{\sum_{i=1}^n X_i}{n} = \frac{5+4+6+8+6+7}{6} = \frac{36}{6} = 6$$

Interpretasi: Rata-rata lama rawat 6 pasien demam tifoid adalah 6 hari.

Pada data berkelompok, rumus rata-rata hitung :

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

Keterangan:

f_i = frekuensi yang sesuai dengan tanda kelas ke-i

M_i = nilai tengah (mid point) setiap interval kelas

Contoh soal:

Hitunglah rerata jika diketahui distribusi frekuensi berat badan responden pada tabel 1, sebagai berikut:

Tabel 1

Distribusi responden berdasarkan berat badan

Berat Badan (kg)	Frekuensi (f)
46 - 50	3
51 - 55	10
56 - 60	8
61 - 65	5
66 - 70	4
jumlah	30

Jawab: Untuk mempermudah menjawabnya, dianjurkan membuat tabel bantu

Berat Badan (kg)	Frekuensi (f)	Mid point (M_i)	$f_i M_i$
46 - 50	3	48	144
51 - 55	10	53	530
56 - 60	8	58	464
61 - 65	5	63	315
66 - 70	4	68	272
jumlah	30		1725

$$\bar{x} = \frac{\sum f_i M_i}{n} = \frac{1725}{30} = 57,5$$

Rerata berat badan adalah 57,5 kg.

Nilai rerata baik digunakan apabila jenis datanya adalah numerik (interval atau rasio), sebaran atau distribusi datanya simetris dan tidak terdapat nilai pencilan (*outlier*).

Kelebihan mean :

- Melibatkan seluruh data dalam setiap perhitungannya
- Mudah dihitung
- Nilai ini banyak dikenal dan tidak asing bagi siapapun

Kekurangan mean:

- Sangat dipengaruhi oleh nilai ekstrem
- Untuk menghitungnya, harus diketahui besaran (nilai) setiap pengamatan.

5.1.2 Median

Nilai yang membagi data terurut menjadi dua bagian yang sama besar didefinisikan sebagai median (Me) atau kuartil kedua (Q_2). Oleh karena itu, nilainya bergantung pada jumlah pengamatan dan bukan pada nilai-nilai ekstrem; sehingga, nilai ini cocok untuk data yang tidak berdistribusi normal atau bersifat asimetris.

Rumus nilai median untuk data tunggal : (Kementrian Pendidikan Tinggi, 2020).

Data diurut dari data terkecil ke terbesar, kemudian

- Untuk Data ganjil, $Me = X_k$, dimana $k = \frac{1}{2}(n+1)$
k adalah letak data
n adalah banyaknya data

- b. Untuk data genap, $Me = \frac{X_k + X_{k+1}}{2}$, dimana $k = n/2$

Contoh soal :

- a. Data ganjil

Diketahui umur pasien : 20 23 26 22 21

Hitunglah nilai mediannya !

Jawab :

1. Urutkan data terkecil hingga data terbesar :
20 21 22 23 26
2. Cari posisi median, $Me = X_k$
 $k = \frac{1}{2} (5+1) = 3$ (posisi data pada data ke-3)
 Median (Me) = 22
 Median umur pasien adalah 22 tahun

- b. Data genap

Diketahui data berat badan pasien:

50 60 45 65 50 55

Hitung nilai median !

Jawab:

Urutkan data : 45 50 50 55 60 65

Posisi median, $Me = \frac{X_k + X_{k+1}}{2}$

$k = n/2 = 6/2 = 3$

$$Me = \frac{X_k + X_{k+1}}{2} = \frac{X_3 + X_{3+1}}{2} = \frac{50 + 55}{2} = 52,5$$

Pada data berkelompok, rumus median (Me), adalah:

Median data berkelompok, dapat ditentukan oleh langkah-langkah:

- Cari posisi median, $Me = \frac{1}{2} n$

- temukan nilai sebelum median tercapai
- frekuensi kumulatif nilai sebelum median tercapai
- interval kelas
- frekuensi kelas dimana nilai median berada
- Hitung nilai median menggunakan rumus:

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

Ket : Me= median

b = batas bawah kelas median

p = panjang kelas median

n =banyaknya pengamatan

Fkum= jumlah semua frekuensi sebelum kelas median

f = frekuensi kelas median

(Luknis, 2008)

Contoh:

Hitung nilai median dari data berikut :

Tabel 2

Distribusi Responden berdasarkan berat badan

Berat Badan (kg)	Frekuensi (f)
46 - 50	3
51 - 55	10
56 - 60	8
61 - 65	5
66 - 70	4
jumlah	30

Jawab:

- Tentukanposisi median, dengan rumus , $n/2 = 30/2 = 15$

- Tabel bantu:

Berat Badan (kg)	Frekuensi (f)	Fkum
46 - 50	3	3
51 - 55	10	13
56 - 60	8	21
61 - 65	5	26
66 - 70	4	30
jumlah	30	

Berdasarkan tabel tersebut, diketahui bahwa kelas median berada pada kelas ke-3 (kelas dengan interval 56 – 60). Karena posisi median, 15 berada di antara nilai 13 – 21.

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

$$Me = 55,5 + 5 \left(\frac{\frac{1}{2}(30) - 13}{8} \right) = 55,5 + 5 \left(\frac{15 - 13}{8} \right)$$

$$Me = 55,5 + 5 \left(\frac{2}{8} \right) = 55,5 + 1,25 = 56,75$$

5.1.3 Modus

Modus adalah nilai yang sering muncul (frekuensi terbanyak). Pada data tunggal maupun data berkelompok, nilai modus (M_o) dicari dengan cara mencari nilai yang sering muncul dalam kelompok data.

Modus tidak dipengaruhi oleh nilai ekstrem atau pencilan (*outlier*). Modus sangat berguna dalam menentukan titik pusat data, khususnya untuk data nominal dan ordinal. (Ratnaningrum, 2021)

Modus untuk data tunggal

Contoh:

Diketahui data berat badan pasien DM di posbindu Mekar Sari, diperoleh data sebagai berikut: 40, 60, 55, 56, 60, 62, 60, 65, 40, 55.

Tentukan modus data tersebut!

Jawab: Nilai modus yaitu 60 karena muncul 3 kali. Artinya pasien DM di posbindu Mekar Sari paling banyak mempunyai berat badan 60 Kg.

Modus untuk data berkelompok

Pada distribusi frekuensi data berkelompok, kelas modus adalah kelas dengan frekuensi terbesar.

Rumus Modus (M_o) :

$$M_o = B_{Mo} + \left(\frac{d_1}{d_1 + d_2} \right) \times i$$

Keterangan:

M_o = *modus*

B_{Mo} = tepi bawah kelas pada lokasi letak modus

d_1 = selisih antara frekuensi kelas modus dengan kelas tepat di bawahnya

d_2 = selisih antara frekuensi kelas modus dengan kelas tepat di atasnya

i = panjang kelas modus

Contoh : data berat badan pasien DM di posbindu Mekar Sari

Berat badan (kg)	Frekuensi (f)	Tepi kelas
		39,5
40 - 49	2	49,5
50 - 59	3	59,5
60 - 69	5	69,5

M_o = pada frekuensi 5 (karena yang terbanyak)

$$B_{M_o} = 59,5$$

$$d_1 = 5 - 3 = 2$$

$$d_2 = 5 - 0 = 5$$

$$i = 10$$

$$M_o = 59,5 + \left(\frac{2}{2+5} \right) \times 10$$

$$M_o = 59,5 + 2,857$$

$$M_o = 62,357$$

Jadi, modus berat badan pasien DM adalah 62,357 kg atau 62,3 kg.

5.2 Ukuran Penyebaran Data (ukuran variasi / dispersi)

5.2.1 Range (jangkauan)

Range (jangkauan) merupakan data terbesar atau tertinggi dikurangi data terendah/ terkecil. Digunakan untuk skala interval dan rasio dari variabel kontinu.

Kelemahan: Tidak dapat digunakan jika terdapat data ekstrim. (Hasbullah, 2010)

Rumus: $R = \text{data tertinggi} - \text{data endah}$

Contoh:

Data tekanan darah sistole masyarakat kota:

140 120 150 140 160 160

Data tekanan darah sistole masyarakat desa:

120 140 140 120 120 150

Range tekanan darah sistole masyarakat kota = $160 - 120 = 40$

Range tekanan darah sistole masyarakat desa = $150 - 120 = 30$

Kesimpulan:

Berdasarkan data tersebut terlihat bahwa tekanan darah sistole masyarakat kota lebih bervariasi dibandingkan dengan tekanan darah sistole masyarakat desa. (Karena rangenya lebih besar).

5.2.2 Varians (Keragaman)

Varians adalah ukuran penyebaran berdasarkan deviasi atau perbedaan nilai pengukuran dari nilai tengahnya, biasanya dinilai dengan mean. Varians dalam data kelompok dihitung dengan menggunakan frekuensi relatif, dengan menghitung selisih dari setiap nilai dengan mean, kemudian mengkuadratkan selisih ini, lalu membaginya dengan jumlah nilai dalam data. Varians merupakan kuadrat dari simpangan baku. Kegunaannya untuk mengetahui tingkat penyebaran atau variasi data. Simbol varians adalah S^2 . (Hastono, 2008)

- a. varians data tunggal

$$S^2 = \frac{\sum (X_i - \bar{x})^2}{n-1} \quad \dots \text{ untuk sampel kecil } (n \leq 30)$$

Keterangan:

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

X_i = nilai setiap sampel

n = banyaknya sampel

contoh : Hitunglah varians dari data : 7 11 8 13
3 12

jawab:

$$\bar{x} = \frac{7+11+8+13+3+12}{6} = 9$$

$$S^2 = \frac{(7-9)^2 + (11-9)^2 + (8-9)^2 + (13-9)^2 + (3-9)^2 + (12-9)^2}{6-1}$$

$$S^2 = \frac{4+4+1+16+36+9}{5} = 14$$

Jadi, varians (S^2) = 14

Namun, Untuk sampel besar ($n > 30$)

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

- b. varians data berkelompok

$$S^2 = \frac{\sum f_i (M_i - \bar{x})^2}{n-1} \quad \dots \text{ jika } (n < 30)$$

Keterangan : f_i = frekuensi kelas interval ke- i

M_i = Nilai tengah kelas interval ke- i

n = jumlah data

Tabel 3

Distribusi responden berdasarkan berat badan di kota X tahun 2025

Berat Badan (kg)	Frekuensi (f)
3 - 4	2
4 - 5	7
5 - 6	10
6 - 7	5
7 - 8	1
jumlah	25

Diketahui mean= 5,3 kg. tentukan variansnya!

Jawab:

Distribusi responden berdasarkan berat badan di kota X tahun 2025

Berat badan	fi	Mi	Mi fi	$(Mi - \bar{x})$	$(Mi - \bar{x})^2$	$(Mi - \bar{x})^2 fi$
3 - 4	2	3,5	7	-1,8	3,24	6,48
4 - 5	7	4,5	31,5	-0,8	0,64	4,48
5 - 6	10	5,5	55	0,2	0,04	0,4
6 - 7	5	6,5	32,5	1,2	1,44	7,2
7 - 8	1	7,5	7,5	2,2	4,84	4,84
	25		133,5			22,84

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

$$S^2 = \frac{22,84}{25-1} = \mathbf{0,95}$$

Untuk sampel besar ($n > 30$)

$$S^2 = \frac{\sum fi (Mi - \bar{x})^2}{n}$$

Contoh:

Tabel 4

Distribusi responden berdasarkan berat badan di kota X tahun 2025

Berat Badan (kg)	Frekuensi (f)
35 - 39	3
40 - 44	6
45 - 49	7
50 - 54	13
55 - 59	9
60 - 64	8
65 - 69	4
jumlah	50

Hitung varians datanya!

Jawab:

Distribusi responden berdasarkan berat badan di kota X tahun 2025

Berat badan	fi	Mi	Mi fi	$(Mi - \bar{x})$	$(Mi - \bar{x})^2$	$(Mi - \bar{x})^2 fi$
35 - 39	3	37	111	-15,9	252,81	758,43
40 - 44	6	42	252	-10,9	118,81	712,86
45 - 49	7	47	329	-5,9	34,81	243,67
50 - 54	13	52	676	-0,9	0,81	10,53
55 - 59	9	57	513	4,1	16,81	151,29
60 - 64	8	62	496	9,1	82,81	662,48
65 - 69	4	67	268	14,1	198,81	795,24
	50		2645		705,67	3334,5

$$\bar{x} = \frac{Mi.fi}{fi} = \frac{2645}{50} = 52,9$$

$$S^2 = \frac{\sum fi (Mi - \bar{x})^2}{n} = \frac{3334,5}{50} = 66,69$$

Varians berat badan = 66,69 kg.

5.2.3 Standar Deviasi (Simpangan Baku)

Standar deviasi atau simpangan baku (s) memberikan gambaran tentang variasi atau penyebaran data di sekitar nilai rata-rata. Semakin besar nilai standar deviasinya, semakin banyak variasi data yang dimiliki. Simbol standar deviasi adalah S. (Pandriadi, 2023)

- a. standar deviasi untuk data tunggal
untuk sampel kecil ($n < 30$)

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

Untuk sampel besar ($n \geq 30$)

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

Contoh :

Tentukan standar deviasi pada contoh soal varians data tunggal sebelumnya !

Jawab:

Dikethui data 7 11 8 13 3 12

Varians (S^2)= 14

Standar deviasi (S) = $\sqrt{14} = 7$

b. standar deviasi untuk data berkelompok

$$S = \sqrt{\frac{\sum fi (Mi - \bar{x})^2}{n-1}} \quad \text{.. jika sampel kecil (n < 30)}$$

Atau

$$S = \sqrt{\frac{\sum fi (Mi - \bar{x})^2}{n}} \quad \text{.. jika sampel besar (n} \geq 30 \text{)}$$

Dengan kata lain , standar deviasi adalah akar dari varians , $S = \sqrt{S^2}$

5.2.4 Koefisien Keragaman

Koefisien keragaman merupakan rasio dari standar deviasi terhadap nilai mean dan dibuat dalam bentuk persentase. Tujuan dari koefisien keragaman adalah untuk membandingkan tingkat variasi anatra dua kelompok data yang memiliki skala yang berbeda. Keterbatasannya adalah hanya relevan bila rata- rata data tidak mendekati nol. (Riyanto, 2013)

$$\text{Rumus, } KV = \frac{S}{\bar{x}} \times 100\%$$

Keterangan: KV = koefisien keragaman

S = Standar deviasi (Simpangan baku)

$\bar{x}$ = rata- rata

Dengan kriteria penilaian:

$KV < 15\%$, KV rendah, menunjukkan data tidak bervariasi.

KV 15%–30%, KV sedang, menunjukkan variasi data ada dan terkendali

$KV > 30\%$, KV tinggi, data sangat bervariasi.

Contoh:

Tinggi rata-rata (mean) dari suatu kelompok pemuda yg berumur 20 tahun = 155,5 cm dengan standar deviasi = 9,5 cm. Dan rata-rata berat badan = 4,98 Kg. dengan standar deviasi = 3,5 Kg

Data mana yg lebih bervariasi?

Jawab:

Diketahui, $S_T = 9,5$ cm dan $X_T = 155,5$ cm

$S_B = 3,5$ kg dan $X_B = 49,8$ Kg

$KV_T = 9,5 / 155,5 \times 100\% = 6,1\%$ dari meannya

$KV_B = 3,5 / 49,8 \times 100\% = 7\%$ dari meannya.

Karena $CV_B > CV_T$, maka data berat pemuda lebih berpencah daripada tingginya.

DAFTAR PUSTAKA

- Hasbullah. 2010. *Modul Biostatistika Deskriptif*. Ternate: Program Studi Kesehatan Masyarakat Fakultas Ilmu Kesehatan. Universitas Muhammadiyah Maluku Utara.
- Kementerian Pendidikan Tinggi, Sains dan Teknologi. 2020. *Ukuran Pemusatan Data*.
https://lmsspada.kemdiktisaintek.go.id/pluginfile.php/684739/mod_resource/content/1/Ukuran%20Pemusatan%20Data.pdf
- Pandriadi, dkk. 2023. *Statistika Dasar*. Jawa Barat: penerbit Widina Media Utama.
<https://repository.uir.ac.id/24256/1/1.%20Statistik%20Dasar..pdf>
- Riyanto, Agus. 2013. *Statistik Deskriptif Untuk Kesehatan*. Yogyakarta: Nuha Medika.
- Sabri, Luknis., Sutanto Priyo Hastono. 2008. *Statistik Kesehatan*. Depok: PT Rajagrafindo Persada.
- Syaban, Mumun., Elly Ratnaningrum. 2021. *Statistika Penelitian*. Bandung: Penerbit Informatika.

BAB 6

DISTRIBUSI PROBABILITAS DAN TEORI DASAR

Oleh Muhammad Al Azhary

Manusia secara alami memiliki keterbatasan dalam memprediksi masa depan, baik dalam jangka pendek maupun panjang. Kondisi inilah yang disebut sebagai ketidakpastian. Sejak zaman dahulu, manusia mencoba mengantisipasinya dengan berbagai cara, mulai dari pendekatan mistis/ramalan hingga metode rasional yang mempelajari pola alam. Meski sering dianggap menyulitkan, ketidakpastian sebenarnya adalah elemen yang membuat hidup lebih dinamis dan bisa menjadi sebuah anugerah

Teori probabilitas hadir sebagai alat ilmiah untuk mengukur ketidakpastian agar lebih terarah dan mudah diprediksi. Teori ini digunakan saat model matematika biasa tidak mampu menyelesaikan masalah yang kompleks, terutama dalam dunia pendidikan yang membutuhkan pertimbangan pragmatis. Statistika adalah bentuk nyata dari penerapan teori ini. Namun, untuk benar-benar memahami hasil dari perhitungan statistik, seseorang harus memahami model probabilitas yang digunakan di baliknya (Darwel, dkk. 2022).

6.1 Pendekatan Probabilitas

6.1.1 Klasik

Pendekatan klasik dalam teori peluang didasarkan pada prinsip bahwa setiap hasil dalam sebuah percobaan memiliki kemungkinan yang identik untuk terjadi (*equally likely*). Artinya, tidak ada satu hasil pun yang lebih diunggulkan dibanding hasil lainnya. Berikut adalah beberapa ilustrasi sederhananya:

Lempar Koin: Saat melempar koin, ada dua kemungkinan hasil: angka atau gambar. Keduanya punya peluang yang sama besar. Jadi, probabilitas munculnya satu sisi (misalnya sisi gambar) adalah 1 dari 2 kemungkinan, atau senilai $1/2$.

Lempar Dadu: Sebuah dadu memiliki enam sisi (angka 1 sampai 6) dengan peluang muncul yang sama. Jika kita mencari peluang munculnya angka genap (2, 4, atau 6), maka ada 3 hasil yang diinginkan dari total 6 kemungkinan. Secara garis besar, dalam kondisi ideal, peluang dihitung dengan membagi jumlah hasil yang diinginkan dengan total seluruh kemungkinan yang ada (Darwel, dkk. 2022).

6.1.2 Frekuensi

Pendekatan ini mendefinisikan peluang berdasarkan hasil eksperimen yang dilakukan secara berulang-ulang dalam jumlah yang sangat banyak di bawah kondisi yang identik.

Pola Kecenderungan: Jika kita melempar koin hanya beberapa kali (misalnya 10 kali), hasilnya mungkin tidak seimbang (misal 6 kali gambar dan 4 kali angka). Namun, jika koin dilempar terus-menerus hingga jumlah yang sangat besar (mendekati tak terhingga),

maka proporsi kemunculan gambar dan angka akan semakin mendekati angka 0,5 atau 50%.

Limit Frekuensi: Secara matematis, peluang dalam pendekatan ini dianggap sebagai "limit" dari frekuensi relatif. Artinya, akurasi nilai peluang sangat bergantung pada seberapa banyak percobaan dilakukan; semakin besar maka, semakin akurat nilai peluangnya.

Peluang Empiris: Karena didasarkan pada observasi atau pengalaman nyata yang telah terjadi, pendekatan ini juga sering disebut sebagai peluang *a posteriori* atau peluang empiris. Contohnya adalah memprediksi kemungkinan sepasang kekasih akan menikah dalam kurun waktu satu tahun berdasarkan data yang ada.

Intinya: Berbeda dengan teori klasik yang bersifat teoretis, pendekatan ini menentukan peluang lewat data nyata dari percobaan berulang (Darwel, dkk. 2022).

6.1.3 Subyektif

Berbeda dengan metode perhitungan rumus atau data statistik yang kaku, pendekatan ini melihat peluang sebagai cerminan dari **tingkat keyakinan** atau rasa percaya diri seseorang terhadap terjadinya suatu peristiwa.

Berikut adalah poin-poin utamanya:

- **Berdasarkan Keyakinan Pribadi:** Peluang ditentukan oleh seberapa besar seseorang percaya bahwa suatu hal akan terjadi (misalnya, keyakinan bahwa seorang pemimpin akan tetap menjabat tahun depan).
- **Nilai yang Bervariasi:** Karena bersifat personal, nilai peluang antara satu orang dengan yang lain bisa berbeda. Hal ini dipengaruhi oleh latar

belakang, pengalaman hidup, sikap, serta data (baik angka maupun informasi kualitatif) yang dimiliki individu tersebut.

- **Rentang Nilai:** Meski bersifat subjektif, nilai peluangnya tetap berada di antara **0 (sangat tidak yakin)** hingga **1 (sangat yakin)**.
- **Pentingnya Konsistensi:** Dalam menetapkan nilai peluang subjektif ini, seseorang harus berhati-hati dan konsisten agar kesimpulan yang diambil tetap logis dan tidak menyesatkan.

Dalam dunia statistika, konsep ini dikenal sebagai **derajat keyakinan (*level of confidence*)** yang menjadi dasar penting dalam pengambilan keputusan (Darwel, dkk. 2022).

6.2 Konsep Dasar

Probabilitas didefinisikan sebagai metrik numerik untuk mengukur peluang terjadinya suatu peristiwa acak yang tidak diprediksi sebelumnya. Karena bersifat acak, kejadian yang telah direncanakan tidak masuk dalam kategori ini. Secara matematis, nilai probabilitas berkisar antara 0 (mustahil terjadi) hingga 1 (pasti terjadi), yang dihitung dengan membandingkan jumlah hasil yang diharapkan terhadap total seluruh kemungkinan yang ada. Dalam konteks praktis, teori peluang membantu manusia mengestimasi hasil dari berbagai situasi. Contoh sederhananya adalah lemparan koin yang memiliki peluang 50% untuk tiap sisinya, atau prakiraan cuaca yang disusun berdasarkan observasi data atmosfer dan rekaman historis

Dalam bidang matematika, konsep ini menjadi fondasi untuk memodelkan fenomena tidak teratur dan memprediksi berbagai kemungkinan keluaran. Sementara dalam statistika, probabilitas berfungsi sebagai alat ukur ketidakpastian serta

validasi sejauh mana sebuah model mampu menjelaskan data yang ada. Pemanfaatan probabilitas sangat luas, mencakup disiplin ilmu sosial, sains murni, hingga teknik. Di sektor ekonomi dan bisnis, metode ini digunakan untuk menganalisis risiko investasi dan proyeksi laba-rugi. Sedangkan dalam dunia medis, probabilitas krusial untuk mengukur tingkat keberhasilan suatu terapi atau tindakan pengobatan terhadap pasien (Tantik, 2023).

Probabilitas suatu peristiwa ditinjau dari hubungan antar peristiwa yang mungkin terjadi diuraikan dalam tabel berikut

Tabel 6. 1 Hubungan Antar Peristiwa

Jenis Hubungan	Rumus Probabilitas Peristiwa Acak	Penjelasan Peristiwa
Bebas/ Independen	$P(A \cap B) = P(A) \times P(B)$ $\cap = \text{Irisan (Intersection)}$ $P(A \cup B) = P(A) \times P(B) - P(A \cap B)$ $\cup = \text{Gabungan (Union)}$	Probabilitas Peristiwa A dan Peristiwa B terjadi Bersama- sama Probabilitas peristiwa A atau peristiwa B akan terjadi
Saling meniadakan/ (<i>mutually exclusive</i>)	$P(A \cap B) = 0$ $P(A \cup B) = P(A) + P(B)$	Probabilitas peristiwa A dan peristiwa B tidak mungkin terjadi Bersama-sama. Probabilitas peristiwa A atau peristiwa B saja

Jenis Hubungan	Rumus Probabilitas Peristiwa Acak	Penjelasan Peristiwa
		yang terjadi karena saling meniadakan
Bersyarat/ (<i>conditional event</i>)	$P(A \cap B) = P(A) \times P(B/A)$	Probabilitas peristiwa A dan peristiwa B terjadi Bersama jika peristiwa A lebih dulu terjadi.

Sumber: Tantik

Ketika kita melempar dua buah uang logam secara bersamaan, setiap logam memiliki dua sisi: Angka (A) dan Gambar (G). Karena setiap sisi memiliki peluang yang sama untuk muncul, maka probabilitas masing-masing sisi pada setiap koin adalah 0,50 (50%). Karena kedua koin tersebut tidak saling memengaruhi (*independen*), terdapat empat kombinasi hasil yang mungkin muncul:

Tabel 6. 2 Kombinasi 2 Koin

Kombinasi Sisi	Frekuensi	Probabilitas (P)
AA (Angka - Angka)	1x	$1/4 = 25\%$
AG (Angka - Gambar)	1x	$1/4 = 25\%$

Kombinasi Sisi	Frekuensi	Probabilitas (P)
GA (Gambar - Angka)	1x	$1/4 = 25\%$
GG (Gambar - Gambar)	1x	$1/4 = 25\%$
Total	4x	100%

Sumber: Tantik

Setiap kombinasi memiliki peluang yang sama besar, yaitu 25%, sehingga total probabilitas dari seluruh ruang sampel tersebut adalah 100%. Berapa probabilitas minimal muncul satu sisi angka (A)? Minimal muncul satu sisi angka berarti pada komposisi AG, GA, dan AA maka besar probabilitasnya : $P(1 \leq A) = P(AG) + P(GA) + P(AA) = 25\% + 25\% + 25\% = 75\%$

Contoh probabilitas untuk hubungan antar-peristiwa yang meniadakan (*mutually exclusive*).

Satu uang logam dilempar ke atas sekali, maka probabilitas muncul sisi angka (A) dan sisi gambar (G) Bersama-sama adalah peristiwa pasti tidak terjadi. Karena sisi uang logam hanya muncul angka atau gambar saja. Berarti $P(A \cap G) = 0$, sedangkan probabilitas muncul sisi angka (A) atau sisi gambar (G) sebesar $P(A \cup G) = P(A) + P(G) = 0,50 + 0,50 = 1,00$ atau 100%

Contoh probabilitas untuk hubungan antar-peristiwa yang bersyarat (*conditional event*)

Si Darban kemungkinan diterima di perguruan tinggi untuk jenjang S1 sebesar 40% . Setelah diterima di jenjang S1, kemungkinan lulus jenjang S1 tersebut

adalah 60%, Jadi probabilitas si Darban diterima di perguruan tinggi dan lulus S1 adalah :

- Peristiwa diterima di perguruan tinggi diberi simbol T
- Peristiwa lulus jenjang S1 diberi simbol L
- $P(T \text{ dan } L) = P(T \cap L) = P(T) \times P(L/T) = 40\% \times 60\% = 24\%$
(Tantik, 2023)

6.3 Distribusi Probabilitas dalam Kesehatan

6.3.1 Konsep Dasar Distribusi Normal

Dalam statistika kesehatan, distribusi normal merupakan model probabilitas kontinu yang paling dominan karena relevansinya dengan berbagai variabel biologis dan klinis. Ciri khasnya adalah kurva simetris berbentuk lonceng, di mana nilai rata-rata (*mean*), median, dan modus berada di titik pusat yang sama, menunjukkan keseimbangan data.

Parameter krusial dalam distribusi ini adalah standar deviasi, yang mengukur sebaran data dari nilai tengah. Berdasarkan aturan empiris, sekitar 68%, 95%, dan 99,7% data masing-masing tersebar dalam rentang satu, dua, dan tiga standar deviasi dari *mean*. Prinsip ini sangat membantu praktisi kesehatan dalam menentukan apakah suatu parameter klinis (seperti tekanan darah atau berat badan) masih dianggap normal atau sudah masuk kategori menyimpang untuk keperluan pengambilan keputusan medis (La Ode, dkk. 2026).

6.3.2 Analisis Distribusi Binomial dan Poisson dalam Kesehatan

Distribusi binomial adalah model probabilitas diskrit yang digunakan untuk mengevaluasi peristiwa

dengan dua hasil eksklusif, yaitu "sukses" atau "gagal", dalam serangkaian percobaan yang jumlahnya tetap. Syarat utamanya adalah probabilitas keberhasilan tetap konsisten dan setiap percobaan bersifat independen. Dalam praktiknya, model ini sering digunakan untuk menganalisis efektivitas terapi atau hasil uji diagnostik.

Sementara itu, distribusi Poisson digunakan untuk memodelkan jumlah kejadian dalam interval waktu atau ruang yang spesifik, terutama jika peluang kejadiannya kecil dan terjadi secara konstan. Model ini sangat relevan dalam epidemiologi untuk memantau insiden penyakit langka atau efek samping obat. Perbedaan mendasar keduanya terletak pada basis pengukurannya: binomial berfokus pada jumlah percobaan yang pasti, sedangkan Poisson berfokus pada frekuensi kejadian dalam rentang waktu atau area tertentu (La Ode, dkk. 2026).

6.3.3 Implementasi Distribusi Probabilitas dalam Sektor Kesehatan

Distribusi probabilitas berfungsi sebagai kerangka matematis yang krusial untuk memodelkan dan memprediksi fenomena kesehatan. Melalui pemilihan model yang akurat, praktisi kesehatan dapat mengkalkulasi risiko serta merancang strategi intervensi yang lebih efisien. Ketepatan pemilihan distribusi ini ditentukan oleh profil data dan jenis peristiwa yang dianalisis.

Sebagai ilustrasi, distribusi binomial efektif untuk mengukur efikasi vaksin pada sekelompok pasien, sementara distribusi Poisson lebih tepat digunakan untuk memproyeksikan frekuensi infeksi nosokomial di rumah sakit dalam periode tertentu. Di sisi lain, distribusi normal menjadi fondasi utama dalam analisis parameter klinis pada populasi luas. Bagi mahasiswa kesehatan, pemahaman ini menegaskan bahwa statistik

bukan sekadar teori angka, melainkan instrumen vital dalam pengambilan keputusan klinis dan perencanaan program berbasis bukti (La Ode, dkk. 2026).

6.4 Konsep Dasar Probabilitas dalam Kesehatan

6.4.1 Konsep Probabilitas dan Risiko dalam Ranah Kesehatan

Peluang atau probabilitas merupakan parameter matematis dengan rentang nilai 0 hingga 1 yang mengukur potensi terjadinya suatu fenomena. Dalam studi kesehatan, konsep ini krusial untuk memetakan kemungkinan morbiditas, efikasi terapi, maupun efek samping klinis, serta menjadi fondasi bagi pengambilan keputusan berbasis bukti.

Sementara itu, risiko didefinisikan sebagai probabilitas terjadinya dampak kesehatan yang merugikan pada populasi dan periode waktu tertentu. Berbeda dengan peluang umum, risiko memiliki dimensi klinis yang lebih spesifik karena dipengaruhi oleh determinan seperti usia, gender, dan faktor lingkungan. Penguasaan terhadap kedua aspek ini memungkinkan tenaga medis untuk melakukan stratifikasi risiko, menyusun program preventif yang terukur, dan memberikan edukasi yang tepat kepada pasien (La Ode, dkk. 2026).

6.4.2 Estimasi Probabilitas Kejadian Penyakit dalam Populasi

Kalkulasi probabilitas penyakit bertujuan untuk mengukur prevalensi atau insidensi suatu kondisi medis dalam kelompok masyarakat tertentu. Secara teknis, nilai ini diperoleh dengan membandingkan jumlah kasus

yang ditemukan dengan total populasi yang memiliki risiko terpapar. Dalam epidemiologi, metode ini menjadi instrumen utama untuk memetakan beban kesehatan masyarakat secara kuantitatif.

Penerapan probabilitas ini juga mencakup ranah klinis, khususnya dalam mengevaluasi profil keamanan obat atau efektivitas prosedur terapi. Sebagai contoh, jika terdapat 50 kasus baru dari 1.000 penduduk, maka risiko kejadiannya adalah 5%. Data statistik ini menjadi landasan krusial bagi otoritas kesehatan dalam merumuskan kebijakan preventif, pengendalian penyakit, serta evaluasi intervensi medis berbasis data (La Ode, dkk. 2026).

6.4.3 Signifikansi Interpretasi Probabilitas dalam Praktik Klinis

Interpretasi probabilitas dalam ranah klinis tidak terbatas pada angka statistik, melainkan harus diterjemahkan ke dalam tindakan medis yang nyata. Tenaga kesehatan dituntut memiliki kemampuan untuk menilai apakah suatu tingkat risiko memerlukan intervensi segera atau sekadar observasi. Sebagai contoh, temuan probabilitas 20% pada kasus preeklamsia mengharuskan pemantauan yang lebih intensif, sementara risiko efek samping obat yang hanya 1% memungkinkan pemberian terapi dengan pengawasan standar.

Selain sebagai alat diagnostik, pemahaman ini krusial dalam "pengambilan keputusan bersama" (*shared decision-making*) antara tenaga kesehatan dan pasien. Penyampaian risiko yang mudah dipahami memungkinkan pasien terlibat aktif dalam menentukan langkah pengobatan. Oleh karena itu, kompetensi dalam menginterpretasikan probabilitas menjadi standar wajib

dalam praktik kedokteran, keperawatan, dan kebidanan yang berbasis bukti (La Ode, dkk. 2026).

6.5 Permutasi dan Kombinasi

6.5.1 Menyusun Objek dalam Satu Garis (Permutasi)

Teks tersebut menjelaskan cara menghitung banyaknya kemungkinan urutan saat kita menyusun benda dalam sebuah barisan. Berikut adalah poin-poin utamanya:

- **Logika Dasar:** Jika kita memiliki dua objek (A dan B), hanya ada dua urutan yang mungkin: AB dan BA.
- **Pengembangan Objek:** Saat jumlah objek bertambah menjadi tiga (A, B, C), kita bisa memilih salah satu dari tiga objek tersebut untuk mengisi posisi pertama. Setelah satu objek diletakkan, tersisa dua objek yang bisa disusun dalam dua cara. Secara total, ada $(3 \times 2 \times 1 = 6)$ kemungkinan urutan.
- **Pola Umum (Induksi):** Pola ini terus berlanjut. Jika ada empat objek, maka total susunannya adalah $(4 \times 3 \times 2 \times 1 = 24)$ (hasil dari tiga objek sebelumnya), yang menghasilkan 24 cara.
- **Notasi Faktorial:** Untuk memudahkan penulisan perkalian menurun yang panjang ini $(4 \times 3 \times 2 \times 1)$, dunia matematika memperkenalkan simbol faktorial (!).

Kesimpulan:

Banyaknya cara menyusun (n) objek berbeda adalah hasil perkalian berurutan dari angka tersebut turun hingga angka satu (Agus, 2022)

6.5.2 Konsep Kombinasi

Teks tersebut menjelaskan perbedaan mendasar antara menyusun objek dan sekadar memilihnya:

- **Urutan Tidak Penting:** Berbeda dengan permutasi yang memperhatikan urutan, kombinasi digunakan ketika urutan pemilihan tidak menjadi masalah. Fokus utamanya adalah menentukan berapa banyak cara memilih sejumlah objek (k) dari total objek yang tersedia (n).
- **Rumus Kombinasi:** Untuk mencari jumlah cara pemilihan, kita menggunakan rumus:

$${}^nC_k = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

- **Logika di Balik Rumus:** Kombinasi diturunkan dari rumus permutasi yang dibagi kembali dengan (k) . Pembagian ini dilakukan untuk menghilangkan urutan-urutan yang dianggap sama dalam kelompok yang terpilih.
- **Contoh Penerapan:** Jika kita ingin memilih **5 orang panitia** dari total **8 relawan**, kita tidak memedulikan siapa yang terpilih duluan. Maka, penghitungannya adalah: (Agus, 2022)

$$\binom{8}{5} = \frac{8!}{5!3!} = 56$$

DAFTAR PUSTAKA

- Alifariki, L. O. *et al.*, 2026. *Buku Ajar Dasar Dasar Statistik Kesehatan*. Cilacap: PT Media Pustaka Indo.
- Darwel *et al.*, 2022. *Statistik Kesehatan: Teori dan Aplikasi*. Padang: PT Global Eksekutif Teknologi.
- Sumarlin, T., [Tahun Terbit]. *Statistik Probabilitas*. Semarang: Yayasan Prima Agus Teknik
- Wibowo, A., [2022]. *Statistik Probabilitas*. Semarang: Yayasan Prima Agus Teknik & Universitas STEKOM

BAB 7

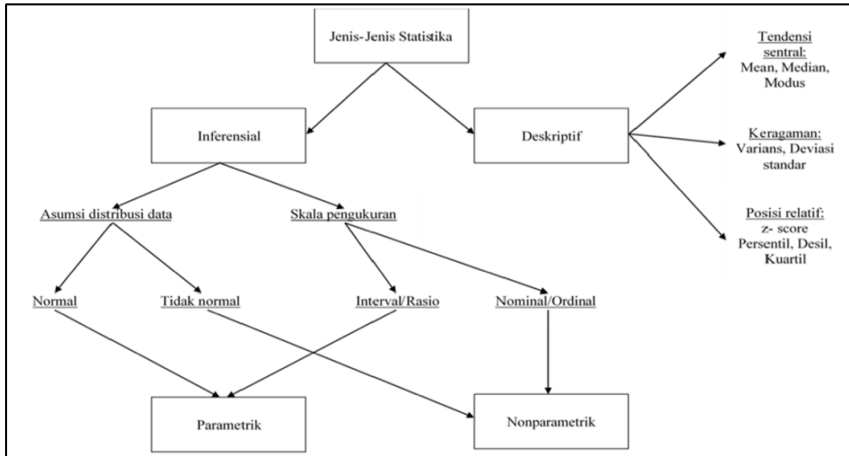
UJI HIPOTESIS PARAMETRIK DAN NONPARAMETRIK

Oleh Nurpatwa Wilda Ningsi

Pengertian statistik (*statistic*) adalah bilangan yang diperoleh melalui proses perhitungan terhadap sekumpulan data yang berasal dari sampel. Penggunaan statistik parametrik dan non parametrik tergantung pada asumsi dan jenis data yang akan dianalisis. Statistik Non Parametrik tidak menuntut terpenuhi banyak asumsi, misalnya data yang dianalisis tidak harus berdistribusi Normal. Oleh karena itu Statistik Parametrik sering disebut sebagai Distribusi Bebas (*Free Distribution*). Statistika nonparametrik biasanya digunakan untuk melakukan analisis pd data kategorial (Nominal/Ordinal). Data berjenis nominal dan ordinal tidak menyebar normal.

Penggunaan metode statistik inferensia dalam penelitian memiliki tiga tujuan utama yaitu melakukan pengujian komparasi/ perbandingan, melihat hubungan/ korelasi/ asosiasi, dan melakukan pembentukan model. Dalam melakukan pengujian komparasi, hal yang dapat dibandingkan ukuran data seperti proporsi data, varian data dan nilai rata-rata data. Baik untuk ketiga ukuran data dapat diuji pada satu populasi dengan membandingkan terhadap suatu nilai acuan. Selain itu dapat dilakukan pada dua populasi yang bersifat dependen/ berpasangan yang diberikan suatu *treatment*/kebijakan maupun pada dua

populasi yang bersifat independen di mana antar nilai/ peluang kejadian populasi tidak saling mempengaruhi. Adapun pemetaan perbedaan antara statistika deskriptif dan statistika inferensial sebagai berikut;



Gambar 7. 1nKategori, Asumsi, dan Jenis Data Statistika

Sumber: Mashuri, 2023

Pengujian hipotesis adalah metode statistika yang digunakan untuk mengevaluasi asumsi tentang populasi berdasarkan data sampel, dengan tujuan menentukan apakah ada cukup bukti untuk menolak hipotesis awal atau menerima alternatifnya. Dalam proses ini, hipotesis nol menyatakan bahwa tidak ada efek atau perbedaan yang signifikan dalam populasi, sedangkan hipotesis alternatif menyatakan adanya efek atau perbedaan yang perlu diuji. Pengujian dilakukan dengan menghitung statistik uji dari data sampel dan membandingkannya dengan distribusi probabilitas yang sesuai, sering kali menggunakan tingkat signifikansi untuk menentukan batas penerimaan atau penolakan hipotesis nol. Jika statistik uji menghasilkan *p-value* kecil yang berada di bawah α , maka hipotesis nol

ditolak, menunjukkan bahwa ada bukti statistik yang cukup untuk mendukung hipotesis alternatif (Mashuri, 2023)

7.1 Statistik Parametrik

Statistik parametrik merupakan bagian dari statistika inferensial yang parameter dari populasinya mengikuti suatu distribusi tertentu, seperti distribusi normal, dan memiliki varians yang homogen. Pengujian hipotesis yang dilakukan melalui statistika parametrik harus memenuhi beberapa asumsi seperti, observasi yang bersifat independen (pemilihan dari sebuah kasus tidak akan mempengaruhi kesempatan kasus-kasus lain untuk

dipilih didalam sampel) (Dodiet Aditya Setyawan, n.d.). Distribusi datanya normal (kurva memiliki kemiringan yang cenderung seimbang, baik pada sisi kiri maupun sisi kanan dan memiliki nilai *skewness* yang rendah atau kemencengan data yang besarnya mendekati nol), pengukuran data berdasarkan nilai, populasinya mempunyai nilai varians yang sama, jumlah sampel yang diambil besar (lebih dari sama dengan 30 atau 80 buah sampel), datanya kuantitatif (berupa angka dan dapat dilakukan operasi matematis), serta memiliki skala pengukuran berupa interval (data mengandung perbedaan, susunan, dan jarak) dan rasio (data mengandung perbedaan, susunan, jarak, serta angka nol mutlak)(Syamsul *et al.*, 2022).

Selain itu, statistika parametrik hanya boleh digunakan jika data memiliki nilai dalam bentuk numerik atau angka nyata. Kelebihan statistika parametrik adalah kesimpulan yang diperoleh lebih kuat jika dibandingkan dengan kesimpulan yang diperoleh dari analisis statistika non parametrik; sedangkan kelemahan statistika parametrik adalah kebutuhan beberapa asumsi untuk melakukan analisis statistika parametrik, dan juga dibutuhkan analisis yang lebih rumit. Terdapat 3 bentuk macam pengujian hipotesis dalam

statistika prametrik diantaranya pengujian satu sampel (*one sample t-test*), pengujian dua sampel bebas (*Independent sample t-test*), dan pengujian dua sampel berpasangan (*paired sample t-test*) (Sulistiyowati & Astuti, 2017).

7.1.1 Pengujian Satu Sampel (*One Sample T-Test*)

One sample t-test berlaku pada satu sampel dan bisa dipakai untuk menguji apakah data sebuah sampel yang diambil menunjang hipotesis yang menyatakan bahwa populasi asal sampel tersebut mengikuti suatu distribusi yang telah ditetapkan. Pengujian beda rata-rata yang menggunakan satu sampel mempunyai tujuan untuk mengetahui apakah sampel tersebut berasal dari suatu populasi yang tertentu. Pada prinsipnya, pengujian satu sampel ingin menguji apakah suatu nilai tertentu (sebagai pembanding) berbeda secara nyata ataukah tidak dengan rata-rata sebuah sampel. Pada umumnya, nilai tertentu di sini adalah sebuah nilai parameter untuk mengukur suatu populasi.

Uji One Sample T-Test adalah metode statistik parametrik yang digunakan untuk mengetahui apakah rata-rata suatu sampel berbeda secara signifikan dengan nilai rata-rata tertentu (nilai pembanding/populasi). Adapun tujuan dilakukan pengujian ini diantaranya;

- a. Mengetahui perbedaan rata-rata sampel dengan nilai standar
- b. Menguji apakah rata-rata sampel telah memenuhi target tertentu
- c. Membantu peneliti dalam pengambilan keputusan berdasarkan data statistik.

Syarat untuk menggunakan *One Sample t-test*, diantaranya;

- a. Data berskala interval atau rasio.

- b. Data diperoleh secara independen.
- c. Sampel hanya satu kelompok.
- d. Data berdistribusi normal.

Variabel dependen dalam one-sample t-test harus diukur dalam skala interval atau rasio dan terdistribusi secara normal (Howell,2017;Ploeger-Lyons, 2017; dalam Mashuri, 2023). Rumus *one sample t-test* adalah sebagai berikut (Privitera, 2015):

$$t = \frac{M - \mu}{\frac{SD}{\sqrt{N}}}$$

Keterangan rumus:

t = Nilai t-test yang dihitung.

M =Rata-rata sampel.

μ = Rata-rata populasi.

SD = Deviasi standar sampel.

N = Jumlah subjek penelitian dalam sampel.

Effect size dalam one-sample t-test bisa mengacu pada rumus di bawah ini (Inayah *et al.*, 2025).

$$t = \frac{M - \mu}{SD}$$

Keterangan rumus:

t = Effect size one-sample t-test.

M = Rata-rata sampel.

μ = Rata-rata populasi.

SD = Deviasi standar sampel.

Tabel 7. 1 Kategorisasi Level Effect Size (d)

Nilai d	Kategori <i>effect size</i>
0.00-0.20	Sangat kecil
0.21-0.79	Sedang
>.80	Sangat besar

7.1.2 Pengujian Dua Sampel Bebas (*Independent Sample t-Test*)

Independent samples t-test, yang disebut juga sebagai *between-subjects t-test*, adalah uji perbedaan univariat untuk menguji perbedaan sebuah variabel dependen (X) atas dasar satu variabel independen (Y). Dalam *independent samples t-test*, terdapat dua kelompok, level perlakuan, atau kategori dalam variabel independen. Dalam statistika parametrik ini, variabel dependen yang diuji berskala interval atau rasio, sementara variabel independen berskala nominal.

Dua kelompok, level, atau kategori variabel independen dalam *independent samples t-test* bersifat “antar-kelompok” (*between-subjects*). Hal ini bersumber pada variasi kelompok, level perlakuan, atau kategori variabel independen yang diberikan kepada atau diisi oleh kelompok subjek yang berbeda. Variasi dalam variabel independen bisa berupa level perlakuan eksperimen (misal: perlakuan A versus perlakuan B), atau bisa berupa kelompok atau kategori yang secara alamiah berbeda (misal: pria versus wanita) atau yang ada di luar kendali peneliti (misal: sekolah akselerasi versus non-akselerasi, dan lain lain) *Independent sample t-test* memiliki beberapa asumsi.

Beberapa asumsi tersebut menyatakan bahwa, pertama, skor variabel dependen dalam dua kelompok sampel yang dibandingkan harus terdistribusi secara normal. Asumsi lainnya adalah kesetaraan varians (*homogeneity of variance*) variabel dependen dalam dua kelompok sampel yang dibandingkan (Graveter dkk., 2020; King dkk., 2021). Rumus independent samples t-test untuk dua sampel berukuran sama (*equal sample sizes*; King dkk., 2021) adalah sebagai berikut:

$$t = \frac{(M_1 - M_2)}{\frac{\sqrt{SD_2X_1 + (N_1 - 1) + SD_2X_2 + (N_2 - 1)}}{N_1 + N_2 - 2}}$$

Keterangan rumus:

t = Koefisien independent sample t-test yang dihitung.

M_1 = Rata-rata sampel/kelompok/perlakuan 1.

M_2 = Rata-rata sampel/kelompok/perlakuan 2.

SD_2X_1 = Varians sampel/kelompok/perlakuan 1.

SD_2X_2 = Varians sampel/kelompok/perlakuan 2.

N_1 = Jumlah subjek pada sampel/kelompok/perlakuan 1.

N_2 = Jumlah subjek pada sampel/kelompok/perlakuan 2.

Sementara itu, persentase peran atau pengaruh variabel independent terhadap variabel dependen dihitung dengan rumus 30 di bawah ini. Tabel 23 menampilkan kategorisasi persentase peran atau pengaruh dalam independent samples t-test (Sulistiyowati & Astuti, 2017):

$$r^2 = \frac{t^2}{t^2 + db}$$

Tabel 7. 2 Kategorisasi Persentase Pengaruh dalam Uji Perbedaan *Independent Samples t-Test*

Nilai r^2	Kategori pengaruh
$r^2 = .01$	Sangat kecil
$r^2 = .09$	Sedang
$r^2 = .25$	Sangat besar

7.1.3 Pengujian Dua Sampel Berkaitan (*Paired Sample T-Test*)

Paired samples t-test merupakan tes univariat untuk menguji perbedaan sebuah variabel dependen atas dasar satu variabel independen. Dalam *paired samples t-test*, variabel independen memiliki atau terdiri dari dua kelompok atau kategori. Sebagai salah satu jenis statistika parametrik, variabel dependen yang diuji dalam *paired samples t-test* bersifat kontinum (interval atau rasio) sementara variabel independen bersifat nominal *paired samples t-test* disebut juga sebagai *repeated measures t-test* dalam artian bahwa variasi dalam variabel independen (jumlah kelompok atau kategori) diisi oleh atau diberikan kepada kelompok subjek yang sama.

Variasi dalam variabel independen tersebut bisa diaplikasikan dalam suatu eksperimen. Dengan desain eksperimen ini, variasi level atau kategori variabel independen diberlakukan dalam bentuk, misalnya, pre-test dan post-test kepada subjek penelitian yang sama secara berulang (*repeated*) *Paired samples t-test* mengasumsikan bahwa variabel dependen terdistribusi

secara normal dan tidak memiliki outliers (Mashuri, 2023).

Rumus *Paired samples t-test* adalah sebagai berikut

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

$$SD_D = \sqrt{\sum (D - M_D)^2 / (n - 1)}$$

Keterangan rumus:

t = Koefisien/nilai paired samples t-test.

M^1 = Rata-rata skor subjek pada pengukuran pertama.

M^2 = Rata-rata skor subjek pada pengukuran kedua.

D = Hasil pengurangan skor pengukuran pertama dan skor pengukuran kedua untuk masing-masing subjek penelitian.

SD_D = Standar Deviasi perbedaan skor pengukuran pertama dan pengukuran kedua.

M_D = Rata-rata perbedaan antara skor subjek pada perlakuan pertama dan kedua.

n = Jumlah pasangan skor.

Pengujian hipotesis dapat bersifat parametrik, jika bergantung pada asumsi distribusi tertentu seperti distribusi normal, atau non-parametrik, jika tidak memerlukan asumsi distribusi spesifik. Metode ini banyak digunakan dalam berbagai bidang penelitian,

termasuk ekonomi, kesehatan, dan ilmu sosial, untuk memastikan bahwa kesimpulan yang diambil didasarkan pada analisis data yang valid dan dapat diandalkan. Hipotesis nol dan hipotesis alternatif adalah dua konsep fundamental dalam pengujian hipotesis statistik yang digunakan untuk menentukan apakah suatu efek atau perbedaan yang diamati dalam data cukup signifikan untuk disimpulkan sebagai sesuatu yang nyata dalam populasi (Syamsul *et al.*, 2022).

Hipotesis nol menyatakan bahwa tidak ada perbedaan atau efek yang signifikan, sering kali mewakili asumsi status quo atau kondisi dasar yang ingin diuji misalnya, bahwa rata-rata dua kelompok adalah sama atau bahwa suatu perlakuan tidak memiliki efek tertentu. Sebaliknya, hipotesis alternatif menyatakan bahwa ada perbedaan atau efek yang signifikan, yang menunjukkan bahwa parameter populasi memiliki nilai berbeda dari yang diasumsikan dalam Hipotesis nol (Faridi *et al.*, 2021).

7.2 Statistik Nonparametrik

Statistik nonparametrik merupakan pendekatan analisis yang tidak mengandalkan asumsi-asumsi kuat seperti kenormalan distribusi atau homogenitas varians. Pendekatan ini sangat berguna ketika peneliti bekerja dengan data yang berskala ordinal atau data numerik yang tidak memenuhi asumsi-asumsi statistik parametrik. Salah satu kekuatan utama dari metode nonparametrik adalah fleksibilitasnya. Teknik ini memungkinkan peneliti untuk tetap melakukan analisis meskipun data yang dimiliki tidak ideal (Jasmin *et al.*, 2023).

Uji nonparametrik merupakan teknik statistik yang digunakan apabila data penelitian tidak memenuhi asumsi parametrik. Uji ini lebih fleksibel karena tidak mengharuskan

data berdistribusi normal. Uji nonparametrik sering digunakan pada data nominal, ordinal, atau penelitian dengan jumlah sampel kecil. Menurut Sidney Siegel, statistik nonparametrik merupakan metode statistik yang tidak bergantung pada asumsi distribusi populasi tertentu. Pengujian hipotesis non-parametrik memiliki berbagai penerapan dalam berbagai bidang karena fleksibilitasnya dalam menganalisis data tanpa asumsi distribusi tertentu. Dalam bidang Kesehatan dan Biomedis; Dalam penelitian medis, uji non-parametrik sering digunakan untuk membandingkan efektivitas

pengobatan ketika data tidak terdistribusi normal atau memiliki sampel kecil. Misalnya, Uji Wilcoxon Rank-Sum digunakan untuk membandingkan dua kelompok pasien berdasarkan tingkat respons terhadap suatu terapi, sementara Uji Friedman dapat digunakan dalam studi klinis untuk menganalisis hasil dari pengobatan yang diterapkan berulang kali pada subjek yang sama (Cong, *et al.*, 2022; Elsinga, *et al.*, 2017). Adapun karakteristik pada uji nonparametrik, diantaranya;

1. Tidak mensyaratkan distribusi normal.
2. Dapat digunakan untuk data nominal dan ordinal.
3. Cocok untuk sampel kecil.
4. Perhitungan lebih sederhana.
5. Lebih fleksibel dibanding uji parametrik.

7.2.1 Uji Wilcoxon Untuk Sampel Berpasangan

Salah satu metode yang sering digunakan untuk menganalisis data berpasangan adalah Uji Wilcoxon Signed- Rank Test. Uji Wilcoxon merupakan alternatif nonparametrik dari uji t berpasangan (*paired t-test*) yang digunakan ketika data tidak memenuhi asumsi normalitas. Metode ini digunakan untuk

membandingkan dua set data berpasangan—misalnya, sebelum dan sesudah perlakuan tertentu—dengan syarat bahwa perbedaan antara pasangan dapat diurutkan dan dinilai secara ordinal.

Contoh penggunaan uji Wilcoxon dapat ditemukan dalam berbagai bidang ilmu, mulai dari psikologi, pendidikan, kesehatan, hingga ekonomi dan bisnis. Misalnya, dalam studi klinis, seorang peneliti mungkin ingin mengetahui apakah tekanan darah pasien berubah secara signifikan setelah diberi obat tertentu. Karena data tekanan darah tidak selalu terdistribusi normal dan sampel bisa kecil, maka uji Wilcoxon menjadi alat yang tepat untuk melakukan pengujian tersebut.

Uji Wilcoxon sering kali diposisikan sebagai alternatif dari uji t berpasangan (*paired t-test*), yang keduanya sama-sama digunakan untuk membandingkan dua kondisi yang diukur pada subjek yang sama. Namun, terdapat sejumlah perbedaan mendasar antara kedua uji ini, baik dari segi asumsi, cara kerja, maupun konteks penggunaannya. Adapun langkah- langkah dalam melakukan uji

Wilcoxon, diantaranya (Ahmadi *et al.*, 2024);

1. Menyusun data berpasangan

Langkah pertama dalam melakukan uji Wilcoxon untuk sampel berpasangan adalah menyusun data secara sistematis.

2. Menghitung selisih dan rangking

Setelah data disusun dalam bentuk pasangan, langkah berikutnya adalah menghitung selisih antara dua nilai dalam setiap pasangan. Selisih ini diperoleh dengan mengurangkan nilai kondisi kedua dari nilai kondisi pertama (atau sebaliknya, tergantung pada arah pengujian).

3. Menentukan nilai statistik wilcoxon

Setelah peringkat positif dan negatif telah ditentukan, langkah berikutnya adalah menghitung nilai statistik Wilcoxon. Nilai ini dikenal sebagai W, yaitu jumlah dari peringkat terkecil antara total rank positif dan total rank negatif. Jika jumlah pasangan data (setelah nilai selisih nol diabaikan) relatif kecil, yaitu di bawah 25, maka perbandingan dilakukan dengan tabel distribusi Wilcoxon. Nilai kritis untuk berbagai jumlah pasangan dan tingkat signifikansi (misalnya $\alpha = 0,05$ atau $0,01$). Dalam hal ini, nilai statistik W akan dikonversi menjadi nilai Z menggunakan rumus:

$$Z = \frac{W - \mu_w}{\sigma W}$$

Dimana:

$\mu_w = \frac{n(n+1)}{4}$ adalah nilai ekspektasi

$\sigma W = \sqrt{\frac{n(n+1)(2n+1)}{24}}$ adalah simpangan baku

n adalah jumlah pasangan data (tanpa nol)

4. Interpretasi hasil

Langkah terakhir dan sangat penting dari uji Wilcoxon adalah interpretasi hasil. Proses ini melibatkan analisis terhadap nilai statistik yang diperoleh dan bagaimana nilai tersebut digunakan untuk membuat kesimpulan terhadap hipotesis penelitian. Jika nilai W yang diperoleh lebih kecil atau sama dengan nilai kritis dari tabel Wilcoxon (untuk n kecil), atau jika nilai Z menunjukkan signifikansi (untuk n besar), maka Keputusan yang diambil adalah menolak hipotesis nol (H_0).

7.2.2 Uji Mann Whitney

Uji Mann-Whitney merupakan salah satu statistic nonparametrik yang banyak digunakan karena fleksibilitas dalam penggunaan, yaitu dapat digunakan untuk data ordinal atau kontinu tanpa memerlukan asumsi normalitas, sederhana dalam proses penghitungan dan luas aplikasinya, cocok digunakan pada berbagai bidang penelitian, termasuk ilmu sosial, pendidikan, biomedis, dan teknik. Namun, uji ini kurang kuat dibandingkan dengan uji parametrik jika asumsi normalitas terpenuhi (Syamsul *et al.*, 2022).

Uji Mann-Whitney tidak menguji perbedaan rata-rata (mean) seperti pada uji parametrik, melainkan membandingkan peringkat (ranking) dari nilai-nilai dalam kedua kelompok. Dengan kata lain, uji ini bertujuan untuk menentukan apakah dua kelompok berasal dari populasi dengan distribusi yang sama atau berbeda. Hipotesis yang diajukan pada Uji Mann-Whitney sedikit berbeda dari uji parametrik seperti uji-t, karena fokusnya adalah pada peringkat nilai, bukan pada rata-rata.

Meskipun Uji Mann-Whitney tidak mensyaratkan asumsi distribusi normal seperti halnya uji parametrik, tetap ada beberapa asumsi dasar yang perlu dipenuhi agar hasil uji tersebut valid dan dapat diinterpretasikan dengan benar. Adapun Langkah- Langkah dalam melakukan Uji Mann Whitney secara manual, diantaranya (Nuryadi *et al.*, 2017);

1. Pisahkan data menjadi dua kelompok independent
2. Gabungkan data dari kedua kelompok
3. Urutkan data secara keseluruhan
4. Tetapkan ranking pada setiap data

5. Lakukan perhitungan statistik

Uji Statistik untuk masing- masing kelompok dapat dihitung menggunakan rumus berikut:

$$U_1 = R_1 - \frac{n_1(n_1 + 1)}{2}$$

$$U_2 = R_2 - \frac{n_2(n_2 + 1)}{2}$$

Dengan:

R_1 : Jumlah ranking total untuk kelompok 1

R_2 : Jumlah ranking total untuk kelompok 2

n_1 : Ukuran sampel kelompok 1

n_2 : Ukuran sampel kelompok 2

6. Tentukan signifikansi statistik

- Untuk Sampel Kecil ($n_1, n_2 \leq 20$): Gunakan tabel distribusi kritis Mann-Whitney untuk menemukan nilai kritis berdasarkan n_1, n_2 dan tingkat signifikansi (α). Bandingkan nilai U yang dihitung dengan nilai kritis dari tabel, jika $U \text{ hitung} \leq U \text{ kritis}$, tolak hipotesis nol (H_0).

- Untuk Sampel Besar ($n_1, n_2 > 20$):

Aproksimasi distribusi normal digunakan karena distribusi U mendekati normal untuk sampel besar. Hitung nilai z-score menggunakan rumus berikut:

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

7. Interpretasi Hasil

Misalkan Hipotesis Nol (H_0) : Median kedua kelompok adalah sama. Hipotesis Alternatif (H_a): Media kedua kelompok berbeda (dapat bersifat satu

arah atau dua arah, tergantung konteks penelitian). Jika hasil uji menunjukkan bahwa nilai U signifikan (tolak H_0), maka disimpulkan bahwa ada perbedaan signifikan antara kedua kelompok.

7.2.3 Uji Kruskal Wallis

Uji Kruskal-Wallis adalah salah satu uji statistik non-parametrik yang digunakan untuk menguji perbedaan antara dua atau lebih grup independen ketika data tidak memenuhi asumsi distribusi normal. Uji ini merupakan alternatif dari uji ANOVA satu arah, yang mengasumsikan bahwa data berasal dari distribusi normal dan memiliki varians yang homogen antar grup. Salah satu alasan utama penggunaan uji Kruskal-Wallis adalah ketika data yang digunakan bersifat ordinal atau tidak memenuhi asumsi distribusi normal.

Data ordinal adalah data yang dapat diurutkan atau diberi peringkat, namun jarak antar peringkat tidak diketahui atau tidak konsisten. Salah satu asumsi utama dalam uji Kruskal-Wallis adalah bahwa data yang dianalisis harus berasal dari kelompok yang independen satu sama lain. Hipotesis dalam uji Kruskal-Wallis memiliki dua komponen utama: hipotesis nol (H_0) dan hipotesis alternatif (H_1). Hipotesis nol dalam uji Kruskal-Wallis menyatakan bahwa tidak ada perbedaan yang signifikan antara grup yang diuji. Artinya, semua grup yang dibandingkan berasal dari populasi yang memiliki distribusi yang sama atau tidak ada perbedaan dalam distribusi peringkat antar grup (Syapitri *et al.*, 2021).

Contoh penerapan langkah-langkah perhitungan uji Kruskal-Wallis dapat diberikan dalam konteks penelitian medis. Misalnya, peneliti ingin menguji apakah ada perbedaan dalam tingkat penyembuhan antara tiga grup pasien yang menerima pengobatan yang berbeda.

- Langkah pertama adalah mengumpulkan data, yaitu tingkat penyembuhan untuk masing-masing pasien dalam tiga grup. Setelah data terkumpul, peneliti memberikan peringkat pada nilai tingkat penyembuhan dari seluruh pasien dalam ketiga grup.
- Selanjutnya, peneliti menghitung jumlah peringkat untuk setiap grup dan menggunakan rumus untuk menghitung statistik H. Setelah itu, peneliti menghitung nilai p dengan membandingkan statistik H dengan distribusi Chi-square dan menentukan apakah perbedaan antara grup tersebut signifikan atau tidak.

7.2.4 Uji Chi-Square

Chi -Square disebut juga dengan Kai Kuadrat. Chi Square adalah salah satu jenis uji komparatif non parametrik yang dilakukan pada dua variabel, di mana skala data kedua variabel adalah nominal. Uji Chi Square digunakan untuk mengetahui apakah terdapat perbedaan yang signifikan antara frekuensi yang diobservasi dengan frekuensi Uji Chi -Square sangat efektif digunakan untuk menganalisis hubungan antar dua variabel kategori dalam tabel kontingensi. Karena tidak bergantung diharapkan dalam suatu distribusi data kategori pada asumsi distribusi normal, uji Chi -Square banyak digunakan dalam penelitian sosial, pendidikan, dan kesehatan yang menggunakan data kualitatif. Oleh karena itu, uji ini menjadi bagian penting dalam metode analisis non- parametrik. Adapun Langkah-langkah uji Kesuaian Chi Square data dari sampel tunggal dan k-Sampel Independen: Berikut Langkah-langkah uji chi square kelompok tunggal:

- a. Menentukan hipotesis nol dan hipotesis alternatif.

Hipotesis nihil atau nol merupakan hipotesis atau dugaan yang biasanya bertentangan dengan pernyataan yang ingin dibuktikan. Sedangkan Hipotesis alternatif merupakan hipotesis yang ingin dibuktikan (Indrawati, 2015).

- b. Perhitungan hasil Chi Square dengan rumus χ^2 dengan menggunakan rumus uji Chi Square adalah

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

Dengan:

O_i : frekuensi observasi pada kategori i

E_i : frekuensi harapan pada kategori j

Σ : Penjumlahan semua kategori k

- c. Menentukan nilai Chi -Square tabel dengan derajat kebebasan (df)=n-1. Huruf n menyatakan jumlah kategori.
- d. Kriteria Pengujian;
1. Jika $\chi^2_{\text{hitung}} \leq \chi^2_{\text{tabel}}$, maka H_0 diterima
 2. Jika $\chi^2_{\text{hitung}} > \chi^2_{\text{tabel}}$, maka H_0 ditolak atau
 3. Jika Sig. $\chi^2_{\text{hitung}} > \alpha$, maka H_0 diterima
 4. Jika Sig. $\chi^2_{\text{hitung}} < \alpha$, maka H_0 ditolak
- e. Menarik Kesimpulan apakah ada hubungan atau ada perbedaan yang signifikan atau tidak sesuai kriteria pengujian. Untuk H_0 ditolak. Berarti ada perbedaan signifikan antara frekuensi yang diamati dan yang diharapkan.

Berikut Langkah-langkah uji chi square untuk 2 kelompok data:

- 1) Merumuskan hipotesis
- 2) Membuat Tabel Kontingensi

Susun data dalam bentuk tabel dua arah yang menunjukkan distribusi frekuensi observasi dari k sampel terhadap kategori yang diuji.

- 3) Menghitung Frekuensi Harapan (*Expected Frequency*)

$$E_{ij} = \frac{n_i \times n_j}{n}$$

Dengan :

n_i : total baris ke-i

n_j : total baris ke-j

n : total seluruh observasi

- 4) Menghitung Chi Square dengan rumus χ^2 dengan rumus

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^k \frac{(O_i - E_i)^2}{E_i}$$

dengan :

O_{ij} : frekuensi observasi dalam baris ke i pada kolom ke j

E_{ij} : frekuensi harapan dalam baris ke i pada kolom ke j

Σ : jumlah semua sel

- 5) Menentukan derajat Kebebasan (df)

$$df = (r - 1)(c - 1)$$

dengan :

r : jumlah baris

c : jumlah kolom

- 6) Menentukan nilai kritis untuk mencari chi square tabel
- 7) Kriteria Keputusan

DAFTAR PUSTAKA

- Ahmadi, R., Murah, & Sulaiman. (2024). *Statistik Nonparametrik* (Mohzana, Ed.). Lembaga Yasin AlSys.
- Dodiet Aditya Setyawan, I. (n.d.). *Statistik Nonparametrik*. Terapi Wicara Poltekkes Kemenkes Surakarta.
- Faridi, A., Ramdany, R., & Lusiana, S. A. (2021). *Metodologi Penelitian Kesehatan*. Yayasan Kita Menulis.
- Inayah, S., Safitri, Y., Sarmada, Z., Andriani, N., Maitri, M., Bunga, A., Cut, M., Gusmanely, M., Muhammad, D., & Karmanto, B. (2025). *STATISTIKA NON PARAMETRIK* (Alfauzain, Ed.). U ME Publishing.
- Jasmin, M., Risnawati, Mk., Rahma Sari Siregar, Mk., St Mutiatu Rahmah, Ms., Wahidah Rohmawati, Mk., Lilis Handayani, Mk., Ronald, Mk., apt Bai Athur Ridwan, Mk., apt Made Ary Sarasmita, Mp., Ns Hana Febriyanti, Mf., Dyah Juliastuti, Mk., Mat, S., Fika Tri Anggraini, dr, Yusnita Anggraeni, Mk. M., & Siska Oktari, Mb. (2023). *METODOLOGI PENELITIAN KESEHATAN*. EUREKA MEDIA AKSARA.
- Mashuri, A. (2023). *Statistika Parametrik Dasar*. h PT Cita Intrans Selaras.
- Nuryadi, Astuti, T. D., Utami, E. S., & Budiantara. (2017). *DASAR-DASAR STATISTIK PENELITIAN*. Sibuku Media. www.sibuku.com
- Sulistiyowati, W., & Astuti, C. C. (2017). *STATISTIK DASAR KONSEP DAN APLIKASINYA*. UMSIDA PRESS.

Syamsul, M., Darwel, Ramlan, P., Samad, M. A., Syakurah, R. A., Ngkolu, N. W., Lestari, P. P., & Rahmawati. (2022). *STATISTIK KESEHATAN TEORI DAN APLIKASI*. PT GLOBAL EKSEKUTIF TEKNOLOGI. www.globaleksekutifteknologi.co.id

Syapitri, H., Amila, & Aritonang, J. (2021). *Metode Penelitian Kesehatan*. AHLIMEDIA PRESS.

www.ahlimediapress.com

BAB 8

ANALISIS KORELASI DAN REGRESI

Oleh Nurul Hidayah

Analisis hubungan antarvariabel dalam penelitian kesehatan digunakan untuk menilai keterkaitan antara variabel independen dan variabel dependen pada suatu populasi. Keterkaitan tersebut dapat berbeda menurut arah, kekuatan, dan pola hubungan, baik positif maupun negatif, kuat maupun lemah, serta linear maupun non-linear, bergantung pada karakteristik data yang dianalisis. Dalam praktiknya, analisis ini sering digunakan untuk mengkaji hubungan faktor demografis, perilaku, lingkungan, atau biologis dengan status kesehatan individu maupun kelompok, misalnya hubungan antara kebiasaan merokok dan penyakit paru atau antara kadar glukosa darah dan risiko komplikasi diabetes melitus. Oleh karena itu, analisis hubungan antarvariabel menjadi dasar penting dalam pengembangan strategi pencegahan, diagnosis, terapi, evaluasi pelayanan kesehatan, serta penguatan praktik kesehatan berbasis bukti (Solikhah; and Amyati., 2022).

Dalam penelitian kesehatan, peneliti sering berhadapan dengan pertanyaan mengenai ada atau tidaknya hubungan antara dua variabel, seberapa kuat hubungan tersebut, serta apakah suatu variabel dapat digunakan untuk memperkirakan nilai variabel lainnya. Pertanyaan tersebut dapat dijawab melalui analisis korelasi dan analisis regresi. Analisis korelasi digunakan untuk mengukur kekuatan dan arah hubungan antara dua atau lebih variabel numerik tanpa

menetapkan hubungan sebab-akibat. Sementara itu, analisis regresi digunakan untuk membangun model matematis yang dapat menjelaskan atau memprediksi nilai variabel dependen berdasarkan satu atau lebih variabel independen (Fitriangga, 2024).

8.1 Korelasi dan Regresi

Korelasi sangat berkaitan dengan regresi, oleh sebab itu regresi dilakukan jika sudah terbukti ada hubungan terlebih dahulu antara variabel independen dan dependen. Jika tidak terbukti tidak ada hubungan maka tidak bisa dilakukan uji regresi. Antar variabel independen harus benar-benar independen (tidak boleh ada korelasi antar variabel independen). Hubungan antara variabel dependen dan independen adalah hubungan yang linier sehingga regresi yang dilakukan adalah regresi linier. Hubungan yang terjadi antara dependen dan independen adalah hubungan secara langsung, tidak boleh ada variabel antara diantara dependen dan independen. Namun dalam beberapa literatur ditemukan bahwa meskipun tidak ada hubungan antara satu variabel independen terhadap dependennya, namun ketika dimasukkan variabel yang lain justru menjadi berhubungan. Oleh sebab itu dalam analisis regresi, independen yang tidak berkorelasi tidak dihilangkan terlebih dahulu, namun dikaji pada kekuatan model yang dibentuk dari hasil analisis regresi (Kartiningrum *et al.*, 2020).

Sebelum membahas lebih lanjut mengenai metode analisis hubungan dalam penelitian kesehatan, penting untuk memahami perbedaan mendasar antara analisis korelasi dan analisis regresi. Kedua metode tersebut sama-sama digunakan untuk menganalisis hubungan antarvariabel, namun memiliki tujuan, interpretasi, dan keluaran analisis yang berbeda. Oleh karena itu, pemahaman mengenai perbedaan kedua metode ini diperlukan agar peneliti dapat

memilih jenis analisis yang sesuai dengan tujuan penelitian dan karakteristik data yang digunakan (Tabel 9.1) (Kartiningrum *et al.*, 2020; Solikhah; and Amyati., 2022).

Tabel 8. 1 Perbedaan Analisis Korelasi dan Regresi

Aspek	Analisis Korelasi	Analisis Regresi
Tujuan dan interpretasi utama	Menilai ada tidaknya hubungan, arah, dan kekuatan hubungan antarvariabel.	Menilai pengaruh variabel independen terhadap dependen serta membentuk model prediksi.
Kedudukan variabel	Variabel diperlakukan setara; tidak ada variabel bebas dan terikat.	Terdapat variabel independen (X) dan variabel dependen (Y).
Hubungan sebab-akibat	Tidak menunjukkan hubungan sebab-akibat.	Dapat menjelaskan hubungan prediktif, tetapi kausalitas tetap bergantung pada desain penelitian.
Koefisien dan arah hubungan	Koefisien berkisar -1 sampai $+1$; tanda positif/negatif menunjukkan arah hubungan.	Koefisien dapat bernilai positif atau negatif dan menunjukkan besar perubahan/pengaruh X terhadap Y.
Output utama	Koefisien korelasi dan nilai p.	Persamaan regresi, koefisien regresi,

Aspek	Analisis Korelasi	Analisis Regresi
		R^2 /adjusted R^2 , OR/HR, dan nilai p.
Jumlah variabel	Umumnya dua variabel, tetapi dapat diperluas.	Satu atau beberapa variabel independen terhadap satu variabel dependen.
Contoh uji	Pearson, Spearman, Kendall's Tau, Phi, Koefisien Kontingensi.	Regresi linier sederhana, regresi linier ganda, regresi logistik, regresi ordinal, regresi Cox.
Contoh penggunaan	Menilai hubungan antara kadar glukosa darah dan tekanan darah.	Memprediksi risiko diabetes berdasarkan indeks massa tubuh, usia, dan kadar glukosa darah.

Eksplorasi Data Awal: Diagram Pencar (*Scatter Plot*)

Sebelum melakukan uji korelasi atau regresi, peneliti perlu melakukan eksplorasi data awal menggunakan diagram pencar (*scatter plot*) untuk melihat pola hubungan antarvariabel secara visual. Diagram ini membantu menilai apakah hubungan bersifat linier, kurvilinear, atau tidak berpola, sekaligus menunjukkan arah, kekuatan hubungan, dan keberadaan nilai pencilan (*outlier*). Pemeriksaan ini penting karena pemilihan metode analisis sangat bergantung pada pola hubungan data. Jika pola hubungan tidak linier atau terdapat outlier ekstrem, maka korelasi Pearson atau regresi linier perlu dipertimbangkan ulang, misalnya dengan transformasi data atau penggunaan metode nonparametrik.

Adapun bagan keputusan pemilihan uji korelasi dan regresi disajikan pada Gambar 8.1.

8.1.1 Korelasi

Pemilihan uji hipotesis korelatif ditentukan oleh karakteristik variabel dan distribusi data yang dianalisis. Hipotesis korelatif umumnya digunakan apabila hubungan yang diteliti melibatkan dua variabel numerik. Pada kondisi tersebut, uji korelasi Pearson digunakan apabila data memiliki distribusi normal dan hubungan antarvariabel bersifat linear. Jika data tetap tidak berdistribusi normal setelah melakukan melakukan transformasi (tidak memenuhi asumsi normalitas), maka analisis alternatif nonparamterik yang digunakan adalah uji korelasi Spearman. Jenis uji korelasi parametrik (Tabel 8.2) dan non paramaterik (Tabel 8.3) dan bagan alur pemilihan korelasi dan regresi dapat dilihat pada Gambar 8.1.

Tabel 8. 2 Jenis Uji Korelasi Parametrik

Uji Korelasi	Jenis Data	Kegunaan & Syarat Utama
Pearson (r)	Numerik–Numerik (interval / rasio)	Mengukur kekuatan & arah hubungan <i>linier</i> . Syarat: kedua variabel berdistribusi normal, tidak ada outlier ekstrem, hubungan linier.
Korelasi Parsial	Numerik–Numerik (interval / rasio)	Mengukur korelasi antara dua variabel setelah mengendalikan variabel ketiga (variabel perancu / confounding). Berbasis Pearson → syarat sama.
Korelasi Ganda (R)	Numerik–Numerik	Mengukur korelasi antara satu variabel dependen

Uji Korelasi	Jenis Data	Kegunaan & Syarat Utama
	(interval / rasio)	dengan beberapa variabel independen sekaligus. Digunakan dalam regresi linier berganda.
Point Biserial (r_{pb})	Dikotomi sejati–Numerik (dikotomi alami)	Kasus khusus Pearson: variabel X dikotomi alami (mis. jenis kelamin), variabel Y kontinu normal. Hasil ekuivalen dengan Pearson pada data yang sama.
Eta (η)	Nominal / Kategorik–Numerik	Mengukur kekuatan hubungan <i>lengkung (non-linier)</i> antara variabel kategorik dan numerik. Tidak mengasumsikan linearitas. Berbasis ANOVA.

* Catatan: Pearson adalah uji korelasi parametrik utama. Semua uji di atas mensyaratkan asumsi distribusi (umumnya normalitas).

Tabel 8. 3 Jenis Uji Korelasi Non Parametrik (Saputra, 2016)

Uji Korelasi	Jenis Data	Kegunaan & Keterangan
Spearman Rank (ρ)	Ordinal–Ordinal atau Numerik tidak normal	Mengukur hubungan monoton (searah/berlawanan secara konsisten). Alternatif non-parametrik dari Pearson. Paling umum digunakan.
Kendall Tau (τ)	Ordinal–Ordinal	Mengukur hubungan monoton. Lebih robust dibanding Spearman untuk sampel kecil

Uji Korelasi	Jenis Data	Kegunaan & Keterangan
		dan data dengan banyak <i>ties</i> (nilai sama). Terdiri dari: τ -a, τ -b, τ -c.
Gamma (γ)	Ordinal–Ordinal	Alternatif Kendall untuk tabel kontingensi ordinal dengan banyak <i>ties</i> . Tidak mempertimbangkan pasangan yang seri (<i>tied pairs</i>) dalam perhitungan.
Somers' d	Ordinal–Ordinal	Versi asimetris dari Kendall Tau-b. Digunakan jika variabel dependen (Y) jelas ditentukan. Menilai seberapa baik X memprediksi Y.
Phi (ϕ)	Nominal dikotomi–Nominal dikotomi (tabel 2×2)	Mengukur hubungan antara dua variabel dikotomi. Ekuivalen dengan Pearson pada data biner. Nilai: –1 s.d. +1.
Cramer's V	Nominal–Nominal (tabel r×k, r atau k > 2)	Generalisasi Phi untuk tabel lebih dari 2×2. Nilai 0–1 (tidak ada arah). Digunakan saat salah satu atau kedua variabel memiliki >2 kategori.
Koefisien Kontingensi (C)	Nominal–Nominal (semua ukuran tabel)	Mengukur keeratan hubungan pada tabel kontingensi nominal. Berbasis chi-square (χ^2). Kelemahan: nilai maksimumnya <1 tergantung ukuran tabel. Cramer's V lebih dianjurkan.

Uji Korelasi	Jenis Data	Kegunaan & Keterangan
Rank Biserial (r_{rb})	Dikotomi–Ordinal atau Dikotomi–Numerik tidak normal	Mengukur hubungan antara variabel dikotomi (mis. kelompok perlakuan vs kontrol) dengan variabel ordinal/rank. Digunakan bersama uji Mann-Whitney.
Lambda (λ)	Nominal–Nominal	Mengukur kemampuan prediksi asimetris variabel nominal. Menunjukkan seberapa besar error berkurang saat menggunakan X untuk memprediksi Y.

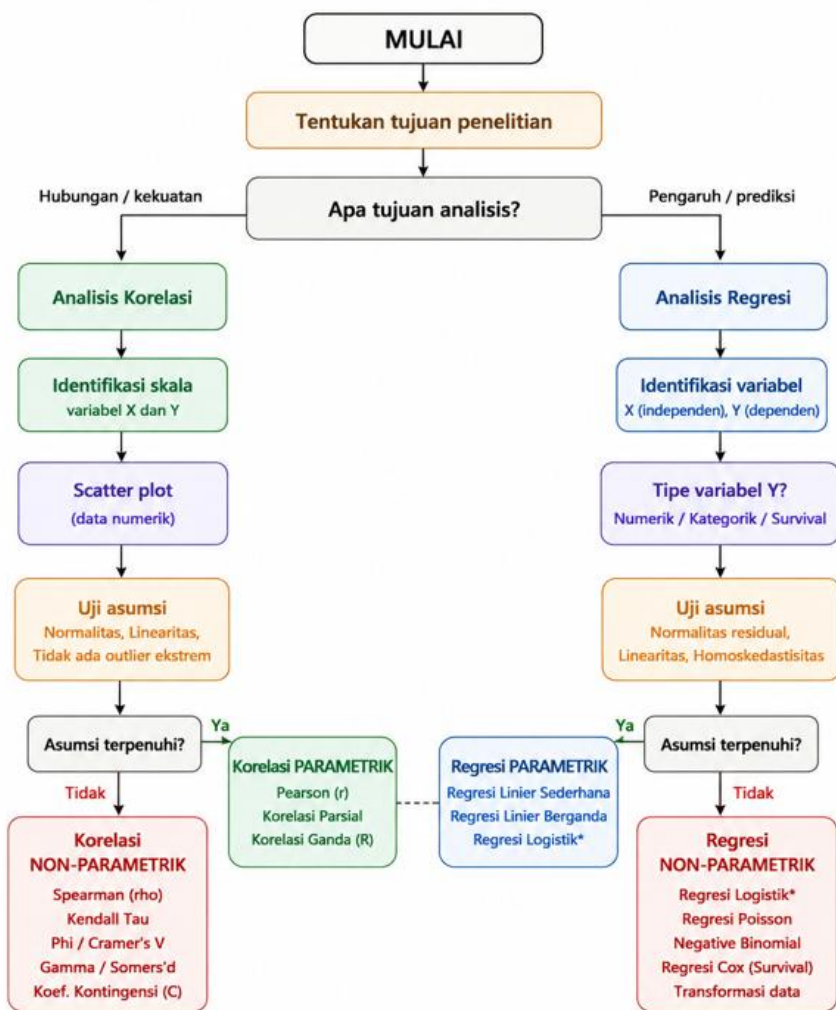
* Phi (ϕ) adalah kasus khusus dari Cramer's V untuk tabel 2×2. Koefisien Kontingensi (C) memiliki keterbatasan, Cramer's V lebih disarankan untuk pelaporan ilmiah modern.

Tabel 8. 4 Panduan Interpretasi Uji Korelasi (Dahlan, 2014)

Parameter	Nilai	Interpretasi
Kekuatan korelasi	0,0 - <0,2	Sangat lemah
	0,2 - <0,4	Lemah
	0,4 - <0,6	Sedang
	0,6 - <0,8	Kuat
	0,8 – 1,00	Sangat kuat
Arah korelasi	Positif	Semakin tinggi variabel A semakin tinggi variabel B

Parameter	Nilai	Interpretasi
	Negatif	Semakin tinggi variabel A semakin rendah variabel B
Nilai p	Nilai p >0,05	Korelasi tidak bermakna
	Nilai p <0,05	Korelasi bermakna
Kemaknaan klinis	r yang diperoleh < r minimal	Korelasi tidak bermakna
	r yang diperoleh > r minimal	Korelasi bermakna

Koefisien korelasi Pearson, yang umumnya dilambangkan dengan huruf r , digunakan untuk menilai kekuatan dan arah hubungan antara dua variabel numerik. Selain itu, uji ini juga dapat digunakan untuk mengetahui besarnya kontribusi variabel independen terhadap variabel dependen. Nilai koefisien korelasi Pearson berada pada rentang -1 hingga $+1$. Semakin mendekati nilai $+1$ atau -1 , maka hubungan antarvariabel semakin kuat, sedangkan nilai yang mendekati 0 menunjukkan hubungan yang semakin lemah. Nilai positif menunjukkan hubungan searah, sedangkan nilai negatif menunjukkan hubungan berlawanan arah (Saputra, 2016; Solikhah; and Amyati., 2022).



Gambar 8. 1 Bagan Keputusan Pemilihan Uji Korelasi dan Regresi (Dahlan, 2014; Kartiningrum *et al.*, 2020)

Beberapa contoh penerapan analisis korelasi, antara lain:

1. Peneliti menggunakan analisis korelasi Pearson untuk menilai hubungan antara indeks massa tubuh (IMT) dan tekanan darah sistolik pada populasi dewasa. Hasil analisis digunakan untuk mengetahui apakah peningkatan IMT berkaitan dengan peningkatan tekanan darah. Analisis ini membantu mengidentifikasi obesitas sebagai faktor yang berhubungan dengan hipertensi.
2. Peneliti menerapkan uji korelasi Spearman untuk mengevaluasi hubungan antara tingkat stres dan kualitas tidur pada mahasiswa. Analisis dilakukan karena data skor stres tidak berdistribusi normal. Hasil penelitian dapat memberikan gambaran mengenai pengaruh kondisi psikologis terhadap kualitas tidur.
3. Peneliti menggunakan analisis korelasi untuk menilai hubungan antara durasi aktivitas fisik dan kadar glukosa darah puasa pada pasien diabetes melitus tipe 2. Korelasi ini digunakan untuk mengetahui apakah peningkatan aktivitas fisik berkaitan dengan penurunan kadar glukosa darah. Hasil penelitian dapat mendukung pentingnya aktivitas fisik dalam pengendalian diabetes.

8.1.2 Uji Regresi

Uji regresi linier

Regresi linier sederhana menggambarkan hubungan antara satu variabel independen (X) dan satu variabel dependen (Y) dalam bentuk garis lurus. Model ini memungkinkan kita memprediksi nilai Y untuk nilai X tertentu. Adapun persamaan model regresi linier:

$$Y = a + bX$$

Keterangan:

Y = nilai prediksi variabel dependen

a = konstanta (intercept)

b = koefisien regresi (slope) (perubahan Y untuk setiap kenaikan 1 satuan X)

X = nilai variabel independent

Asumsi klasik regresi linier perlu diperiksa sebelum peneliti menginterpretasikan hasil model, karena pelanggaran asumsi dapat memengaruhi validitas estimasi koefisien, nilai signifikansi, dan kesimpulan penelitian. Pemeriksaan ini mencakup linearitas, normalitas residual, homoskedastisitas, dan independensi residual (Dahlan, 2014; Kartiningrum *et al.*, 2020).

Gambar 8. 2 Asumsi Klasik Regresi Linier dan Cara Pengujiannya

Asumsi	Cara Pengujian
Linearitas	Scatter plot X vs Y; Uji ANOVA Linearity di SPSS
Normalitas residual	Histogram residual; P-P Plot; Uji Kolmogorov-Smirnov
Homoskedastisitas	Scatter plot: Predicted vs Residual; pola acak menunjukkan asumsi terpenuhi
Independensi residual	Uji Durbin-Watson; nilai 1,5–2,5 menunjukkan tidak ada autokorelasi

Koefisien Determinasi (R^2)

Koefisien determinasi (R^2) merupakan ukuran yang digunakan untuk menunjukkan kemampuan model regresi dalam menjelaskan variasi variabel dependen (Y). Nilai R^2 berkisar antara 0 sampai 1, di mana semakin mendekati 1 menunjukkan bahwa model memiliki kemampuan yang semakin baik dalam menjelaskan data. Sebagai contoh, apabila diperoleh nilai R^2 sebesar 0,60, maka model regresi mampu menjelaskan 60% variasi pada variabel dependen, sedangkan sisanya sebesar 40% dipengaruhi oleh faktor atau variabel lain di luar model penelitian. Oleh karena itu, koefisien determinasi digunakan untuk menilai tingkat kecocokan (*goodness of fit*) model regresi terhadap data yang dianalisis (Kartiningrum *et al.*, 2020).

Uji Signifikansi Model

Uji signifikansi model dilakukan untuk menilai apakah model regresi dan koefisien regresi yang dihasilkan memiliki makna secara statistik. Dalam analisis regresi, pengujian ini umumnya dilakukan menggunakan uji F dan uji t. Uji F (ANOVA) digunakan untuk mengevaluasi signifikansi model regresi secara keseluruhan, yaitu untuk mengetahui apakah variabel independen secara simultan berpengaruh terhadap variabel dependen. Sementara itu, uji t digunakan untuk menilai signifikansi masing-masing koefisien regresi secara parsial. Keputusan pengujian didasarkan pada nilai *p-value*, di mana nilai $p < 0,05$ menunjukkan bahwa model atau koefisien regresi signifikan secara statistik.

Tabel 8. 5 Jenis Model Regresi Alternatif Berdasarkan Tipe Variabel Dependen (Y) (Dahlan, 2016; Kartiningrum *et al.*, 2020; Fitriangga, 2024)

Jenis Regresi	Klasifikasi	Variabel Dependen (Y)	Kegunaan
Regresi Logistik Biner	GLM / Parametrik	Dikotomi (2 kategori: 0/1)	Memprediksi probabilitas kejadian; menghitung odds ratio
Regresi Logistik Multinomial	GLM / Parametrik	Nominal >2 kategori (tanpa urutan)	Outcome kategorik tidak berurutan lebih dari 2 kelompok
Regresi Logistik Ordinal	GLM / Parametrik	Ordinal >2 kategori (ada urutan)	Outcome kategorik berurutan / berjenjang
Regresi Poisson	GLM / Parametrik	Data hitung (count), mean $\approx$ varians	Frekuensi kejadian; jika overdispersi $\rightarrow$ gunakan Negative Binomial
Regresi Negative Binomial	GLM / Parametrik	Data hitung dengan overdispersi (varians > mean)	Alternatif Poisson saat overdispersi; pemodelan varians lebih fleksibel

Jenis Regresi	Klasifikasi	Variabel Dependen (Y)	Kegunaan
Regresi Cox	Semi parametrik	Waktu hingga kejadian (time-to-event)	Faktor yang memengaruhi hazard / waktu survival; tidak perlu asumsi distribusi waktu
Quantile Regression	Non parametrik	Numerik kontinu (semua kuantil, terutama median)	Regresi pada median atau kuantil lain; robust terhadap outlier dan distribusi tidak normal
LOESS / LOWESS	Non parametrik	Numerik kontinu	Regresi lokal tanpa asumsi bentuk kurva; untuk hubungan non-linier kompleks

Keterangan:

GLM/ Parametrik (Generalized Linear Model): mengasumsikan distribusi dari keluarga eksponensial. Dipilih berdasarkan *tipe Y*, bukan karena asumsi normalitas gagal; Semiparametrik (tidak mengasumsikan distribusi waktu survival); hanya mengasumsikan proportional hazard; Nonparametrik tidak mengasumsikan distribusi data maupun bentuk hubungan tertentu; LOESS = Locally Estimated Scatterplot Smoothing; LOWESS = Locally Weighted Scatterplot Smoothing

Analisis Regresi Linier Berganda

Regresi linier berganda merupakan pengembangan dari regresi linier sederhana yang digunakan untuk menganalisis hubungan antara satu variabel dependen dengan dua atau lebih variabel independen. Model ini dinyatakan dalam persamaan $\hat{Y} = a + b_1X_1 + b_2X_2 + \dots + b_kX_k$, dengan $\hat{Y}$ sebagai nilai prediksi variabel dependen, a sebagai konstanta, serta b_1 sampai b_k sebagai koefisien regresi parsial dari setiap variabel independen. Koefisien b_i menunjukkan besarnya perubahan nilai Y akibat perubahan satu satuan pada X_i , dengan variabel independen lainnya dikendalikan atau dianggap konstan. Dengan demikian, regresi linier berganda memungkinkan peneliti menilai kontribusi masing-masing variabel prediktor secara simultan dalam menjelaskan variasi variabel dependen (Kartiningrum *et al.*, 2020).

Dalam regresi linier berganda, salah satu asumsi penting yang perlu diperhatikan adalah tidak adanya multikolinearitas antarvariabel independen. Multikolinearitas terjadi ketika dua atau lebih variabel independen saling berkorelasi kuat sehingga dapat menyebabkan estimasi koefisien regresi menjadi tidak stabil dan sulit diinterpretasikan. Pemeriksaan multikolinearitas umumnya dilakukan menggunakan nilai *Variance Inflation Factor* (VIF) dan *tolerance*. Nilai VIF lebih dari 10 atau nilai *tolerance* kurang dari 0,10 menunjukkan adanya indikasi multikolinearitas yang perlu diperhatikan dalam model.

Pada regresi linier berganda, ukuran kecocokan model lebih tepat dinilai menggunakan *Adjusted R²* dibandingkan R^2 biasa. Hal ini karena R^2 cenderung meningkat setiap kali variabel independen baru ditambahkan ke dalam model, meskipun variabel

tersebut tidak relevan secara substantif. *Adjusted R²* memperhitungkan jumlah variabel independen dalam model sehingga memberikan estimasi yang lebih objektif terhadap kemampuan model dalam menjelaskan variasi variabel dependen. Oleh karena itu, *Adjusted R²* lebih disarankan dalam evaluasi model regresi berganda, terutama ketika model melibatkan banyak prediktor (Dahlan, 2014).

Beberapa contoh penerapan analisis regresi sederhana, antara lain:

1. Peneliti menggunakan regresi linier sederhana untuk memprediksi tekanan darah sistolik berdasarkan usia pada populasi lansia. Variabel usia digunakan sebagai prediktor, sedangkan tekanan darah sebagai variabel dependen. Model regresi membantu memperkirakan perubahan tekanan darah yang terjadi seiring bertambahnya usia.
2. Peneliti menerapkan regresi linier sederhana untuk menganalisis hubungan antara kadar vitamin D dan kepadatan tulang pada wanita menopause. Analisis ini digunakan untuk mengetahui seberapa besar perubahan kepadatan tulang akibat peningkatan kadar vitamin D. Hasil penelitian dapat digunakan sebagai dasar intervensi kesehatan tulang.
3. Peneliti menggunakan regresi linier sederhana untuk menilai pengaruh durasi olahraga terhadap penurunan kadar kolesterol total pada pasien hiperlipidemia. Model regresi digunakan untuk memperkirakan besarnya perubahan kadar kolesterol berdasarkan peningkatan durasi aktivitas fisik.

Adapun contoh penerapan regresi linier berganda, antara lain:

1. Peneliti menggunakan regresi linier berganda untuk menganalisis pengaruh usia, indeks massa tubuh, dan aktivitas fisik terhadap tekanan darah sistolik. Analisis ini memungkinkan peneliti menilai kontribusi masing-masing faktor secara simultan terhadap perubahan tekanan darah.
2. Peneliti menerapkan regresi linier berganda untuk mengevaluasi faktor-faktor yang memengaruhi kadar glukosa darah puasa pada pasien diabetes melitus tipe 2. Variabel independen yang dianalisis meliputi pola makan, aktivitas fisik, dan kepatuhan konsumsi obat. Model regresi digunakan untuk menentukan faktor yang paling dominan memengaruhi kadar glukosa darah.
3. Peneliti menggunakan regresi linier berganda untuk menilai pengaruh konsumsi natrium, kebiasaan merokok, dan tingkat stres terhadap risiko hipertensi pada pekerja industri. Analisis ini membantu menjelaskan hubungan kompleks antara faktor gaya hidup dan kondisi kesehatan kardiovaskular.

Uji regresi logistik

Regresi logistik merupakan metode analisis statistik yang digunakan untuk menganalisis hubungan antara variabel dependen kategorik dengan satu atau lebih variabel independen, baik berskala kategorik maupun kontinu. Berdasarkan jenis variabel dependennya, regresi logistik dibedakan menjadi regresi logistik biner, multinomial, dan ordinal. Berbeda dengan regresi linear (*Ordinary Least Squares/OLS*), regresi logistik digunakan untuk variabel dependen kategorik, terutama dikotomik seperti ya/tidak, baik/buruk, atau

sakit/tidak sakit, serta tidak mensyaratkan residual berdistribusi normal karena model didasarkan pada distribusi logistik (Solikhah; and Amyati., 2022).

Regresi logistik menggunakan transformasi logit untuk memperkirakan peluang (*odds*) suatu kejadian sehingga hubungan antara variabel independen dan dependen tidak harus linier secara langsung. Selain itu, regresi logistik tidak memerlukan asumsi normalitas multivariat maupun homoskedastisitas, dan variabel independen dapat berupa data numerik maupun kategorik yang bersifat saling eksklusif. Dalam penerapannya, regresi logistik digunakan untuk menilai pengaruh variabel prediktor terhadap variabel respons, baik secara parsial maupun simultan, yang umumnya dievaluasi menggunakan uji Chi-Square model (Fitriangga, 2024).

Beberapa contoh penerapan analisis regresi sederhana, antara lain:

1. Peneliti menggunakan regresi logistik biner untuk menganalisis faktor-faktor yang memengaruhi kejadian hipertensi pada populasi dewasa. Variabel dependen dikategorikan menjadi hipertensi dan tidak hipertensi, sedangkan variabel independen meliputi usia, indeks massa tubuh, kebiasaan merokok, dan aktivitas fisik. Analisis ini digunakan untuk menentukan faktor risiko yang paling berpengaruh terhadap kejadian hipertensi.
2. Peneliti menerapkan regresi logistik untuk mengevaluasi faktor-faktor yang berhubungan dengan kepatuhan minum obat pada pasien diabetes melitus tipe 2. Variabel dependen dibagi menjadi patuh dan tidak patuh, sedangkan variabel independen meliputi tingkat pendidikan, dukungan keluarga, lama menderita diabetes, dan jumlah obat yang dikonsumsi. Model regresi logistik digunakan

untuk menghitung *odds ratio* dari masing-masing faktor prediktor.

3. Peneliti menggunakan regresi logistik dalam penelitian epidemiologi untuk menilai faktor-faktor yang memengaruhi kejadian infeksi saluran pernapasan akut (ISPA) pada anak usia prasekolah. Variabel dependen berupa kejadian ISPA (ya/tidak), sedangkan variabel independen mencakup status imunisasi, paparan asap rokok, kepadatan hunian, dan kebiasaan mencuci tangan. Analisis ini membantu mengidentifikasi faktor risiko yang berkontribusi terhadap peningkatan kejadian ISPA pada anak.

DAFTAR PUSTAKA

- Dahlan, S.M. (2014) *Statistik untuk Kedokteran dan Kesehatan*. 6th edn. Edited by A. Susila. Jakarta: Epidemiologi Indonesia.
- Dahlan, S.M. (2016) *Analisis Multivariat Regresi Logistik*. Cetakan Ke. Jakarta: Epidemiologi Indonesia.
- Fitriangga, A. (2024) *Biostatistik untuk Mahasiswa Kesehatan*. I. Medan: PT. Media Penerbit Indonesia.
- Kartiningrum, E.D. *et al.* (2020) *Aplikasi Regresi dan Korelasi Dalam Analisis Data Hasil Penelitian*. 1st edn. Edited by R.L. Mahmudah. Mojokerto: STIKES Majapahit Mojokerto.
- Saputra, R. (2016) *Buku Ajar Biostatistik Dasar Dalam Ilmu Kesehatan Masyarakat*. 1st edn, *Distribusi frekuensi*. 1st edn. Batam.
- Solikhah; and Amyati. (2022) *Biostatistik (Sebuah Aplikasi SPSS dalam Bidang Kesehatan dan Kedokteran)*. I. Edited by R.A. Nugroho. Yogyakarta: Jejak Pustaka.

BAB 9

BIOSTATISTIK DALAM EPIDEMIOLOGI DAN KESEHATAN MASYARAKAT

Oleh Dwi Yulia Maritasari

Biostatistik adalah cabang ilmu statistik yang menerapkan prinsip analisis data salah satunya dibidang kesehatan masyarakat. Cabang ilmu statistik ini berperan dalam mendeskripsikan masalah kesehatan, menilai efektivitas intervensi, dan mendukung pengambilan keputusan berbasis *evidence based decision making*/berbasis bukti. Biostatistik mengubah data menjadi sebuah informasi yang digunakan untuk pengambilan keputusan dalam dunia kedokteran maupun kesehatan masyarakat (Ernawati *et al.*, 2025). Dalam perkembangan ilmu kesehatan modern, biostatistik tidak hanya berfungsi sebagai alat analisis data, akan tetapi menjadi komponen penting dalam perencanaan penelitian, pengukuran derajat kesehatan, identifikasi determinan kesehatan, evaluasi intervensi kesehatan, serta merumuskan kebijakan kesehatan (Rosner, 2006). Biostatistik memungkinkan peneliti mengubah sebuah data menjadi informasi ilmiah yang dapat digunakan dalam memahami masalah kesehatan secara objektif dan terukur.

Biostatistik memiliki peranan penting dalam ilmu epidemiologi mengingat epidemiologi merupakan ilmu yang mempelajari distribusi dan faktor-faktor yang

mempengaruhi kejadian penyakit. Faktor risiko dan kejadian penyakit dapat dianalisis secara sistematis dan objektif melalui pendekatan statistik. Epidemiologi memerlukan metode statistik yang tepat untuk mengidentifikasi hubungan antara faktor risiko dengan kejadian penyakit. Sehingga dapat dikatakan bahwa, biostatistik dan epidemiologi merupakan disiplin ilmu yang saling melengkapi dalam menghasilkan kerangka konseptual dalam memahami masalah kesehatan di populasi (Friis & Sellers, 2020).

Dalam konteks kesehatan masyarakat, biostatistik berperan penting dalam menghasilkan informasi mengenai berbagai indikator kesehatan yang menjadi dasar dalam perencanaan dan evaluasi program kesehatan, seperti angka insiden, prevalensi, mortalitas, case fatality rate (CFR), angka harapan hidup yang diperoleh melalui analisis statistik. Analisis statistik yang akurat dapat mengidentifikasi tren penyakit, prediksi KLB, dan evaluasi intervensi kesehatan yang sangat penting bagi pemerintah dan pemangku kepentingan informasi tersebut digunakan untuk menentukan prioritas program, alokasi sumber daya, dan merancang strategi yang efektif (Rothman *et al.*, 2008).

Dengan demikian, biostatistik merupakan fondasi metodologi yang penting dalam epidemiologi dan kesehatan masyarakat. Penerapan metode statistik yang tepat dapat mengolah data kesehatan menjadi bukti ilmiah yang mendukung pengembangan kebijakan dan peningkatan derajat kesehatan masyarakat.

9.1 Konsep Dasar Biostatistik

Statistika merupakan pengetahuan yang berkaitan dengan proses pengumpulan data, pengolahan data, dan analisisnya yang digunakan untuk menarik kesimpulan berdasarkan data. Pada penelitian kuantitatif yang datanya berupa angka-angka, data tersebut belum memberikan

informasi jika belum dilakukan analisis yang tepat. Pengetahuan mengenai cara analisis data tersebut yang dinamakan statistika atau ilmu statistika (Ahmad & Jaya, 2021). Berdasarkan informasi yang dihasilkan melalui data, maka statistika dapat dibedakan menjadi 2, yaitu :

1. Statistika deskriptif, yaitu statistik yang digunakan untuk menggambarkan masalah atau pengamatan yang tidak sampai pada penarikan kesimpulan. Statistik deskriptif ini hanya menggambarkan data apa adanya dengan menunjukkan distribusi dari data yang disajikan dalam bentuk nilai rata-rata, median, standar deviasi, varians, nilai minimum, nilai maksimum, dll.
2. Statistik inferensial, yaitu statistika yang dilakukan dengan menganalisis data dari sampel, kemudian hasilnya digeneralisasikan ke populasi.

Biostatistika merupakan ilmu statistik yang berfokus pada penerapan dalam penelitian dan praktik kesehatan. Biostatistik digunakan untuk menganalisis data yang berkaitan dengan penyakit, faktor resiko, efektivitas intervensi kesehatan, evaluasi program kesehatan, serta berbagai masalah kesehatan lainnya (Daniel & Cross, 2018). Biostatistik berperan penting dalam mengubah data menjadi sumber informasi yang bermakna sebagai dasar pengambilan keputusan berbasis bukti. Biostatistika membantu peneliti merancang penelitian, menentukan ukuran sampel, menguji hipotesis, dan menarik kesimpulan yang valid dan dapat dipertanggungjawabkan secara ilmiah (Rosner, 2006).

Biostatistik mencakup seluruh tahapan pengolahan data kesehatan, mulai dari perencanaan penelitian hingga interpretasi hasil analisis. Dalam praktiknya, biostatistik digunakan untuk merancang penelitian kesehatan, menentukan teknik pengambilan sampel, mengembangkan instrumen pengumpulan data, mengolah dan membersihkan data (data cleaning), menganalisis data menggunakan metode statistik yang sesuai, menafsirkan hasil analisis, dan

menyajikan hasil penelitian dalam bentuk tabel, grafik, maupun laporan ilmiah (Pagano *et al.*, 2022).

9.2 Hubungan Biostatistik dalam Epidemiologi

Biostatistika merupakan ilmu yang menerapkan statistik dalam bidang biologi maupun bidang kesehatan. Biostatistik digunakan untuk merancang suatu penelitian, mengumpulkan, menganalisis, dan menginterpretasikan data agar didapatkan kesimpulan yang valid sebagai dasar pengambilan keputusan ilmiah (Bhattacharya, 2021). Penerapan biostatistik tidak hanya sebatas angka dan grafik, akan tetapi menggali makna yang lebih mendalam dari data yang tersedia. Biostatistika digunakan untuk menjawab permasalahan penelitian, dan jawaban dari pertanyaan tersebut membutuhkan analisis data yang sistematis. Kekuatan dari biostatistik adalah kemampuan dalam memberikan landasan yang rasional, karena angka-angka memberikan kepastian yang akurat (Islamiyati & Sadiman, 2025).

Epidemiologi merupakan ilmu yang mempelajari distribusi, penyebab, dan pencegahan di masyarakat. Ilmu epidemiologi membantu dalam pemahaman terhadap kesehatan dan faktor-faktor yang mempengaruhinya, baik di tingkat individu maupun di populasi. Epidemiologi merupakan cabang ilmu kesehatan yang berfokus pada analisis distribusi dan determinan berbagai masalah kesehatan dalam populasi serta dampaknya terhadap masyarakat. Sebagai pilar utama kesehatan masyarakat, epidemiologi menyediakan kerangka ilmiah untuk mengevaluasi faktor risiko, memahami mekanisme terjadinya penyakit, dan menghasilkan bukti yang mendukung kebijakan serta intervensi kesehatan yang efektif. Dengan demikian, epidemiologi berkontribusi dalam pengembangan strategi pencegahan, pengendalian, dan

pengobatan penyakit yang berbasis bukti ilmiah (Karlinda & Putri, 2025).

Dalam epidemiologi dan kesehatan masyarakat, ruang lingkup biostatistik juga mencakup pengukuran frekuensi penyakit, identifikasi faktor risiko, evaluasi program kesehatan, analisis tren penyakit, serta permodelan kejadian penyakit dalam populasi (Friis & Sellers, 2020). Epidemiologi dan biostatistik merupakan disiplin yang saling melengkapi. Epidemiologi menyediakan kerangka konseptual untuk memahami distribusi dan determinan penyakit, sedangkan biostatistik menyediakan metode kuantitatif untuk menguji dan memvalidasi hubungan tersebut secara ilmiah (Rothman *et al.*, 2008).

9.2.1 Epidemiologi sebagai Ilmu Dasar Kesehatan Masyarakat

Epidemiologi merupakan salah satu disiplin ilmu dalam kesehatan masyarakat karena berkaitan dengan distribusi, frekuensi, dan determinan masalah kesehatan. Epidemiologi berperan dalam mengidentifikasi masalah kesehatan, faktor resiko, perencanaan dan evaluasi program kesehatan, dan pengambilan keputusan berbasis evidence based. Metode epidemiologi mampu menjelaskan hubungan sebab akibat antara faktor resiko dan penyakit sehingga memiliki posisi yang penting dalam kesehatan masyarakat. Epidemiologi tidak hanya berfungsi menjelaskan hubungan sebab akibat penyakit saja, akan tetapi menjadi alat dalam pembangunan kesehatan masyarakat dalam upaya peningkatan drajat kesehatan masyarakat (Gouda & Powles, 2014).

Epidemiologi memiliki peran utama dalam kesehatan masyarakat sebagai dasar dalam mengidentifikasi masalah kesehatan, perencanaan program, pelaksanaan intervensi, evaluasi, dan

pengambilan kebijakan kesehatan berbasis bukti. Berikut peran epidemiologi dalam kesehatan masyarakat (Friis & Sellers, 2020).

1. Mengidentifikasi masalah kesehatan masyarakat
2. Mengidentifikasi faktor resiko penyakit
3. Penyediaan data untuk perencanaan program kesehatan
4. Mendukung sistem surveilans dan kewaspadaan dini
5. Evaluasi program intervensi kesehatan
6. Dasar pengambilan kebijakan kesehatan
7. Memprediksi masalah kesehatan di masa yang akan datang
8. Landasan ilmiah dalam pengembangan ilmu kesehatan masyarakat

9.2.2 Peran Biostatistik dalam Epidemiologi

Biostatistik dan epidemiologi merupakan dua disiplin ilmu yang saling berkaitan erat. Kedua bidang ilmu tersebut saling melengkapi dalam menghasilkan bukti ilmiah yang valid dalam upaya mendukung pengambilan keputusan berdasarkan evidence based. Epidemiologi berkaitan dengan desain penelitian, pengukuran variabel, identifikasi faktor resiko, serta mengukur hubungan sebab akibat. Sedangkan, biostatistika berkaitan dengan metode kuantitatif analisis data, pengolahan data, serta menginterpretasikan data yang berkaitan dengan kesehatan (Goldstein *et al.*, 2020). Perkembangan epidemiologi modern tidak dapat dipisahkan dari kemajuan metode biostatistik yang memungkinkan peneliti memperoleh hasil penelitian yang lebih akurat, objektif, dan dapat dipertanggungjawabkan secara ilmiah (Friis & Sellers, 2020). Berikut peran biostatistika

dalam epidemiologi sebagai berikut (Celentano *et al.*, 2023).

1. Pengukuran distribusi penyakit melalui indikator epidemiologis, seperti prevalensi, insidensi, dan mortalitas.
2. Menganalisis tingkat risiko serta kekuatan hubungan antara faktor paparan dan kejadian penyakit.
3. Menguji hipotesis penelitian sehingga dapat menentukan signifikansi suatu temuan ilmiah.
4. Menilai validitas hasil penelitian dengan mengidentifikasi dan mengendalikan berbagai sumber kesalahan, termasuk bias dan faktor perancu (*confounding factors*).
5. Mengevaluasi keberhasilan dan efektivitas program kesehatan serta menyediakan bukti ilmiah yang akurat sebagai dasar dalam perumusan kebijakan.
6. Pengambilan keputusan kesehatan masyarakat berbasis bukti (*evidence-based public health*).

9.2.3 Biostatistik dalam Evidence-Based Public Health

Evidence-Based Public Health (EBPH) merupakan pendekatan dalam kesehatan masyarakat yang berfokus pada penggunaan bukti ilmiah empiris yang tersedia untuk mendukung pengambilan keputusan, perencanaan program, implementasi intervensi, dan evaluasi kebijakan kesehatan. Komponen utama EBPH meliputi penggunaan bukti ilmiah yang valid, pemanfaatan data dan sistem informasi secara sistematis, evaluasi program yang tepat, serta penerapan hasil penelitian dalam praktik kesehatan masyarakat (Brownson *et al.*, 2009a). Melalui metode statistik, data yang dikumpulkan dari survei kesehatan, surveilans epidemiologi, penelitian observasional, maupun eksperimental dapat dianalisis secara objektif sehingga

menghasilkan bukti yang valid dan dapat dipertanggungjawabkan secara ilmiah (Matranga *et al.*, 2021).

Evidence-Based Public Health merupakan proses pengambilan keputusan dibidang kesehatan masyarakat yang didasarkan pada bukti ilmiah menggunakan data yang sistematis, sehingga biostatistik menjadi landasan yang memungkinkan bukti ilmiah tersebut dihasilkan secara objektif. Peran utama biostatistik dalam Evidence-Based Public Health sebagai berikut (Brownson *et al.*, 2009b).

1. Mengidentifikasi masalah kesehatan masyarakat melalui prevalensi, insidensi, mortalitas, dan indikator kesehatan lainnya.
2. Mengidentifikasi faktor resiko dan determinan kesehatan melalui metode analisis statistik.
3. Mengevaluasi efektivitas program dan kebijakan kesehatan sehingga dapat program dapat diukur secara objektif.
4. Analisis biostatistik membantu pemangku kepentingan untuk pengambilan keputusan berbasis bukti ilmiah yang terukur.
5. Memprediksi tren dan resiko kesehatan .
6. Biostatistik berperan dalam penentuan desain penelitian, ukuran sampel, teknik pengambilan sampel, analisis dan interpretasi data sehingga menghasilkan data yang dapat dipertanggungjawabkan secara ilmiah.

Biostatistik merupakan komponen yang penting dalam *evidence-based public health* karena menyediakan metode ilmiah untuk mengubah data menjadi informasi yang digunakan dalam pengambilan keputusan dan penyusunan kebijakan kesehatan. Tanpa adanya biostatistik, maka proses pengambilan keputusan

berbasis bukti tidak dapat berjalan optimal (Brownson *et al.*, 2009b).

9.3 Data dalam Epidemiologi

Data merupakan komponen yang penting dalam epidemiologi untuk menggambarkan distribusi penyakit serta faktor-faktor resiko yang mempengaruhinya. Epidemiologi tidak hanya mempelajari kejadian penyakit, akan tetapi juga bergantung pada kualitas data yang diperoleh selama proses pengumpulan data, pengolahan, dan analisis data. Data epidemiologi merupakan informasi mengenai distribusi dan faktor resiko penyakit yang diperoleh dari berbagai sumber seperti kegiatan surveilans, survei populasi, rekan medis, hingga statistik vital. Data tersebut digunakan untuk mengidentifikasi pola kejadian penyakit berdasarkan orang, tempat, dan waktu maupun analisis hubungan faktor resiko dengan kejadian penyakit (Celentano *et al.*, 2023). Data epidemiologi yang baik harus memenuhi prinsip validitas, reliabilitas, akurasi, kelengkapan (completeness), konsistensi, serta ketepatan waktu (timeliness). Data yang tidak memenuhi karakteristik tersebut berpotensi menimbulkan bias sehingga dapat memengaruhi interpretasi hasil penelitian maupun penyusunan kebijakan kesehatan (Aschengrau & Seage, 2013).

Berdasarkan sifatnya, data dibedakan menjadi data kualitatif dan data kuantitatif. Data kualitatif merupakan data kategorik yang menggambarkan karakteristik individu tanpa memiliki nilai numerik, misalnya jenis kelamin, status merokok, jenis pekerjaan, atau status imunisasi. Sebaliknya, data kuantitatif merupakan data yang dinyatakan dalam bentuk angka sehingga dapat dilakukan analisis statistik (Szklo *et al.*, 2014). Berdasarkan sumbernya, data epidemiologi terdiri atas data primer dan data sekunder. Data

primer dikumpulkan secara langsung oleh peneliti melalui wawancara, observasi, pemeriksaan fisik, pengukuran antropometri, maupun pemeriksaan laboratorium. Data sekunder diperoleh dari sumber yang telah tersedia sebelumnya, seperti sistem surveilans penyakit, laporan rumah sakit, registrasi penyakit, sensus penduduk, maupun data rutin dari instansi pemerintah (Friis & Sellers, 2020).

Dengan demikian, kualitas data dan ketepatan pengukuran menjadi fondasi utama dalam menghasilkan bukti ilmiah yang valid. Data yang berkualitas akan meningkatkan ketepatan estimasi frekuensi penyakit, identifikasi faktor risiko, evaluasi intervensi, serta penyusunan kebijakan kesehatan masyarakat berbasis bukti.

9.4 Ukuran Epidemiologi dalam Biostatistik

Keakuratan pengukuran dalam epidemiologi dipengaruhi oleh kejelasan definisi kasus dan ketepatan penentuan populasi berisiko. Definisi kasus yang terstandar diperlukan untuk menghindari kesalahan interpretasi serta memastikan validitas perbandingan data antarwaktu maupun antarpopulasi. Oleh karena itu, sebelum menetapkan suatu kasus, penting untuk memahami prosedur dan kriteria diagnostik yang digunakan dalam mengidentifikasi penyakit, sehingga proses pengumpulan dan analisis data epidemiologi dapat dilakukan secara valid dan reliabel. Selain itu, Penentuan populasi berisiko merupakan aspek yang sangat penting dalam pengukuran dan penelitian epidemiologi. Populasi berisiko adalah kelompok individu yang memiliki peluang untuk mengalami atau terserang suatu penyakit tertentu. Identifikasi yang tepat terhadap populasi ini memungkinkan peneliti memperoleh gambaran yang akurat mengenai distribusi masalah kesehatan serta besarnya risiko penyakit dalam suatu populasi. (Najmah, 2015).

9.4.1 Pengukuran Frekuensi Penyakit

1. Rasio

Rasio adalah perbandingan suatu peristiwa (event) sebagai numerator (a) dan peristiwa lainnya yang berhubungan sebagai denominator (b).

$$ratio = \frac{a}{b}$$

2. Proporsi

Proporsi adalah bagian dari suatu peristiwa atau ukuran yang membandingkan suatu peristiwa sebagai numerator (a) dan peristiwa lainnya sebagai denominator yang mengandung peristiwa numerator (a+b) dikali dengan konstanta (100).

$$Proporsi = \frac{a}{a+b} \times k$$

3. Rate

Rate adalah suatu ukuran frekuensi kejadian pada populasi tertentu selama periode waktu tertentu. Rate sangat berguna untuk membandingkan frekuensi penyakit di lokasi, waktu atau antar kelompok orang yang berbeda dengan ukuran populasi yang berbeda.

Beberapa ahli epidemiologi menyatakan bahwa rate menggambarkan seberapa cepat penyakit terjadi dalam suatu populasi.

Rate

$$= \frac{\text{jumlah orang yang mengalami peristiwa pada periode tertentu}}{\text{jumlah populasi beresiko sesuai numerator dalam waktu tertentu}} \times k$$

K = 100, 1000, 10.000, 100.000 dapat dipilih salah satu sesuai kebutuhan untuk menghindari pecahan.

9.4.2 Insidens dan Prevalens

1. Insidens (*Insidance*)

Insidens adalah angka kasus baru suatu penyakit dari populasi yang berisiko selama periode waktu tertentu. Insidens dibagi menjadi 2, yaitu *incidence rate* dan *incidence cumulative*.

Incidence rate menunjukkan kecepatan kejadian suatu penyakit dan digunakan untuk menghitung insiden dimana kasus baru muncul diamati suatu populasi berisiko dengan periode pengamatan yang berbeda-beda.

$$IR = \frac{\text{Jumlah kasus baru}}{\text{populasi yang berisiko}} \times k$$

Keterangan :

IR = *Incidence Rate*

Jumlah Kasus Baru = Total orang yang baru terdiagnosis penyakit pada periode yang diamati.

Total *Person - Time* = Total jumlah waktu setiap orang yang berisiko diamati

K = Konstanta 100, 1000, 10.000, 100.000 (biasanya disesuaikan dengan skala populasi)

Incidence cumulative merupakan perbandingan antara jumlah kasus baru yang terjadi selama periode tertentu dengan jumlah penduduk yang berisiko pada awal periode pengamatan. Ukuran ini sering digunakan untuk mengukur risiko individu terkena penyakit dalam suatu populasi

$$CI = \frac{\text{Jumlah kasus baru selama periode tertentu}}{\text{jumlah populasi berisiko pada awal periode}} \times 100\%$$

2. Prevalensi (*Prevalence*)

Prevalensi adalah proporsi seluruh kasus penyakit (kasus lama dan baru) yang terdapat dalam populasi pada waktu tertentu.

$$\text{Prevalensi} = \frac{\text{jumlah seluruh kasus}}{\text{jumlah populasi}} \times k$$

K = Konstanta 100, 1000, 10.000, 100.000 (biasanya disesuaikan dengan skala populasi)

9.4.3 Ukuran Mortalitas dan Morbiditas

Mortalitas merupakan ukuran yang berkaitan dengan kejadian kematian dalam suatu populasi.

1. Angka Kematian Kasar (*Crude Death Rate*)

Angka kematian kasar (*crude death rate*) merupakan ukuran yang digunakan untuk memperkirakan proporsi kematian yang terjadi dalam suatu populasi selama periode waktu tertentu. Ukuran ini dapat digunakan untuk menggambarkan seluruh kejadian kematian maupun kematian yang disebabkan oleh faktor atau penyakit tertentu. Perhitungan angka kematian kasar dilakukan dengan menggunakan rumus sebagai berikut.

$$\text{CDR} = \frac{\text{jumlah kematian}}{\text{jumlah populasi pada pertengahan tahun}} \times 1000$$

2. Angka Kematian Spesifik (*Specific Death Rate*)

Angka Kematian Spesifik (*Specific Death Rate*) digunakan untuk mengukur kematian berdasarkan karakteristik tertentu, seperti umur, jenis kelamin, atau penyebab kematian. Perhitungan Angka Kematian Spesifik (*Specific Death Rate*) menggunakan rumus sebagai berikut.

$$\text{SDR} = \frac{\text{jumlah kematian pada kelompok usia dan jenis kelamin}}{\text{jumlah populasi pada kelompok usia dan jenis kelamin}} \times 1000$$

3. Angka Kematian Neonatal (Neonatal Mortality Rate)
 Angka kematian neonatal atau bayi baru lahir adalah jumlah kematian bayi usia kurang dari 28 hari pada periode tertentu, biasanya dalam satu tahun per 1.000 kelahiran hidup pada tahun yang sama.
 Perhitungan angka kematian neonatal diformulasikan pada rumus sebagai berikut.

$$\text{Angka Kematian Neonatal} = \frac{\text{Jumlah Kematian bayi dibawah 28 hari}}{\text{Kelahiran hidup}} \times 1000$$

4. Angka Kematian Bayi (*Infant Mortality Rate/IMR*)
 Mengukur jumlah kematian bayi usia kurang dari satu tahun per 1.000 kelahiran hidup. Perhitungan angka kematian bayi diformulasikan pada rumus sebagai berikut.

$$\text{Angka Kematian Bayi} = \frac{\text{Jumlah Kematian bayi} \leq 1 \text{ tahun}}{\text{Kelahiran hidup}} \times 1000$$

Morbiditas menunjukkan tingkat kesakitan dalam suatu populasi.

1. Attack Rate

Digunakan terutama pada investigasi wabah atau Kejadian Luar Biasa (KLB)

$$\text{Attack Rate} = \frac{\text{jumlah kasus selama wabah}}{\text{jumlah penduduk yang terpapar}} \times 100\%$$

2. *Case Fatality Rate (CFR)*

CFR adalah proporsi penderita penyakit tertentu yang meninggal akibat penyakit tersebut dalam periode tertentu.

$$CFR = \frac{\text{jumlah kematian akibat penyakit}}{\text{jumlah kasus penyakit}} \times 100\%$$

9.4.4 Pengukuran Risiko Penyakit

Pengukuran risiko penyakit digunakan untuk menilai hubungan antara faktor risiko dan kejadian penyakit.

1. *Relative Risk (RR)*

RR digunakan pada studi kohort untuk membandingkan risiko penyakit antara kelompok terpapar dan tidak terpapar. Relative risk disebut juga rasio insidensi kumulatif/Cummulative Incidence Ratio (CIR). Berikut rumus Risk Rasio.

$$RR = \frac{\text{Insidensi kumulatif kelompok terpapar } (\frac{a}{N_1})}{\text{Insidensi kumulatif kelompok tidak terpapar } (\frac{b}{N_0})}$$

Interpretasi:

RR = 1 → tidak ada hubungan

RR > 1 → faktor risiko meningkatkan risiko penyakit

RR < 1 → faktor bersifat protektif

2. *Odds Ratio (OR)*

Odds Ratio (OR) umumnya digunakan dalam penelitian dengan desain kasus-kontrol untuk mengukur kekuatan hubungan antara faktor risiko dan kejadian penyakit. Odds merupakan perbandingan antara dua kemungkinan pada variabel yang bersifat dikotomi. Pada kelompok kasus, odds dihitung sebagai perbandingan antara jumlah individu yang terpapar dengan jumlah

individu yang tidak terpapar. Sementara itu, pada kelompok kontrol, odds merupakan perbandingan antara jumlah individu yang terpapar dengan jumlah individu yang tidak terpapar pada kelompok kontrol. Nilai OR kemudian diperoleh dari perbandingan odds pada kelompok kasus terhadap odds pada kelompok kontrol.

$$OR = \frac{\text{Odds Kasus } (\frac{a}{b})}{\text{Odds Kontrol } (\frac{c}{d})}$$

Interpretasi:

OR = 1 → tidak ada hubungan

OR > 1 → terdapat hubungan positif

OR < 1 → faktor bersifat protektif

DAFTAR PUSTAKA

- Ahmad, A., & Jaya, I. (2021). *Biostatistik: Statistik dalam penelitian kesehatan*. Prenada Media. <https://www.google.com/books?hl=id&lr=&id=PNpBEAAAQBAJ&oi=fnd&pg=PA1&dq=Biostatistik:+Statistik+dalam+Penelitian+Kesehatan&ots=NYgRoHNln0&sig=xsPQHOjRTkG7AqFnb8JrAwxsjtU>
- Aschengrau, A., & Seage, G. R. (2013). *Essentials of Epidemiology in Public Health*. Jones & Bartlett Learning.
- Bhattacharya, P. K. (2021). Research Methodology in the Health Sciences: A Quick Reference Guide. (No Title). <https://cir.nii.ac.jp/crid/1130854726071122967>
- Brownson, R. C., Fielding, J. E., & Maylahn, C. M. (2009a). Evidence-Based Public Health: A Fundamental Concept for Public Health Practice. *Annual Review of Public Health*, 30(1), 175–201. <https://doi.org/10.1146/annurev.publhealth.031308.100134>
- Brownson, R. C., Fielding, J. E., & Maylahn, C. M. (2009b). Evidence-Based Public Health: A Fundamental Concept for Public Health Practice. *Annual Review of Public Health*, 30(1), 175–201. <https://doi.org/10.1146/annurev.publhealth.031308.100134>

- Celentano, D. D., Szklo, M., & Farag, Y. (2023). *Gordis epidemiology e-book: Gordis epidemiology e-book*. Elsevier health sciences. https://www.google.com/books?hl=id&lr=&id=07PxEAAAQBAJ&oi=fnd&pg=PP1&dq=Gordis+epidemiology+e-book:+Gordis+epidemiology+e-book&ots=LxYxOTd1n1&sig=vuwPjmuWj_vCYWQMr_AWHauo4OI
- Daniel, W. W., & Cross, C. L. (2018). *Biostatistics: A foundation for analysis in the health sciences*. John Wiley & Sons. <https://www.google.com/books?hl=id&lr=&id=PON1DwAAQBAJ&oi=fnd&pg=PR7&dq=Biostatistics:+A+Foundation+for+Analysis+in+the+Health+Sciences.&ots=a82rhRjjLq&sig=eSmpZl7bn20g22hqYkpqLroddM8>
- Ernawati, R., Susanto, J., Nuritasari, N. R. T., Kep, M., Kurdaningsih, N. S. V., Kep, M., Pratama, B. A., Putra, O., Lesmono, W. D., & Siregar, R. (2025). *Biostatistik Modern*. PT Bukuloka Literasi Bangsa. <https://www.google.com/books?hl=id&lr=&id=weCkEQAAQBAJ&oi=fnd&pg=PA1&dq=Biostatistik+Modern&ots=N3MajOWV3z&sig=ieS8FXyUqsxGyrgBdN2KEAn90A4>
- Friis, R. H., & Sellers, T. (2020). *Epidemiology for Public Health Practice*. Jones & Bartlett Learning.
- Goldstein, N. D., LeVasseur, M. T., & McClure, L. A. (2020). On the convergence of epidemiology, biostatistics, and data science. *Harvard Data Science Review*, 2(2), 10–1162.
- Gouda, H. N., & Powles, J. W. (2014). The science of epidemiology and the methods needed for public health assessments: A review of epidemiology textbooks. *BMC Public Health*, 14(1), 139. <https://doi.org/10.1186/1471-2458-14-139>

- Islamiyati, A. K., & Sadiman, A. K. (2025). *Biostatistik Kesehatan*. PT Bukuloka Literasi Bangsa.
https://books.google.com/books?hl=en&lr=&id=meCkEQAAQBAJ&oi=fnd&pg=PA1&dq=Biostatistik+Kesehatan&ots=WpUsgdHqkL&sig=WjbHE41V7rAluU_ILUNBqY3_t7o
- Karlinda, K., & Putri, S. I. (2025). *EPIDEMIOLOGI KESEHATAN*.
<https://media.neliti.com/media/publications/632137-epidemiologi-kesehatan-7bb6f9b4.pdf>
- Matranga, D., Bono, F., & Maniscalco, L. (2021). Statistical advances in epidemiology and public health. In *International journal of environmental research and public health* (Vol. 18, Issue 7, p. 3549). MDPI.
<https://www.mdpi.com/1660-4601/18/7/3549>
- Najmah, N. (2015). *Epidemiologi: Untuk Mahasiswa Kesehatan Masyarakat*. Jakarta: Raja Grafindo Persada.
- Pagano, M., Gauvreau, K., & Mattie, H. (2022). *Principles of biostatistics*. Chapman and Hall/CRC.
<https://api.taylorfrancis.com/content/books/mono/download?identifierName=doi&identifierValue=10.1201/9780429340512&type=googlepdf>
- Rosner, B. A. (2006). *Fundamentals of biostatistics* (Vol. 6). Thomson-Brooks/Cole Belmont, CA.
<https://www.academia.edu/download/62914619/Fundamentals-Of-Biostatistics-by-Bernard-Rosner.pdf>
- Rothman, K. J., Greenland, S., & Lash, T. L. (2008). *Modern epidemiology* (Vol. 3). Wolters Kluwer Health/Lippincott Williams & Wilkins Philadelphia.
[https://www.annemergmed.com/article/S0196-0644\(08\)01394-2/abstract](https://www.annemergmed.com/article/S0196-0644(08)01394-2/abstract)

Szklo, M., Nieto, F. J., & Epidemiology, F. (2014). Beyond the basics. *Ed2014*.
https://samples.jbpub.com/9781284116595/9781284116595_FMxx_Secure.pdf

BIODATA PENULIS



Ryryn Suryaman Prana Putra, SKM, M.Kes.
Dosen Program Studi Ilmu Kesehatan Masyarakat
Fakultas Kesehatan Masyarakat
Universitas Hasanuddin

Penulis lahir di Soppeng dimana pada tahun 2014 menyelesaikan Pendidikan Strata 2 di Prodi Kesehatan Masyarakat Departemen Administrasi dan Kebijakan Kesehatan Universitas Hasanuddin Makassar. Pada tahun 2011 sampai 2015 penulis menjadi asisten dosen, staf administrasi, dan pengelola jurnal artikel ilmiah di Fakultas Kesehatan Masyarakat Departemen Administrasi dan Kebijakan Kesehatan Universitas Hasanuddin Makassar. Sejak tahun 2015 sampai 2024 penulis bekerja sebagai Dosen Tetap pada Program Studi S-1 Administrasi Rumah Sakit Institut Ilmu Kesehatan Pelamonia Makassar dan pernah menjabat sebagai Sekertaris Institusi, Operator PDPT Kampus, Ketua LPPM, dan Ketua Bidang Penelitian pada Lembaga Penelitian dan Pengabdian kepada Masyarakat Institut Ilmu Kesehatan Pelamonia Kesdam XIV Hasanuddin Makassar. Saat ini penulis bekerja sebagai Dosen di Universitas Hasanuddin Fakultas Kesehatan Masyarakat Departemen Administrasi dan Kebijakan Kesehatan. Buku yang telah ditulis oleh penulis adalah Buku Monograf

"Pengaruh Penerapan Pelayanan Prima terhadap Kepuasan Pasien Rawat Jalan di RSUD Labuang Baji Makassar" (ISBN. 978-623-6957-55-4), Buku Monograf "Analisis GAP Harapan dan Kenyataan Pasien JKN terhadap Kualitas Pelayanan Rawat Jalan dengan Metode SERVQUAL dan Importance Performance Analysis di RSUD H. Padjonga Daeng Ngalle Kabupaten Takalar" (ISBN. 978-623-5856-22-3), Buku Manajemen Pemasaran (ISBN. 978-623-5981-69-7), Buku Manajemen Rumah Sakit (Teori Dan Aplikasi) (ISBN. 978-623-362-599-9), Buku Metodologi Penelitian Kuantitatif Dan Kualitatif (ISBN. 978-623-362-634-7), Buku Hukum Kesehatan dan Sumber Daya Manusia Kesehatan (ISBN. 978-623-362-757-3), Buku Hukum Kesehatan (ISBN. 978-623-8051-31-1), Buku Ekonomi Kesehatan (ISBN. 978-623-8102-16-7), Buku Gizi Kronis pada Anak Stunting (ISBN. 978-623-8102-02-0), Buku Penelitian Ilmu Kesehatan (ISBN. 978-623-8102-35-8), Buku Kebijakan Kesehatan (ISBN. 978-623-198-064-9), Buku Administrasi dan Kebijakan Kesehatan (ISBN. 978-623-198-099-1), Buku Manajemen dan Analisis Data (ISBN. 978-623-198-259-9), Buku Dasar-Dasar Ilmu Kesehatan Masyarakat (ISBN. 978-623-198-417-3), Buku Sistem Informasi Rumah Sakit (ISBN. 978-623-198-518-7), Buku Pentingnya Administrasi Rumah Sakit yang Tertata dengan Baik : Panduan Lengkap untuk Mengelola Rumah Sakit (ISBN. 978-623-198-518-7), Buku Manajemen Logistik Kesehatan (ISBN. 978-623-198-974-1), Buku Manajemen Bencana di Indonesia (ISBN. 978-623-508-622-4), Buku Mengelola Risiko : Lingkungan Kesehatan Publik di Tengah Bencana dan Urbanisasi (ISBN. 978-623-10-9849-8), Buku Metode Penelitian Kesehatan (ISBN. 978-623-508-660-6), Buku Kesehatan Masyarakat (Teori dan Penerapan) (ISBN. 978-623-512-614-2), Buku Kepemimpinan dalam Organisasi Era Teknologi (ISBN. 978-634-7380-98-2), Buku Manajemen Pemasaran Modern: Strategi, Taktik, dan Implementasi yang Sukses (ISBN. 978-634-255-453-1), dan Buku Metodologi Penelitian Terpadu (ISBN. 978-634-04-9749-6)

BIODATA PENULIS



Dr. RAHMAWATI, S. Si, M. Kes.

Dosen Tetap Program Studi D3 Teknologi Laboratorium
Medis (TLM)

POLITEKNIK MUHAMMADIYAH MAKASSAR

Penulis lahir di Desa Kasiwang Kecamatan Suli Kabupaten Luwu pada tanggal 12 April 1980, merupakan anak kedua dari dua bersaudara dari pasangan Abd. Muis dan Rapidah Hasan. Penulis adalah dosen tetap yayasan pada Program Studi D3 Teknologi Laboratorium Medis (TLM) Politeknik Muhammadiyah Makassar. Menyelesaikan pendidikan sekolah dasar di SD Negeri 20 Cimpu Kecamatan Suli Kabupaten Luwu pada tahun 1992. Pada tahun 1995 telah menyelesaikan pendidikan sekolah menengah pertama pada SMP Negeri 2 Kabupaten Takalar. Kemudian pada tahun 1998 menyelesaikan pendidikan sekolah menengah atas pada Sekolah Menengah Farmasi (SMF) Depkes Ujungpandang, Makassar.

Penulis menyelesaikan pendidikan S1 pada Jurusan Kimia Fakultas MIPA Universitas Hasanuddin pada tahun 2003. Pada tahun 2004 melanjutkan S2 pada Jurusan

Kesehatan Lingkungan Fakultas Kesehatan Masyarakat (FKM) Universitas Hasanuddin dan berhasil menyelesaikan studi pada tahun 2006. Kemudian pada tahun 2013 melanjutkan S3 pada jurusan Ilmu Kimia Fakultas MIPA Universitas Hasanuddin dan berhasil meraih gelar Doktor pada tahun 2018.

Sebagai dosen, penulis terus berkontribusi dan berperan aktif dalam penerapan tridarma perguruan tinggi. Penulis menekuni bidang ilmu kimia analitik, biokimia, pengantar laboratorium medis, toksikologi klinik, dan metodologi penelitian dan statistika. Penulis juga aktif dalam kegiatan penelitian dan pengabdian kepada masyarakat, khususnya dalam bidang kimia kesehatan dan menulis dengan tulisan telah terbit di beberapa buku, jurnal nasional dengan publikasi ilmiah yang dapat diakses melalui ID SINTA 6102341 dan *Google Scholar*.

Jabatan yang pernah diduduki oleh penulis sebagai Kepala Lembaga Pengembangan Pendidikan dan saat ini sebagai Sekertaris Lembaga Penjaminan Mutu dan Pengembangan Pendidikan (LPMPP) Politeknik Muhammadiyah Makassar.

BIODATA PENULIS



Bdn. Asruria Sani Fajriah, SST., MKM
Dosen Program Studi S1 Kebidanan
Fakultas Kebidanan
Universitas STRADA Indonesia

Penulis lahir di Kediri tanggal 29 Juni 1992. Penulis adalah dosen tetap pada Program Studi Studi S1 Kebidanan Fakultas Keperawatan dan Kebidanan, Universitas STRADA Indonesia. Menyelesaikan pendidikan D3 pada Jurusan Kebidanan, kemudian melanjutkan S1/D4 pada Jurusan Bidan Pendidik, dan melanjutkan pendidikan S2 pada Jurusan Ilmu Kesehatan Masyarakat. Penulis menekuni bidang kesehatan masyarakat dan riset. Penulis juga sebagai Chief Editor Journal for Quality in Public Health. Pernah mengisi sebagai pembicara dan moderator di beberapa seminar online kolaborasi berbagai institusi negeri dan swasta di Indonesia. Pernah menulis buku diantaranya (1) Dasar Biostatistik Untuk Peneliti; (2) Manajemen Kebidanan: Dari Kehamilan Hingga Nifas; (3) Promosi Kesehatan: Menuju Masyarakat Sehat Dan Aktif; (4) Pelayanan Kesehatan Reproduksi Lansia dalam Kebidanan Komunitas; dan (5) Pemberdayaan Perempuan dalam Kesehatan Maternal.

BIODATA PENULIS**Resti Ariani, M.Biomed**

Dosen Program Studi D3 Teknologi Bank Darah

Penulis lahir di Jepara, menyelesaikan pendidikan Sarjana (S1) Biologi di Universitas Negeri Semarang dan Magister (S2) Ilmu Biomedis di Universitas Diponegoro. Saat ini, menjabat sebagai dosen tetap pada Program Studi D3 Teknologi Bank Darah, Politeknik Bina Trada Semarang. Pada praktiknya sebagai dosen, sehari-hari mengampu sejumlah mata kuliah, antara lain Imunologi, Serologi, Komponen Darah, serta Infeksi Menular Lewat Transfusi Darah. Aktif dalam penelitian di bidang transfusi darah, darah, dan komponen darah, juga mencatatkan sejumlah kekayaan intelektual, baik paten maupun hak cipta nasional melalui tulisan dan karya lainnya seperti *Dasar-Dasar Biomedik, Kajian Dasar Imunologi Klinis dan Fitoterapi Imunomodulator, Dasar-Dasar Hematologi*.

BIODATA PENULIS



Suryani Mansyur, SKM., M.Kes

Dosen Program Studi Gizi Program Sarjana
Fakultas Ilmu Kesehatan, Universitas Muhammadiyah
Maluku Utara

Penulis lahir di Maros tanggal 10 Februari 1987. Penulis adalah dosen tetap pada Program Studi Gizi Program Sarjana Fakultas Ilmu Kesehatan, Universitas Muhammadiyah Maluku Utara. Menyelesaikan pendidikan S1 pada prodi kesehatan masyarakat peminatan gizi Fakultas Kesehatan Masyarakat (FKM), UNHAS (2008) dan melanjutkan S2 pada konsentrasi gizi pada kampus yang sama (2012). Penulis pernah menjabat sebagai Sekretaris Prodi Kesehatan Masyarakat Fakultas Ilmu Kesehatan, dan sekarang menjabat sebagai Ketua Biro Penerimaan Mahasiswa Baru Universitas Muhammadiyah Maluku Utara. Penulis juga aktif dalam kegiatan ilmiah dan organisasi keprofesian yaitu sebagai sekretaris Pimpinan Wilayah Imani Prokami Maluku Utara (Perhimpunan tenaga Kesehatan Profesi Kesehatan Muslim Indonesia) serta Sekretaris Pengda IAGIKMI Maluku Utara. Sehari-harinya bekerja sebagai dosen tetap pengampuh mata kuliah Biostatistika dasar, Ilmu Gizi dasar, Gizi Kesehatan Masyarakat, Manajemen KIE dan Program Gizi, serta Manajemen dan Analisis Data Dasar. Selain itu penulis juga aktif dalam menulis jurnal dan *book chapter*.

BIODATA PENULIS



Dr. dr. Muhammad Al Azhary, M.K.M.

Dosen Program Studi Keperawatan
Stikes Widya Dharma Husada Tangerang Selatan Kota

Penulis lahir di Balikpapan tanggal 18 Oktober 1981. Penulis adalah dosen tetap pada Program Studi Keperawatan Stikes Widya Dharma Husada Tangerang Selatan Kota. Menyelesaikan pendidikan Dokter Umum dan melanjutkan S2 pada Jurusan Kesehatan Masyarakat serta S3 pada Jurusan Manajemen.

BIODATA PENULIS



Nurpatwa Wilda Ningsi, SKM.,M.Kes

Dosen Program Studi Gizi
Fakultas Ilmu Kesehatan Universitas Sulawesi Barat

Penulis lahir di Ujung Pandang, pada 2 April 1995. Wilda menyelesaikan pendidikan S1 Universitas Muslim Indonesia konsentrasi Epidemiologi dan pendidikan magister dari Universitas Hasanuddin pada konsentrasi Epidemiologi. Wanita yang kerap disapa Wilda ini adalah anak dari pasangan Drs. Adam Malik (ayah) dan Nurmaida (ibu). Nurpatwa Wilda Ningsi berasal dari keluarga yang bergelut di bidang pendidikan, Bapak dan Ibunya aktif sebagai pengajar di Institusi Pendidikan dibawah naungan dinas pendidikan. Orang tuanya menjadi role model dalam menggapai cita-cita dan tanggungjawabnya sebagai dosen. Adapun riwayat pekerjaannya; tahun 2020 hingga sekarang wilda bekerja di Fakultas Ilmu Kesehatan Universitas Sulawesi Barat sebagai dosen Program Studi Gizi.

BIODATA PENULIS



Apt. Nurul Hidayah, S.Farm., M.Si.

Dosen Program Studi Diploma Tiga Farmasi
Jurusan Farmasi, Poltekkes Kemenkes Medan

Nurul Hidayah lahir di Medan pada 16 Oktober 1989. Ia menyelesaikan pendidikan Sarjana di Fakultas Farmasi, Universitas Sumatera Utara (USU) pada tahun 2012, kemudian melanjutkan pendidikan Magister di Sekolah Farmasi, Institut Teknologi Bandung (ITB) dan lulus pada tahun 2015. Ia memulai karier sebagai dosen di Fakultas Farmasi Universitas Tjut Nyak Dhien Medan pada tahun 2015. Sejak tahun 2018 hingga saat ini, ia menjadi dosen tetap di Jurusan Farmasi Poltekkes Kemenkes Medan.

BIODATA PENULIS



Dwi Yulia Maritasari, SKM., M.KM

Dosen Program Studi Kesehatan Masyarakat
Fakultas Kesehatan Universitas Mitra Indonesia

Penulis lahir di Bandar Lampung, 14 Juli 1990. Sebelumnya mengikuti Pendidikan Program Sarjana Ilmu Kesehatan Masyarakat (S1) di Universitas Sriwijaya, Indralaya dan melanjutkan Program Magister Ilmu Kesehatan Masyarakat (S2) Peminatan Gizi Masyarakat di Universitas Indonesia, Depok. Penulis merupakan dosen tetap di Program Studi Kesehatan Masyarakat Universitas Mitra Indonesia sejak tahun 2016 hingga sekarang.

Jabatan yang pernah diduduki oleh penulis selama berkarir antara lain sebagai Ketua UPT Penelitian dan Pengabdian kepada Masyarakat (PPM) Fakultas Kesehatan, dan Sekretaris Program Studi Magister Kesehatan Masyarakat Universitas Mitra Indonesia. Sejak tahun 2016 hingga saat ini, penulis mengampu mata kuliah biostatistika, dasar ilmu gizi masyarakat, epidemiologi gizi, epidemiologi penyakit tropis, serta metodologi penelitian pada program studi jenjang S1. Berbagai karya tulis ilmiah dan penelitian

telah banyak dilakukan oleh penulis dibidang kesehatan masyarakat, dan penulis juga aktif mengikuti kegiatan seminar dan pelatihan dalam rangka meningkatkan kompetensi di bidang ilmu kesehatan masyarakat.