



selat media



Dasar-Dasar Biostatistik

untuk Kesehatan Masyarakat

Editor :
Bangkit Ary Pratama, SKM., M.Kes.

DASAR - DASAR BIOSTATISTIK UNTUK MASYARAKAT

Bangkit Ary Pratama, SKM., M.Kes., dkk.

Editor: Bangkit Ary Pratama, SKM., M.Kes.

DASAR - DASAR BIOSTATISTIK UNTUK MASYARAKAT

Penulis : Bangkit Ary Pratama, SKM., M.Kes, Hardian Bimanto, S.KM., M.Kes, Desie Rahmawati, S.KM., M.Kes, Shiva Aflahiyah, S.Tr.Keb., M.K.M, Fitri Eka Sari S, M.Kes, Zakiyah, SST., M.K.M, Tri Budiarti, S.K.M, M.K.M, Agus Erwin Ashari, SKM, M.Kes, Intan Pujilestari, SST., M.K.M, Izal Puji Santoso, SKM., M.Kes

ISBN : 9786347465771
Penyunting Naskah : Selat Media Patners
Tata Letak : Selat Media Patners
Desain Sampul : Novikean Keysah Sanisri

Tebal : x + 197 halaman
Ukuran : 16 x 23 cm

Cetakan 1

Yogyakarta, Juni 2026

© Hak cipta dilindungi oleh undang-undang

Berlaku selama 50 (lima puluh) tahun sejak ciptaan tersebut pertama kali dilakukan pengumuman. Dilarang mengutip atau memperbanyak sebagian atau seluruh isi buku ini tanpa izin tertulis dari penerbit. Ketentuan Pidana Sanksi Pelanggaran Pasal 2 UU Nomor 19 Tahun 2002 Tentang Hak Cipta. Barang siapa dengan sengaja dan tanpa hak melakukan perbuatan sebagaimana dimaksud dalam Pasal 2 ayat (1) atau Pasal 49 ayat (1) dan ayat (2) dipidana dengan pidana penjara masing-masing paling singkat 1 (satu) bulan dan/atau denda paling sedikit Rp1.000.000,00 (satu juta rupiah), atau pidana penjara paling lama 7 (Tujuh) tahun dan/atau denda paling banyak Rp5.000.000.000,00 (lima miliar rupiah).

Barang siapa dengan sengaja menyerahkan, menyiarkan, memamerkan, mengedarkan atau menjual kepada umum suatu ciptaan atau barang hasil pelanggaran Hak Cipta atau Hak Terkait sebagaimana dimaksud pada ayat (1) dipidana dengan pidana penjara paling lama 5 (lima) tahun dan/atau denda paling banyak Rp500.000.000,00 (lima ratus juta rupiah).

Diterbitkan oleh:

Penerbit Selat Media Patners

Anggota IKAPI No. 165/DIY/2022

Kertopaten RT.02 Wirokerten, Banguntapan Bantul Yogyakarta

Mitra Penerbit: **Penerbit PT Bukuloka Literasi Bangsa**

Kompleks Business Park Kebon Jeruk Blok I No. 21, Jl. Meruya Ilir Raya No. 88, Meruya Utara, Kec. Kembangan, Kota Jakarta Barat, DKI Jakarta 11620

KATA PENGANTAR

Puji syukur kami panjatkan ke hadirat Tuhan Yang Maha Esa atas terselesaikannya buku Dasar-Dasar Biostatistik untuk Masyarakat. Buku ini disusun sebagai buku ajar yang ditujukan bagi mahasiswa, tenaga kesehatan, serta pembaca umum yang ingin memahami penerapan biostatistik dalam konteks kesehatan masyarakat secara praktis dan komprehensif.

Biostatistik merupakan salah satu pilar penting dalam pengambilan keputusan berbasis bukti di bidang kesehatan. Namun, tidak sedikit mahasiswa maupun praktisi yang masih menganggap biostatistik sebagai disiplin yang sulit dijangkau. Buku ini hadir untuk menjembatani kesenjangan tersebut dengan menyajikan konsep-konsep biostatistik secara sistematis, mulai dari pengertian dasar, jenis dan sumber data kesehatan, pengolahan serta penyajian data, ukuran pemusatan dan penyebaran, distribusi data, hingga pemanfaatan biostatistik dalam pemantauan dan evaluasi program kesehatan. Pembahasan mengenai etika dan validitas data serta pemanfaatan teknologi digital dalam analisis biostatistik turut melengkapi cakupan buku ini agar sesuai dengan kebutuhan era kesehatan modern.

Penyusunan materi mengacu pada literatur ilmiah terkini sehingga diharapkan dapat menjadi referensi yang relevan dan aplikatif bagi kepentingan akademik maupun praktik di lapangan.

Kami menyadari bahwa buku ini masih memiliki keterbatasan. Oleh karena itu, kritik dan saran yang membangun dari pembaca sangat kami harapkan demi perbaikan pada edisi selanjutnya. Semoga buku ini dapat memberikan manfaat yang sebesar-besarnya bagi pengembangan ilmu kesehatan masyarakat di Indonesia.

Yogyakarta, Juni 2026

Tim Penulis

DAFTAR ISI

KATA PENGANTAR..... V

BAB 1 KONSEP DASAR BIOSTATISTIK DALAM KESEHATAN MASYARAKAT..... 1

1.1 PENGERTIAN BIOSTATISTIK 1

1.2 PERAN BIOSTATISTIK DALAM KESEHATAN MASYARAKAT 4

1.3 KONSEP DASAR STATISTIK YANG DIGUNAKAN DALAM BIOSTATISTIK ... 7

1.4 TEKNIK PENGUMPULAN DATA DALAM BIOSTATISTIK 11

DAFTAR PUSTAKA 15

BAB 2 JENIS DAN SUMBER DATA KESEHATAN 17

2.1 DEFINISI DAN KONSEP DATA KESEHATAN..... 17

2.2 JENIS-JENIS DATA KESEHATAN 20

2.3 SUMBER DATA KESEHATAN 22

2.4 TEKNOLOGI DAN SISTEM INFORMASI UNTUK MANAJEMEN DATA KESEHATAN..... 24

2.5 TANTANGAN DAN PELUANG DALAM PENGGUNAAN DATA KESEHATAN26

DAFTAR PUSTAKA 30

BAB 3 PENGOLAHAN DAN PENYAJIAN DATA STATISTIK 32

3.1 KONSEP DASAR PENGOLAHAN DATA 32

3.2 PENYAJIAN DATA STATISTIK 37

3.3 ANALISIS DESKRIPTIF DAN INFERENSIAL 41

3.4 PRAKTIK TERBAIK DALAM PENGOLAHAN DATA..... 44

DAFTAR PUSTAKA 49

BAB 4 UKURAN PEMUSATAN DAN PENYEBARAN DATA

..... 50

4.1 KONSEP DASAR UKURAN PEMUSATAN DATA 51

4.2 JENIS-JENIS UKURAN PEMUSATAN 52

4.3 KONSEP UKURAN PENYEBARAN DATA 62

4.4 JENIS-JENIS UKURAN PENYEBARAN 63

DAFTAR PUSTAKA 74

BAB 5 DISTRIBUSI DATA DAN INTERPRETASI HASIL ... 75

5.1 KONSEP DISTRIBUSI DATA 75

5.2 ANALISIS DESKRIPTIF DATA 81

5.3 ANALISIS INFERENSIAL DATA 85

5.4 INTERPRETASI HASIL STATISTIK 90

DAFTAR PUSTAKA 96

BAB 6 ANALISIS HUBUNGAN ANTAR VARIABEL DALAM KESEHATAN 97

6.1 KONSEP DASAR ANALISIS HUBUNGAN VARIABEL KESEHATAN..... 97

6.2 DESAIN PENELITIAN UNTUK MENGANALISIS HUBUNGAN VARIABEL 100

6.3 MENGIDENTIFIKASI DAN MENGENDALIKAN BIAS DALAM ANALISIS
HUBUNGAN VARIABEL..... 103

6.4 METODE STATISTIK UNTUK MENGANALISIS HUBUNGAN VARIABEL . 107

6.5 APLIKASI PRAKTIS DALAM PENELITIAN KESEHATAN POPULASI 110

DAFTAR PUSTAKA 115

BAB 7 PENGGUNAAN BIOSTATISTIK DALAM PEMANTAUAN KESEHATAN MASYARAKAT 116

7.1 PENGERTIAN BIOSTATISTIK DALAM KESEHATAN MASYARAKAT 116

7.2 SURVEILANS KESEHATAN MASYARAKAT 119

7.3 ANALISIS DATA EPIDEMIOLOGI 122

7.4 APLIKASI SISTEM INFORMASI DALAM PEMANTAUAN KESEHATAN	125
7.5 IMPLEMENTASI PROGRAM KESEHATAN BERBASIS BIOSTATISTIK.....	128
DAFTAR PUSTAKA	134
BAB 8 APLIKASI BIOSTATISTIK DALAM EVALUASI PROGRAM KESEHATAN	135
8.1 PENGANTAR BIOSTATISTIK	135
8.2 DESAIN DAN PENGUMPULAN DATA	138
8.3 ANALISIS STATISTIK	141
8.4 INTERPRETASI HASIL DAN EVALUASI PROGRAM.....	143
DAFTAR PUSTAKA	149
BAB 9 ETIKA DAN VALIDITAS DATA DALAM BIOSTATISTIK	151
9.1 FONDASI ETIKA DALAM PENELITIAN BIOSTATISTIK.....	151
9.2 VALIDITAS DATA DALAM KONTEKS BIOSTATISTIK	154
9.3 RELIABILITAS DATA DAN KONSISTENSI PENGUKURAN	157
9.4 MANAJEMEN INTEGRITAS DATA DAN PELAPORAN ETIS.....	159
9.5 ISU ETIKA KONTEMPORER DALAM BIOSTATISTIK DAN PENELITIAN KESEHATAN.....	161
DAFTAR PUSTAKA	167
BAB 10 PEMANFAATAN TEKNOLOGI DIGITAL DALAM ANALISIS BIOSTATISTIK KESEHATAN	169
10.1 PENGENALAN TEKNOLOGI DIGITAL DALAM ANALISIS BIOSTATISTIK KESEHATAN.....	169
10.2 SOFTWARE DAN PLATFORM ANALISIS DATA BIOSTATISTIK.....	173
10.3 ARTIFICIAL INTELLIGENCE DAN ADVANCED ANALYTICS.....	177
10.4 IMPLEMENTASI PRAKTIS DAN STUDI KASUS	181
DAFTAR PUSTAKA	187

PROFIL PENULIS188

BAB 1

KONSEP DASAR

BIOSTATISTIK DALAM

KESEHATAN MASYARAKAT

Bangkit Ary Pratama, SKM.,M.Kes.

1.1 Pengertian Biostatistik

Biostatistik merupakan cabang ilmu yang berfokus pada penerapan metode statistik untuk memahami, menganalisis, dan menginterpretasikan data yang berkaitan dengan kesehatan masyarakat, epidemiologi, serta penelitian klinis. Disiplin ini menempati posisi strategis di antara statistik dan ilmu kesehatan, yang berfungsi menyediakan kerangka kuantitatif bagi peneliti dan pengambil kebijakan guna menghasilkan keputusan akurat berbasis bukti. Kegiatan biostatistik mencakup pengumpulan, pengolahan, perhitungan nilai statistik, hingga pengujian hipotesis untuk menilai hubungan antarvariabel kesehatan. Informasi mentah dapat diolah menjadi temuan bermakna melalui pendekatan ini, seperti prediksi tren penyakit, identifikasi faktor risiko, dan evaluasi efektivitas intervensi. Data kunjungan pasien ISPA di puskesmas, sebagai

ilustrasi, dianalisis untuk memprediksi lonjakan kasus pada musim pancaroba sehingga persediaan obat dapat disiapkan lebih awal. Penjaminan mutu penelitian kesehatan juga sangat bergantung pada biostatistik, khususnya dalam penetapan desain studi yang valid dan pengendalian bias. Proses ilmiah yang mendasari pengembangan ilmu kesehatan masyarakat secara keseluruhan bertumpu pada biostatistik, bukan hanya pada tahap analisis data.

Evaluasi program dan intervensi kesehatan secara sistematis dimungkinkan berkat landasan objektif yang disediakan biostatistik. Peta distribusi penyakit berdasarkan usia atau lingkungan dapat disusun menggunakan data insidensi dan prevalensi, sehingga intervensi dapat diarahkan pada kelompok prioritas. Hubungan kausal antara faktor risiko dan kejadian penyakit dinilai menggunakan analisis regresi atau model survival analysis. Kebijakan kesehatan yang diambil melalui pendekatan berbasis bukti ini dipastikan memiliki dukungan data kuantitatif yang valid. Riset klinis memanfaatkan biostatistik untuk mengevaluasi keamanan serta efektivitas intervensi medis melalui rancangan uji klinis yang ketat. Kesimpulan yang dihasilkan dari pendekatan tersebut memiliki kekuatan ilmiah memadai untuk pengambilan keputusan. Literatur ilmiah menegaskan bahwa biostatistik merupakan bagian integral dari metodologi penelitian modern karena menyediakan alat untuk memastikan temuan penelitian dapat digeneralisasi (Gerstman, 2015).

Teknologi informasi memperluas peran biostatistik melalui teknik analisis lanjutan seperti *big data analytics* dan *machine learning*. Identifikasi pola yang sebelumnya tidak terlihat kini dimungkinkan berkat kemampuan pengolahan data dalam jumlah besar. Rekam medis elektronik jutaan pasien, contohnya, diolah menggunakan algoritma machine learning untuk mendeteksi pola resistensi antibiotik di berbagai wilayah. Prinsip dasar biostatistik tetap menjadi fondasi penting dalam memastikan integritas analisis, terlepas dari berbagai inovasi tersebut. Pemahaman mengenai konsep ini sangat diperlukan bagi tenaga kesehatan dan peneliti agar mampu menilai kualitas bukti ilmiah serta merancang program yang efektif. Biostatistik merupakan disiplin strategis dalam pembangunan kesehatan masyarakat yang bertumpu pada data dan bukti ilmiah (Rosner, 2015). Pemanfaatan biostatistik tidak lagi sekadar dipandang sebagai alat analisis teknis, melainkan telah menjadi disiplin strategis dalam pembangunan kesehatan masyarakat yang bertumpu pada data dan bukti ilmiah. Integrasi metode kuantitatif yang tepat memastikan bahwa setiap kebijakan kesehatan yang diambil memiliki tingkat akurasi tinggi dan efisiensi yang terukur. Kemampuan mengolah data mentah menjadi informasi kebijakan merupakan kunci utama dalam menjawab tantangan kesehatan global yang semakin kompleks.

1.2 Peran Biostatistik dalam Kesehatan Masyarakat

Ilmu biostatistik memiliki peran strategis dalam menunjang praktik dan kebijakan kesehatan masyarakat. Pendekatan kuantitatif memungkinkan analisis mendalam terhadap distribusi penyakit dan efektivitas intervensi. Pemanfaatan data statistik yang akurat membantu praktisi kesehatan mengambil keputusan berbasis bukti serta memprediksi tren masalah kesehatan secara presisi (Kusnanto & Dika, 2021).

1.2.1 Analisis Distribusi Penyakit

Identifikasi pola distribusi penyakit melalui pemanfaatan data epidemiologi menjadi kontribusi utama biostatistik dalam sistem kesehatan. Epidemiolog menerapkan berbagai metode kuantitatif, termasuk incidence rate dan prevalence, guna memberikan gambaran komprehensif mengenai persebaran penyakit di tengah populasi. Hasil analisis ini menjadi basis fundamental dalam menetapkan prioritas program serta melakukan evaluasi terhadap keberhasilan intervensi yang telah dijalankan. Deteksi dini terhadap tren peningkatan penyakit tidak menular, seperti hipertensi, dapat dilakukan secara lebih akurat melalui analisis data longitudinal yang konsisten. Pemetaan tingkat prevalensi diabetes melitus di tingkat provinsi yang diuraikan per kabupaten/kota membantu pemerintah dalam menentukan alokasi pendanaan program pencegahan secara proporsional. Strategi serupa diterapkan pada penanganan penyakit menular di wilayah tertentu. Distribusi logistik berupa kelambu dan obat-obatan di daerah endemis malaria

difokuskan pada titik koordinat dengan densitas kasus tertinggi berdasarkan data biostatistik (Setiawan & Hermawan, 2022). Pendekatan berbasis spasial ini memastikan bahwa sumber daya kesehatan yang terbatas dapat dimanfaatkan secara optimal untuk menurunkan beban penyakit di masyarakat.

1.2.2 Identifikasi Faktor Risiko

Identifikasi faktor risiko yang memengaruhi derajat kesehatan masyarakat merupakan fungsi vital dari implementasi biostatistik. Peneliti menggunakan parameter kuantitatif seperti *odds ratio* dan *relative risk* guna mengukur kekuatan asosiasi antara variabel lingkungan atau perilaku terhadap peluang terjadinya penyakit. Penentuan rancangan intervensi yang spesifik serta efektivitas pemantauan pada kelompok rentan sangat bertumpu pada hasil pengukuran tersebut. Besaran risiko kanker paru pada kelompok perokok aktif dibandingkan dengan kelompok non-perokok, sebagai ilustrasi, dihitung menggunakan *nilai relative risk* sebagai basis data pendukung untuk memperkuat regulasi kawasan tanpa rokok (Setiawan & Hermawan, 2022). Langkah tersebut memastikan kebijakan publik memiliki landasan ilmiah yang kuat dalam melindungi kesehatan populasi. Penerapan biostatistik dalam mengidentifikasi determinan kesehatan juga terlihat pada analisis masalah gizi kronis. Penghitungan risiko kejadian stunting pada balita berdasarkan variabel kualitas air bersih dan tingkat pendapatan keluarga di kawasan pemukiman padat menjadi bukti nyata pemanfaatan statistik untuk menemukan solusi kesehatan

masyarakat (Khoirunnisa & Dewi, 2022). Temuan ini memungkinkan pemerintah merancang program intervensi gizi yang lebih terarah dan efisien sesuai dengan kondisi sosiokultural wilayah setempat.

1.2.3 Evaluasi Intervensi dan Kebijakan Kesehatan

Sistematikasi pengukuran efektivitas program kesehatan menjadi keharusan guna memetakan dampak nyata terhadap populasi sasaran. Evaluasi kuantitatif pada berbagai inisiatif, mulai dari program berbasis komunitas hingga kampanye promosi kesehatan, dapat terlaksana secara akurat berkat dukungan biostatistik. Metode ilmiah seperti *randomized controlled trials* (RCT) dan *cost-effectiveness analysis* digunakan untuk memperoleh gambaran objektif mengenai keberhasilan suatu program (Ramadani et al., 2023).

Analisis terhadap efektivitas program pemberian tablet tambah darah pada kelompok remaja putri dalam upaya menurunkan prevalensi anemia, sebagai ilustrasi, dilakukan guna menentukan keberlanjutan kebijakan tersebut (Kusnanto & Dika, 2021). Langkah ini memastikan bahwa sumber daya negara dialokasikan pada program yang memiliki bukti efikasi klinis dan sosiologis yang nyata. Program intervensi gizi lainnya, seperti pemberian makanan tambahan pada balita dengan status gizi kurang, dievaluasi tingkat keberhasilannya melalui perbandingan kenaikan berat badan sebelum dan sesudah intervensi menggunakan uji statistik yang sesuai. Hasil evaluasi kuantitatif ini menjadi dasar utama dalam

merumuskan rekomendasi kebijakan kesehatan publik yang berkelanjutan dan berbasis bukti (Sari & Putra, 2021). Penggunaan data statistik menjamin bahwa setiap kebijakan kesehatan tidak hanya bersifat reaktif, tetapi juga memiliki akuntabilitas ilmiah yang tinggi.

Peranannya yang komprehensif memposisikan biostatistik tidak sekadar sebagai alat bantu teknis, melainkan fondasi utama dalam pengambilan keputusan yang rasional dan efisien di sektor kesehatan masyarakat. Penguasaan biostatistik merupakan keterampilan wajib bagi profesional kesehatan masyarakat di tengah era kebijakan berbasis data saat ini guna menjamin akuntabilitas dan efektivitas program kesehatan.

1.3 Konsep Dasar Statistik yang Digunakan dalam Biostatistik

1.3.1 Distribusi dan Probabilitas

Pemahaman pola kejadian dalam populasi bertumpu pada distribusi data yang menjadi fondasi krusial dalam kajian biostatistik. Acuan utama sering kali menggunakan distribusi normal mengingat karakteristiknya yang sesuai dengan pola berbagai variabel biologis manusia. Peneliti mampu mengenali keberadaan gejala ekstrem serta variabilitas dalam kelompok melalui pemahaman mendalam terhadap penyebaran data tersebut.

Konsep *probability* atau probabilitas digunakan sebagai alat ukur untuk menentukan peluang terjadinya suatu peristiwa kesehatan. Estimasi mengenai risiko seseorang terkena penyakit

berdasarkan faktor risiko spesifik dihitung menggunakan pendekatan probabilitas dalam konteks medis. Peluang seorang perokok berat mengalami penyakit jantung koroner dalam kurun waktu sepuluh tahun ke depan, sebagai contoh, diestimasi secara kuantitatif menggunakan skor risiko kardiovaskular (Nugroho & Hartati, 2022).

Kombinasi antara pemahaman distribusi data dan teori probabilitas menjadi kebutuhan mutlak dalam perencanaan intervensi kesehatan yang berbasis bukti. Langkah ini memastikan bahwa setiap tindakan preventif maupun kuratif didasarkan pada perhitungan risiko yang presisi dan valid secara statistik. Penerapan prinsip-prinsip ini membantu praktisi kesehatan masyarakat dalam mengalokasikan sumber daya secara lebih efektif pada kelompok populasi yang memiliki tingkat risiko tertinggi.

1.3.2 Inferensi Statistik dan Pengujian Hipotesis

Penarikan kesimpulan umum yang bersumber dari data sampel terbatas merupakan tujuan utama dalam implementasi inferensi statistik. Relevansi proses ini dinilai sangat tinggi mengingat keterbatasan penelitian kesehatan yang sering kali tidak mampu menjangkau keseluruhan populasi secara total. Keputusan ilmiah yang berpijak pada data dihasilkan melalui serangkaian teknik inferensi, mencakup estimasi parameter serta uji signifikansi.

Hypothesis testing atau pengujian hipotesis menjadi salah satu metode primer dalam kerangka kerja inferensi tersebut. Langkah-langkah yang melibatkan perumusan hipotesis serta

penggunaan uji statistik dilakukan secara ketat guna menilai hubungan atau kaitan antarvariabel kesehatan. Dugaan mengenai pengaruh olahraga rutin terhadap penurunan kadar kolesterol darah, sebagai ilustrasi, dibuktikan validitasnya melalui aplikasi uji statistik t-test pada kelompok sampel yang representatif (Sari & Putra, 2021). (Sari & Putra, 2021).

Validitas akhir dari setiap keputusan ilmiah sangat bergantung pada tingkat signifikansi atau *p-value* serta kekuatan uji statistik (*statistical power*) yang diterapkan dalam analisis. Penjagaan terhadap standar-standar ini memastikan bahwa kesimpulan riset kesehatan masyarakat dapat dipertanggungjawabkan secara metodologis dan dapat dijadikan rujukan dalam penyusunan intervensi medis maupun kebijakan publik. Pemanfaatan inferensi statistik yang tepat meminimalkan risiko penarikan kesimpulan yang salah akibat faktor kebetulan dalam pengambilan sampel.

1.3.3 Analisis Korelasi dan Regresi

Teknik analitik korelasi dan regresi digunakan secara luas guna mengevaluasi hubungan antara dua atau lebih variabel kesehatan. Parameter korelasi berfungsi untuk mengukur kekuatan serta arah hubungan antarvariabel, sementara pemodelan hubungan guna kepentingan estimasi atau prediksi masa depan dilakukan melalui analisis regresi. Evaluasi mengenai kaitan antara berbagai faktor risiko terhadap kondisi kesehatan tertentu dalam ranah biostatistik sangat bertumpu pada penggunaan kedua metode ini.

Prediksi mengenai probabilitas kejadian diabetes berdasarkan variabel faktor risiko individu dapat dilakukan melalui penerapan model regresi logistik. Pendekatan ini memungkinkan praktisi kesehatan masyarakat untuk mengidentifikasi kelompok populasi yang memiliki risiko tertinggi sehingga langkah preventif dapat diambil sedini mungkin. Kaitan antara variabel fisik juga dapat dimodelkan secara matematis untuk melihat pengaruh satu variabel terhadap variabel lainnya. Pengaruh peningkatan Indeks Massa Tubuh (IMT) terhadap kenaikan tekanan darah sistolik, sebagai contoh, dianalisis menggunakan model regresi linear (Utami & Ramadhan, 2023). Pemilihan model regresi yang tepat sangat krusial agar hasil analisis dapat diterjemahkan menjadi strategi intervensi yang efektif dan tepat sasaran. Akurasi dalam pemodelan statistik ini menjamin bahwa program kesehatan yang dirancang memiliki landasan ilmiah kuat dan dapat dipertanggungjawabkan hasilnya bagi peningkatan derajat kesehatan masyarakat.

Penguasaan terhadap berbagai konsep statistik tersebut memungkinkan praktisi kesehatan serta peneliti memformulasikan keputusan yang lebih akurat, meningkatkan derajat validitas studi, dan memperkuat sokongan terhadap kebijakan kesehatan berbasis bukti. Kemampuan mengaplikasikan biostatistik merupakan kompetensi esensial bagi profesional di bidang kesehatan masyarakat, epidemiologi, serta riset klinis guna memastikan setiap program kesehatan memiliki landasan ilmiah yang kuat.

1.4 Teknik Pengumpulan Data dalam Biostatistik

1.4.1 Survei sebagai Metode Pengumpulan Data

Teknik survei digunakan sebagai metode utama untuk memperoleh data kuantitatif dari populasi tertentu. Informasi mengenai karakteristik kesehatan dan faktor risiko dikumpulkan peneliti melalui kuesioner terstruktur. Pengumpulan data dalam jumlah besar dapat dilakukan dengan waktu singkat dan biaya efisien menggunakan survei. Survei Status Gizi Indonesia (SSGI), misalnya, dilaksanakan secara nasional untuk memotret angka prevalensi stunting pada balita serta menyediakan data dasar bagi kementerian kesehatan untuk mengevaluasi capaian target nasional. Gambaran komprehensif tentang tren kesehatan masyarakat yang dihasilkan metode ini berguna untuk analisis epidemiologis dan perencanaan kebijakan (Muntaner et al., 2020; Khoirunnisa & Dewi, 2022).

1.4.2 Wawancara dan Penggalan Informasi Mendalam

Perolehan data kualitatif yang mendalam dilakukan melalui teknik wawancara, terutama untuk memahami dimensi faktor sosial serta perilaku kesehatan. Implementasi teknik ini dapat dirancang secara terstruktur maupun tidak terstruktur guna menyesuaikan dengan kebutuhan spesifik riset. Interaksi langsung antara peneliti dan narasumber memungkinkan proses penggalan makna dari setiap jawaban secara mendetail dan komprehensif. Faktor sosial yang menghambat partisipasi kelompok pria dalam program Keluarga Berencana (KB), sebagai ilustrasi, digali melalui metode wawancara

mendalam guna merumuskan strategi komunikasi yang lebih tepat (Suryani & Wibowo, 2021). Pendekatan ini memastikan bahwa kebijakan kesehatan tidak hanya berpijak pada angka, tetapi juga mempertimbangkan nilai-nilai sosiokultural masyarakat. Penelitian biostatistik yang memfokuskan kajian pada determinan sosial sangat terbantu oleh integrasi metode ini dalam memperkuat analisis data kuantitatif. Penjelasan yang kaya dari hasil wawancara memberikan konteks yang diperlukan untuk memahami pola-pola statistik yang ditemukan di lapangan, sehingga solusi kesehatan masyarakat yang dihasilkan menjadi lebih holistik.

1.4.3 Eksperimen dalam Penelitian Kesehatan

Pembuktian hubungan sebab akibat antarvariabel kesehatan dilakukan melalui penerapan teknik pengumpulan data eksperimen secara ketat. Peneliti melakukan manipulasi pada variabel tertentu guna mengamati efeknya terhadap hasil kesehatan melalui rancangan studi seperti *randomized controlled trial* (RCT). Kesimpulan kausal yang kuat dapat dihasilkan berkat perolehan data objektif yang bersumber dari prosedur eksperimental tersebut.

Efikasi vaksin baru dalam mencegah infeksi virus tertentu, sebagai ilustrasi, diuji dengan membandingkan kelompok penerima vaksin dan kelompok plasebo secara acak (Ramadani et al., 2023; Muntaner et al., 2020). Prosedur ini memastikan bahwa perbedaan hasil kesehatan yang ditemukan benar-benar bersumber dari intervensi yang diberikan, bukan karena faktor kebetulan atau bias eksternal. Metode eksperimen ini menjadi standar utama dalam

penilaian efektivitas intervensi kesehatan masyarakat maupun berbagai tahapan uji klinis. Implementasi rancangan eksperimental yang tepat memungkinkan otoritas kesehatan untuk mengadopsi teknologi medis atau kebijakan pencegahan penyakit dengan tingkat kepercayaan ilmiah yang tinggi. Keputusan untuk mengimplementasikan suatu pengobatan massal atau program vaksinasi nasional selalu didasarkan pada hasil evaluasi empiris yang diperoleh melalui metode ini.

1.4.4 Observasi dalam Analisis Pola Kesehatan

Pencatatan perilaku atau fenomena kesehatan tanpa melibatkan intervensi langsung dilakukan melalui penerapan teknik observasi. Metode ini sering dimanfaatkan secara luas dalam studi epidemiologi serta penelitian terkait perilaku kesehatan masyarakat. Kondisi riil di lapangan dapat ditangkap secara akurat oleh peneliti guna menghasilkan data yang relevan bagi analisis pola kesehatan populasi.

Pengamatan terhadap kepatuhan pedagang kaki lima dalam menjaga kebersihan alat makan, sebagai ilustrasi, merupakan aplikasi metode observasi guna menilai risiko penyakit tular makanan (Khoirunnisa & Dewi, 2022). Langkah ini memungkinkan identifikasi bahaya kesehatan lingkungan secara langsung tanpa bergantung pada laporan subjektif dari responden. Kepatuhan tenaga medis dalam menerapkan prosedur cuci tangan lima momen di ruang ICU dicatat secara sistematis menggunakan lembar pemeriksaan (*checklist*) tanpa sepengetahuan subjek guna menghindari bias

perilaku. Teknik ini menempati posisi yang sangat penting karena mampu mengukur perilaku yang sulit diperoleh secara akurat melalui instrumen survei atau wawancara (Suryani & Wibowo, 2021). Hasil observasi memberikan data objektif yang krusial dalam upaya peningkatan mutu layanan kesehatan serta pengendalian infeksi di fasilitas medis.

Latihan Soal

1. Apa yang dimaksud dengan biostatistik dalam bidang kesehatan masyarakat?
2. Bagaimana biostatistik digunakan untuk mengidentifikasi faktor risiko penyakit?
3. Apa saja teknik pengumpulan data yang sering digunakan dalam penelitian kesehatan?
4. Mengapa pengujian hipotesis penting dalam analisis data biostatistik?
5. Mengapa evaluasi program kesehatan membutuhkan analisis biostatistik?

DAFTAR PUSTAKA

- Gerstman, B. B. (2015). *Basic biostatistics: Statistics for public health practice* (2nd ed.). Jones & Bartlett Learning.
- Khoirunnisa, D., & Dewi, R. (2022). Teknik pengumpulan data dalam penelitian kesehatan masyarakat. *Jurnal Kesehatan Masyarakat*, 14(2), 85–94.
- Kusnanto, H., & Dika, A. M. (2021). Peran biostatistik dalam penentuan kebijakan kesehatan berbasis data. *Jurnal Kesehatan Masyarakat Indonesia*, 16(2), 88–94.
- Muntaner, C., et al. (2020). Data collection methods in epidemiological research: An overview. *Journal of Public Health Research*, 9(3), 233–241.
- Nugroho, A., & Hartati, S. (2022). Dasar-dasar biostatistik dalam analisis kesehatan masyarakat. *Jurnal Statistika dan Kesehatan*, 9(1), 15–25.
- Ramadani, F., et al. (2023). Experimental methods for evaluating health interventions. *Indonesia Journal of Health Research*, 7(1), 22–30.
- Rosner, B. (2015). *Fundamentals of biostatistics* (8th ed.). Cengage Learning.
- Sari, M. A., & Putra, R. N. (2021). Pengujian hipotesis dalam penelitian kesehatan: Konsep dan aplikasi. *Jurnal Biometrika Kesehatan*, 8(2), 112–120.

- Setiawan, B., & Hermawan, R. (2022). Aplikasi biostatistik dalam epidemiologi penyakit tidak menular. *Media Ilmiah Kesehatan*, 11(1), 55–63.
- Suryani, T., & Wibowo, A. (2021). Pendekatan kualitatif dalam studi kesehatan masyarakat. *Jurnal Ilmu Kesehatan Masyarakat*, 12(1), 41–50.
- Utami, R., & Ramadhan, D. (2023). Korelasi dan regresi dalam penelitian epidemiologi. *Jurnal Ilmu Kesehatan Komunitas*, 11(1), 41–52.

BAB 2

JENIS DAN SUMBER DATA KESEHATAN

Hardian Bimanto, S.KM., M.Kes.

Data kesehatan merupakan aset berharga dalam sistem kesehatan modern yang memungkinkan pengambilan keputusan berbasis bukti (*evidence-based*) untuk meningkatkan kualitas pelayanan kesehatan. Memahami jenis dan sumber data kesehatan sangat penting bagi para ahli epidemiologi, administrator kesehatan, peneliti, dan pengambil kebijakan dalam merancang program kesehatan yang efektif dan terarah. Tulisan ini membahas secara komprehensif berbagai jenis data kesehatan dan sumbernya, dengan penekanan pada pentingnya kualitas data, interoperabilitas, dan aplikasinya dalam praktik kesehatan modern.

2.1 Definisi dan Konsep Data Kesehatan

2.1.1 Pengertian Data Kesehatan

Data kesehatan adalah informasi terstruktur yang dikumpulkan, disimpan, dan dianalisis untuk mendukung pengambilan keputusan dalam bidang kesehatan. Data ini mencakup

berbagai informasi tentang penyakit, risiko kesehatan, perilaku kesehatan, dan capaian kesehatan pada tingkat individu, komunitas, atau populasi. Dalam konteks *real-world data (RWD)*, data kesehatan yang dikumpulkan dalam praktik klinis sehari-hari menjadi sumber informasi yang semakin penting untuk penelitian dan pengawasan kesehatan (Bourke et al., 2020). Data kesehatan dapat berbentuk kuantitatif maupun kualitatif, dan berperan penting dalam epidemiologi deskriptif, investigasi etiologi, pengawasan (*surveillance*) penyakit, dan evaluasi program kesehatan.

2.1.2 Karakteristik Data Kesehatan Berkualitas

Data kesehatan yang berkualitas tinggi harus memenuhi beberapa kriteria penting. Pertama, akurasi data menunjukkan tingkat kebenaran informasi yang dicatat dan disimpan. Kedua, kelengkapan data berarti semua informasi yang diperlukan telah dikumpulkan tanpa ada bagian yang hilang. Ketiga, ketepatan waktu (*timeliness*) mengacu pada seberapa cepat data dapat diakses dan digunakan untuk pengambilan keputusan. Keempat, konsistensi data memastikan informasi yang sama tidak ada perbedaan ketika diakses dari berbagai sumber (Ehsani-Moghaddam et al., 2019). Kelima, keunikan data mencegah duplikasi atau reduplikasi informasi yang tidak perlu. Keenam, koherensi data memastikan bahwa informasi dari berbagai sumber dapat disatukan dan diintegrasikan dengan baik.

2.1.3 Pentingnya Penjaminan Kualitas Data

Penjaminan kualitas data (*data quality assurance*) merupakan proses berkelanjutan untuk memantau dan meningkatkan kualitas data yang dikumpulkan. Dalam sistem kesehatan, buruknya kualitas data dapat mengakibatkan keputusan klinis yang tidak tepat, penelitian yang tidak valid, dan kebijakan kesehatan yang kurang efektif (Njuguna et al., 2019). Oleh karena itu, setiap organisasi kesehatan harus memiliki protokol standar untuk penilaian kualitas data, termasuk verifikasi sumber data, pengecekan konsistensi internal, dan validasi silang dengan sumber data eksternal. Implementasi sistem manajemen data yang baik, pelatihan staf yang memadai, dan penggunaan teknologi otomatisasi dapat meningkatkan kualitas data secara signifikan.

2.1.4 Pentingnya Privasi dan Keamanan Data Kesehatan

Privasi dan keamanan data kesehatan adalah aspek kritis yang tidak boleh diabaikan. Data kesehatan mengandung informasi pribadi yang sensitif, termasuk riwayat medis, diagnosis, dan informasi genetik. Perlindungan data ini memerlukan kerangka kerja keamanan yang ketat, enkripsi data, kontrol akses yang tepat, dan kepatuhan terhadap regulasi perlindungan privasi lokal dan internasional (Tertulino et al., 2023). Organisasi kesehatan harus mengimplementasikan standar keamanan informasi seperti yang didefinisikan dalam standar internasional ISO/IEC 27001 dan memastikan bahwa semua personel yang menangani data kesehatan telah menerima pelatihan tentang privasi dan keamanan data.

2.2 Jenis-Jenis Data Kesehatan

2.2.1 Data Demografis dan Sosioekonomis

Data demografis meliputi informasi tentang karakteristik populasi seperti usia, jenis kelamin, status perkawinan, tingkat pendidikan, dan pekerjaan. Data ini penting untuk memahami komposisi populasi dan mengidentifikasi disparitas kesehatan antar kelompok. Data sosioekonomis mencakup informasi tentang pendapatan, status pekerjaan, kondisi perumahan, dan akses ke layanan kesehatan. Penelitian menunjukkan bahwa faktor sosioekonomis sangat memengaruhi status kesehatan dan akses terhadap layanan kesehatan. Data demografis dan sosioekonomis sering digunakan sebagai variabel perancu dalam analisis epidemiologi dan dapat membantu mengidentifikasi kelompok populasi yang rentan mengalami masalah kesehatan.

2.2.2 Data Klinis dan Catatan Medis Elektronik

Catatan medis elektronik (*electronic health records*, EHR) atau *electronic medical records* (EMR) merupakan kumpulan data klinis yang disimpan dalam format digital. EHR berisi informasi tentang riwayat medis pasien, diagnosis, resep obat, hasil laboratorium, temuan pemeriksaan fisik, dan rencana perawatan. Data klinis ini sangat berharga untuk penelitian epidemiologi, karena memungkinkan para peneliti untuk mengidentifikasi pola penyakit, faktor risiko, dan hasil kesehatan pada skala populasi yang besar (Sarwar et al., 2022). Tantangan utama dalam menggunakan data EHR adalah variabilitas dalam cara data dicatat di antara

penyedia layanan kesehatan yang berbeda, serta kebutuhan untuk standarisasi dan interoperabilitas antar sistem.

2.2.3 Data Laboratorium dan Hasil Pemeriksaan Diagnostik

Data laboratorium mencakup hasil dari berbagai tes darah, urine, dan tes diagnostik lainnya yang dilakukan untuk mengkonfirmasi diagnosis atau memonitor penyakit. Data ini sering diintegrasikan dengan EHR untuk memberikan gambaran lengkap tentang status kesehatan pasien. Hasil pemeriksaan diagnostik seperti sinar-X, *computed tomography (CT)* scan, dan resonansi magnetik (*magnetic resonance imaging, MRI*) juga merupakan bagian penting dari data klinis. Data laboratorium penting untuk penelitian etiologi penyakit dan dapat digunakan untuk mengidentifikasi kelompok berisiko tinggi dalam populasi.

2.2.4 Data Kematian dan Vital Statistik

Data mortalitas mencakup informasi tentang jumlah kematian, penyebab kematian, dan karakteristik demografis dari orang-orang yang meninggal. Vital statistik adalah catatan resmi kejadian demografi penting seperti kelahiran, kematian, pernikahan, dan perceraian. Sistem registrasi vital (*vital registration system*) yang efektif adalah fondasi dari *surveillance* kesehatan masyarakat yang baik. Data kematian digunakan untuk menghitung *age-standardized mortality rate (ASMR)* dan mengidentifikasi tren penyakit penyebab kematian utama dalam populasi. Kualitas data mortalitas sangat tergantung pada keakuratan diagnosis penyakit penyebab kematian yang dicatat oleh penyedia layanan kesehatan.

2.3 Sumber Data Kesehatan

2.3.1 Data dari Sistem Informasi Kesehatan di Fasilitas Pelayanan Kesehatan

Rumah sakit, klinik, dan fasilitas kesehatan lainnya mengumpulkan data kesehatan sebagai bagian dari operasi klinis rutin mereka. Data ini mencakup informasi tentang setiap kunjungan pasien, diagnosis, pengobatan, dan hasil kesehatan. Banyak fasilitas kesehatan modern telah mengimplementasikan sistem informasi kesehatan terintegrasi yang memungkinkan penyimpanan dan pengambilan data yang efisien. Namun, ada tantangan dalam menggunakan data ini untuk penelitian dan *surveillance*, karena data sering dikumpulkan untuk tujuan administratif dan klinis, bukan untuk penelitian (Florentino et al., 2024). Untuk mengatasi hal ini, peneliti harus bekerja sama dengan fasilitas kesehatan untuk memastikan bahwa data dikumpulkan dengan standar yang konsisten dan dapat divalidasi.

2.3.2 Data Registri Penyakit dan Registri Berbasis Populasi

Registri penyakit adalah sistem pengumpulan data yang terstruktur untuk mencatat informasi tentang individu dengan penyakit tertentu atau kelompok penyakit tertentu. Registri berbasis populasi (*population-based registry*) mencakup semua kasus penyakit dalam wilayah geografis tertentu selama periode waktu tertentu, sedangkan registri berbasis fasilitas hanya mencakup pasien yang datang ke fasilitas kesehatan tertentu (Ali et al., 2021). Registri penyakit yang terkenal termasuk registri kanker, registri penyakit

langka, dan registri penyakit kardiovaskular. Data dari registri ini sangat berharga untuk memahami beban penyakit, faktor risiko, hasil kesehatan, dan efektivitas intervensi kesehatan.

2.3.3 Data dari Program Vaksinasi dan Imunisasi

Program vaksinasi dan imunisasi mengumpulkan data tentang cakupan vaksinasi, jenis vaksin yang diberikan, efek samping vaksin, dan hasil kesehatan terkait imunisasi. Data ini penting untuk memantau efektivitas program imunisasi dan mengidentifikasi kelompok populasi yang belum mencapai cakupan vaksinasi yang optimal. Sistem pencatatan vaksinasi yang baik memungkinkan pengidentifikasian kelompok usia atau wilayah geografis dengan cakupan vaksinasi rendah, sehingga intervensi dapat ditargetkan dengan lebih efektif.

2.3.4 Data Pengawasan Penyakit Menular dan Nonmenular

Sistem *surveillance* penyakit menular mengumpulkan data tentang kejadian penyakit menular yang dilaporkan oleh fasilitas kesehatan dan laboratorium diagnostik. Data ini digunakan untuk mendeteksi wabah, mengidentifikasi tren penyakit, dan memandu responss kesehatan masyarakat. Untuk penyakit nonmenular seperti kanker, diabetes, dan penyakit jantung, sistem *surveillance* mengumpulkan data insidensi, prevalensi, mortalitas, dan faktor risiko untuk memahami beban penyakit dan merencanakan intervensi pencegahan. Kualitas data *surveillance* sangat penting karena data ini digunakan untuk pengambilan keputusan kebijakan kesehatan tingkat nasional.

2.4 Teknologi dan Sistem Informasi untuk Manajemen Data Kesehatan

2.4.1 Catatan Medis Elektronik dan Sistem Informasi Kesehatan Terintegrasi

Electronic health records (EHR) dan sistem informasi kesehatan terintegrasi (*integrated health information systems*) memungkinkan penyimpanan dan pengaksesan data kesehatan yang efisien. Sistem ini memfasilitasi komunikasi dan kolaborasi antar penyedia layanan kesehatan, meningkatkan koordinasi perawatan pasien, dan mengurangi risiko kesalahan medis. EHR yang baik dirancang dengan mempertimbangkan kebutuhan pengguna klinis, kemudahan penggunaan, dan integrasi dengan sistem lain (Palojoki et al., 2024). Tantangan dalam implementasi EHR termasuk biaya awal yang tinggi, kurangnya interoperabilitas antar sistem dari vendor yang berbeda, dan kebutuhan untuk melatih staf tentang penggunaan sistem baru.

2.4.2 Interoperabilitas dan Standar Data Kesehatan

Interoperabilitas adalah kemampuan sistem informasi yang berbeda untuk berkomunikasi dan bertukar data secara efektif. Dalam konteks kesehatan, interoperabilitas memungkinkan data dari berbagai fasilitas kesehatan dan sistem informasi untuk diintegrasikan dan dianalisis bersama. Standar data kesehatan seperti *Health Level 7 (HL7)*, *SNOMED CT (Systematized Nomenclature of Medicine Clinical Terms)*, dan *DICOM (Digital Imaging and Communications in Medicine)* memfasilitasi

pertukaran data yang efisien dan konsisten. Dengan meningkatkan interoperabilitas, sistem kesehatan dapat memberikan informasi yang lebih lengkap dan akurat kepada penyedia layanan kesehatan, yang pada akhirnya meningkatkan kualitas perawatan pasien.

2.4.3 Platform Digital dan *Dashboard* Kesehatan

Platform digital dan *dashboard* kesehatan memungkinkan visualisasi dan analisis data kesehatan secara real-time. Platform ini dapat diakses oleh berbagai pemangku kepentingan (*stakeholders*), termasuk pengambil kebijakan, administrator kesehatan, dan klinis, untuk memantau indikator kesehatan kunci dan mengidentifikasi area yang memerlukan perhatian khusus. *Dashboard* yang dirancang dengan baik dapat meningkatkan transparansi, akuntabilitas, dan responss cepat terhadap masalah kesehatan *emerging*. Selama pandemi COVID-19, banyak negara mengembangkan *dashboard* untuk melacak kasus, kematian, dan vaksinasi secara real-time, menunjukkan nilai platform digital dalam komunikasi kesehatan masyarakat.

2.4.4 Keamanan Data dan Arsitektur Teknologi Blockchain

Dengan meningkatnya volume dan sensitivitas data kesehatan yang disimpan secara digital, keamanan data menjadi perhatian utama. Teknologi *blockchain* telah diusulkan sebagai solusi untuk meningkatkan keamanan dan privasi data kesehatan dengan menyediakan sistem penyimpanan yang terdesentralisasi dan tidak dapat diubah (*immutable*) (Gutierrez et al., 2020). Teknologi *blockchain* juga memungkinkan pasien untuk memiliki

kontrol yang lebih besar atas data kesehatan mereka dan memfasilitasi *smart contracts* untuk manajemen izin akses data. Meskipun teknologi ini menjanjikan, tantangan implementasi termasuk skalabilitas, biaya komputasi, dan kebutuhan untuk standarisasi di seluruh sistem kesehatan.

2.5 Tantangan dan Peluang dalam Penggunaan Data Kesehatan

2.5.1 Integrasi Data dari Berbagai Sumber

Integrasi data dari berbagai sumber kesehatan menghadapi tantangan teknis dan metodologi. Perbedaan format data, sistem klasifikasi yang berbeda, dan standar yang tidak konsisten antar fasilitas kesehatan memperumit proses integrasi. Selain itu, data sering kali dikumpulkan untuk tujuan yang berbeda (administratif, klinis, atau penelitian), sehingga mungkin tidak sesuai untuk keperluan lain. Untuk mengatasi tantangan ini, diperlukan investasi dalam infrastruktur teknologi, pengembangan standar data yang konsisten, dan kolaborasi antar organisasi kesehatan untuk berbagi praktik terbaik dalam manajemen data.

2.5.2 Data yang Dihasilkan Pasien dan *Patient-Generated Health Data*

Dengan meningkatnya penggunaan perangkat kesehatan *wearable* dan aplikasi kesehatan di *smartphone*, pasien sekarang dapat mengumpulkan dan berbagi data kesehatan mereka sendiri. Data yang dihasilkan pasien (*patient-generated health data*) mencakup informasi tentang aktivitas fisik, pola makan, kualitas

tidur, tingkat stres, dan gejala kesehatan (Bourke et al., 2020). Data ini dapat memberikan wawasan berharga tentang faktor-faktor yang memengaruhi kesehatan pasien dan memungkinkan intervensi yang dipersonalisasi. Namun, tantangan dalam menggunakan data yang dihasilkan pasien termasuk variabilitas dalam akurasi perangkat, keterbatasan pada jenis data yang dapat dikumpulkan, dan kebutuhan untuk mengintegrasikan data ini dengan data klinis tradisional.

2.5.3 Penggunaan *Non-Traditional Data* dan *Big Data* dalam Kesehatan

Sumber data non-tradisional seperti media sosial, pencarian internet, data mobilitas, dan data penjualan obat dapat memberikan sinyal awal tentang wabah penyakit atau masalah kesehatan *emerging*. Selama pandemi COVID-19, banyak negara menggunakan data non-tradisional untuk melengkapi data *surveillance* tradisional dan mendapatkan gambaran yang lebih lengkap tentang penyebaran penyakit (Bolt et al., 2024). Penggunaan *big data* dan teknik analitik lanjutan seperti *machine learning* memungkinkan identifikasi pola dan prediksi yang tidak mungkin dilakukan dengan metode analisis tradisional. Namun, penggunaan data non-tradisional juga menimbulkan pertanyaan tentang bias data, validitas, privasi, dan kebutuhan untuk konsultasi etika.

2.5.4 Penelitian dan Inovasi dalam Penggunaan Data Kesehatan

Data kesehatan yang tersedia saat ini memberikan peluang besar untuk penelitian dan inovasi dalam bidang kesehatan. Dengan menggunakan data dari registri penyakit, EHR, dan sistem *surveillance*, peneliti dapat melakukan studi kohort besar, uji coba terkontrol berbasis registri, dan penelitian komparatif efektivitas dengan biaya yang jauh lebih rendah daripada penelitian tradisional. Kemajuan dalam teknologi analitik, termasuk *artificial intelligence (AI)* dan *machine learning*, membuka peluang baru untuk penemuan pengetahuan dari data kesehatan yang ada. Misalnya, model *machine learning* dapat digunakan untuk memprediksi hasil kesehatan, mengidentifikasi pasien berisiko tinggi, dan mengoptimalkan alokasi sumber daya kesehatan.

Latihan Soal (Esai)

1. Jelaskan berbagai jenis data kesehatan, termasuk data kuantitatif dan kualitatif, serta contoh penggunaannya dalam penelitian dan praktik kesehatan masyarakat.
2. Sebutkan sumber data kesehatan primer dan sekunder. Bagaimana masing-masing sumber digunakan untuk pemantauan kesehatan dan pengambilan keputusan?
3. Uraikan perbedaan antara data individu (misal rekam medis) dan data populasi (misal survei kesehatan nasional).

Bagaimana keduanya saling melengkapi dalam analisis kesehatan?

4. Jelaskan pentingnya kualitas data kesehatan, termasuk akurasi, konsistensi, dan keterkinian informasi untuk perencanaan program kesehatan.
5. Sebutkan tantangan dalam pengumpulan dan pengelolaan data kesehatan, seperti keterbatasan akses, kerahasiaan, dan integrasi data lintas sumber.

DAFTAR PUSTAKA

- Ali, S., Bryce, J., Kodra, Y., Taruscio, D., Persani, L., & Ahmed, S. F. (2021). The Quality Evaluation of Rare Disease Registries—An Assessment of the Essential Features of a Disease Registry. *International Journal of Environmental Research and Public Health*, 18(22), 11968. <https://doi.org/10.3390/ijerph182211968>
- Bolt, K., Gil-González, D., & Oliver, N. (2024). Unconventional data, unprecedented insights: leveraging non-traditional data during a pandemic. *Frontiers in Public Health*, 12, 1350743. <https://doi.org/10.3389/fpubh.2024.1350743>
- Bourke, A., Dixon, W., Roddam, A., Lin, K. J., Hall, G., Curtis, J., ... & Bartels, D. (2020). Incorporating patient generated health data into pharmacoepidemiological research. *Pharmacoepidemiology and Drug Safety*, 29(11), 1437-1449. <https://doi.org/10.1002/pds.5169>
- Ehsani-Moghaddam, B., Martin, K., & Queenan, J. (2019). Data quality in healthcare: A report of practical experience with the Canadian Primary Care Sentinel Surveillance Network data. *SAGE Open Nursing*, 5, 2377960819887743. <https://doi.org/10.1177/1833358319887743>
- Florentino, P., Bertoldo Junior, J., Barbosa, G. C. G., Cerqueira-Silva, T., Oliveira, V. A., Garcia, M. H. O., ... & Marcilio, I. (2024). Impact of Primary Health Care Data Quality on

- Infectious Disease Surveillance in Brazil: Case Study. *JMIR Public Health and Surveillance*, 10, 67050. <https://doi.org/10.2196/67050>
- Njuguna, C., Vandi, M. A., Mugagga, M., Kanu, J., Liyosi, E., Chimbaru, A., ... & Talisuna, A. (2020). Institutionalized data quality assessments: a critical pathway to improving the accuracy of integrated disease surveillance data in Sierra Leone. *BMC Health Services Research*, 20(1), 51. <https://doi.org/10.1186/s12913-020-05591-x>
- Palojoki, S., Lehtonen, L., & Vuokko, R. (2024). Semantic Interoperability of Electronic Health Records: Systematic Review of Alternative Approaches for Enhancing Patient Information Availability. *JMIR Medical Informatics*, 12(4), 53535. <https://doi.org/10.2196/53535>
- Sarwar, T., Seifollahi, S., Chan, J. A., Zhang, X., Aksakalli, V., Hudson, I., ... & Cavedon, L. (2022). The Secondary Use of Electronic Health Records for Data Mining: Data Characteristics and Challenges. *ACM Computing Surveys*, 55(2), 1-41. <https://doi.org/10.1145/3490234>

BAB 3

PENGOLAHAN DAN PENYAJIAN DATA STATISTIK

Desie Rahmawati, S.KM., M.Kes.

3.1 Konsep Dasar Pengolahan Data

3.1.1 Pengertian dan Tujuan Pengolahan Data

Pengolahan data merupakan rangkaian kegiatan sistematis untuk mengubah data mentah (*raw data*) menjadi informasi yang terorganisir, akurat, dan siap untuk dianalisis lebih lanjut (Mishra & Khan, 2024). Dalam konteks kesehatan masyarakat, pengolahan data merupakan tahapan fundamental yang menentukan kualitas seluruh proses penelitian dan keputusan kesehatan publik. Proses ini mencakup berbagai aktivitas mulai dari pengumpulan data awal, pencatatan data, pemeriksaan kualitas, hingga penyimpanan data dalam format yang terstruktur dan mudah diakses oleh peneliti.

Tujuan utama pengolahan data dalam kesehatan masyarakat adalah mengubah data yang semula tidak terorganisir menjadi informasi bermakna yang dapat digunakan untuk pengambilan

keputusan klinis dan kebijakan kesehatan publik (Antomonov, 2024). Lebih khusus lagi, pengolahan data bertujuan untuk memastikan akurasi tinggi terhadap setiap nilai data, menghilangkan data yang hilang atau data yang salah tercatat, mengidentifikasi dan mengatasi nilai pencilan (*outliers*) yang tidak realistis, mengorganisir data sesuai dengan kategori atau variabel yang relevan dengan penelitian, dan mempersiapkan data dalam format yang siap untuk analisis statistik lebih mendalam. Dengan melakukan pengolahan data yang baik dan sistematis, data mentah yang awalnya tidak memberikan informasi berguna dapat diubah menjadi aset yang sangat berharga untuk pengambilan keputusan kesehatan.

3.1.2 Tahapan Proses Pengolahan Data

Proses pengolahan data dalam kesehatan masyarakat mengikuti tahapan sistematis yang telah teruji efektivitasnya. Tahap pertama adalah persiapan data (*data preparation*), di mana data mentah yang baru dikumpulkan dari lapangan diinspeksi kelengkapannya dan diformat sesuai dengan standar yang telah ditentukan sebelumnya (Junker et al., 2023). Dalam tahap ini, petugas pencatat data melakukan verifikasi awal untuk memastikan setiap formulir atau instrumen pengumpulan data telah diisi dengan lengkap, jelas, dan sesuai protokol yang disepakati. Seluruh formulir harus diperiksa untuk memastikan tidak ada halaman yang hilang atau data penting yang tidak tercatat.

Tahap kedua adalah pemasukan data (*data entry*) ke dalam sistem informasi atau program komputer yang telah disiapkan. Data yang sebelumnya tercatat dalam bentuk kertas atau dokumen elektronik dimasukkan ke dalam *database* digital menggunakan aplikasi khusus seperti *Epidata*, SPSS, atau sistem manajemen data lainnya yang sesuai dengan jenis penelitian (Mishra & Khan, 2024). Pemasukan data harus dilakukan dengan cermat dan teliti untuk meminimalkan kesalahan pengetikan atau pengkodean. Banyak sistem modern dilengkapi dengan fitur validasi otomatis yang mencegah pemasukan data yang tidak sesuai dengan format atau kriteria yang telah ditentukan sebelumnya, sehingga mengurangi risiko kesalahan input.

Tahap ketiga adalah pembersihan data (*data cleaning*). Pada tahap kritis ini, data yang telah dimasukkan ke dalam sistem diperiksa secara menyeluruh untuk mengidentifikasi dan memperbaiki kesalahan, data yang hilang, atau nilai yang tidak masuk akal dalam konteks penelitian (Antomonov, 2024). Kesalahan dapat terjadi karena berbagai alasan termasuk kesalahan penulisan pada formulir asli, kesalahan saat pemasukan data oleh petugas, atau kesalahan dalam transmisi data antar sistem. Pembersihan data melibatkan pengecekan konsistensi data, perbandingan lintas variabel untuk memastikan logika, dan identifikasi nilai-nilai yang tidak realistis atau tidak sesuai dengan kriteria inklusi penelitian. Proses ini sangat penting untuk memastikan integritas data sebelum analisis dilakukan.

Tahap keempat adalah transformasi data. Data yang telah dibersihkan kemudian ditransformasi sesuai dengan kebutuhan spesifik analisis, seperti membuat kategori baru dari variabel kontinyu, menggabungkan beberapa variabel, atau membuat variabel turunan (*derived variables*) yang dihitung dari variabel lain (Junker et al., 2023). Transformasi ini memungkinkan peneliti untuk mengorganisir data dengan cara yang lebih sesuai untuk pertanyaan penelitian. Tahap kelima adalah analisis data awal atau *exploratory data analysis* untuk memahami karakteristik umum data sebelum melakukan analisis statistik yang lebih mendalam dan kompleks.

3.1.3 Sumber Data dan Jenis Data

Dalam kesehatan masyarakat, data dapat berasal dari berbagai sumber yang berbeda. Sumber data primer adalah data yang dikumpulkan secara langsung melalui survei khusus, wawancara mendalam dengan responden, atau pengukuran langsung pada subjek penelitian menggunakan alat ukur standar (Olowe et al., 2024). Data primer memberikan informasi spesifik sesuai kebutuhan penelitian tetapi memerlukan investasi waktu dan sumber daya yang lebih besar. Sumber data sekunder adalah data yang sudah dikumpulkan oleh institusi atau peneliti lain sebelumnya, seperti data dari pencatatan medis rumah sakit, laporan kesehatan publik resmi, atau *database* surveillance penyakit.

Data dalam kesehatan masyarakat dapat diklasifikasikan berdasarkan jenisnya menjadi beberapa kategori dengan

karakteristik yang berbeda. Data nominal adalah data yang hanya dapat dikategorikan tanpa urutan tertentu, seperti jenis kelamin (laki-laki atau perempuan), status pernikahan, atau diagnosis penyakit spesifik (Mishra & Khan, 2024). Data ordinal adalah data yang dapat diurutkan dengan jelas, seperti tingkat pendidikan (SD, SMP, SMA, Universitas) atau tingkat keparahan penyakit (ringan, sedang, berat). Data interval adalah data numerik dengan jarak yang sama antara nilai-nilai yang berturutan, seperti suhu tubuh dalam derajat Celsius atau skor pada skala pengukuran. Data rasio adalah data numerik dengan titik nol yang bermakna secara biologis, seperti berat badan dalam kilogram, tinggi badan dalam sentimeter, atau umur dalam tahun (Antomonov, 2024).

3.1.4 Validasi dan Verifikasi Data

Validasi data adalah proses memastikan bahwa data yang dikumpulkan akurat, lengkap, dan sesuai dengan kriteria yang telah ditentukan sebelumnya pada protokol penelitian (Junker et al., 2023). Verifikasi data adalah proses memeriksa bahwa data telah dicatat dengan benar dan sesuai dengan sumber aslinya tanpa perubahan atau distorsi. Kedua proses ini sangat penting untuk memastikan kualitas tinggi dari data dalam penelitian kesehatan masyarakat karena kesalahan data dapat menyebabkan kesimpulan yang salah.

Proses validasi melibatkan pemeriksaan range nilai (misalnya, umur tidak boleh lebih dari 150 tahun atau kurang dari 0), pemeriksaan konsistensi logis antara variabel (misalnya, ibu rumah

tangga tidak boleh memiliki penghasilan dari pekerjaan formal), dan pemeriksaan kelengkapan data (tidak ada data yang hilang untuk variabel penting dalam penelitian) (Mishra & Khan, 2024). Verifikasi dapat dilakukan dengan membandingkan kembali data yang telah dimasukkan ke dalam sistem dengan formulir asli (*double entry*) atau dengan melakukan pemeriksaan sampel (*spot check*) pada sebagian kecil data. Dalam era digital, banyak program komputer yang dilengkapi dengan fitur validasi otomatis yang dapat mendeteksi kesalahan data secara real-time saat pemasukan data dilakukan, sehingga meningkatkan efisiensi dan akurasi.

3.2 Penyajian Data Statistik

3.2.1 Prinsip Penyajian Data yang Efektif

Penyajian data yang efektif adalah kombinasi seni dan sains dalam mengkomunikasikan informasi statistik dengan cara yang jelas, akurat, dan mudah dipahami oleh target audiens yang beragam (Olowe et al., 2024). Prinsip pertama adalah kejelasan dan kesederhanaan dalam penyajian. Data harus disajikan dalam bentuk yang mudah dipahami oleh pembaca tanpa mengorbankan akurasi dan kelengkapan informasi. Penyajian yang terlalu rumit atau menggunakan istilah teknis yang tidak perlu dapat membingungkan pembaca dan mengurangi efektivitas komunikasi penelitian.

Prinsip kedua adalah akurasi data. Setiap angka, grafik, atau tabel yang disajikan harus mencerminkan data asli tanpa distorsi, pemalsuan, atau kesalahan interpretasi yang dapat menyesatkan

pembaca. Prinsip ketiga adalah relevansi konten yang dipilih. Hanya data yang relevan dengan pertanyaan penelitian atau tujuan komunikasi utama yang harus disajikan dalam laporan, menghindari penyajian data yang tidak penting. Prinsip keempat adalah konsistensi dalam format penyajian (Antomonov, 2024). Format penyajian, unit pengukuran, dan notasi yang digunakan harus konsisten di seluruh laporan atau publikasi untuk menghindari kebingungan pembaca.

3.2.2 Tabel Statistik dan Distribusi Frekuensi

Tabel merupakan salah satu cara paling efektif dan tradisional untuk menyajikan data statistik, terutama ketika ada banyak angka atau kategori data yang perlu ditampilkan (Mishra & Khan, 2024). Tabel distribusi frekuensi menunjukkan berapa kali setiap nilai atau kategori muncul dalam data yang dikumpulkan, memberikan gambaran singkat tentang pola distribusi. Contoh sederhana adalah tabel yang menunjukkan jumlah orang berdasarkan kategori umur, tingkat pendidikan, atau status keluarga.

Tabel yang baik harus memiliki beberapa elemen penting untuk meningkatkan kejelasan. Judul yang jelas menjelaskan isi dan tujuan tabel. Label baris dan kolom yang deskriptif membantu pembaca memahami setiap entri data. Satuan pengukuran harus tercantum dengan jelas. Catatan kaki (*footnotes*) diperlukan untuk menjelaskan istilah atau singkatan khusus yang mungkin tidak diketahui oleh sebagian pembaca (Junker et al., 2023). Dalam kesehatan masyarakat, tabel sering digunakan untuk

membandingkan frekuensi atau persentase antar kelompok, misalnya membandingkan prevalensi penyakit antara wilayah urban dan rural, antara kelompok umur yang berbeda, atau antara laki-laki dan perempuan.

3.2.3 Grafik dan Visualisasi Data

Grafik adalah representasi visual dari data yang dapat membantu pembaca dengan cepat memahami pola, tren, dan hubungan dalam data tanpa perlu membaca angka secara detail (Antomonov, 2024). Berbagai jenis grafik tersedia untuk tujuan yang berbeda sesuai dengan jenis data dan pertanyaan penelitian. Grafik batang (*bar chart*) cocok digunakan untuk membandingkan frekuensi atau nilai antara kategori yang berbeda secara simultan. Grafik garis (*line graph*) sangat cocok untuk menunjukkan tren atau perubahan dari waktu ke waktu dalam penyakit atau indikator kesehatan. Grafik pie atau diagram lingkaran cocok untuk menunjukkan proporsi atau persentase dari berbagai komponen yang membentuk keseluruhan total (Olowe et al., 2024).

Histogram digunakan untuk menampilkan distribusi frekuensi dari data numerik yang kontinyu dalam bentuk batang yang berdekatan. *Scatter plot* atau diagram pencar digunakan untuk menunjukkan hubungan atau korelasi antara dua variabel numerik, membantu peneliti mengidentifikasi pola hubungan. Pemilihan jenis grafik yang tepat sangat penting untuk memastikan bahwa pesan dan temuan utama dari data tersampaikan dengan efektif kepada

pembaca. Grafik yang dipilih harus sederhana, tidak berlebihan dengan elemen dekoratif, dan fokus pada penyajian data yang jelas.

3.2.4 Ukuran Pemusatan dan Penyebaran Data

Ukuran pemusatan (*measure of central tendency*) adalah nilai statistik yang mewakili atau "memusatkan" data, memberikan gambaran singkat tentang nilai karakteristik umumnya (Mishra & Khan, 2024). Rata-rata (*mean*) adalah jumlah semua nilai dibagi dengan banyaknya nilai yang ada. Median adalah nilai tengah ketika data diurutkan dari kecil ke besar. Modus adalah nilai yang paling sering muncul atau memiliki frekuensi tertinggi dalam data. Ketiga ukuran ini memiliki karakteristik berbeda dan dipilih sesuai dengan distribusi data dan tujuan analisis.

Ukuran penyebaran (*measure of dispersion*) menunjukkan seberapa jauh data tersebar dari nilai pusatnya, memberikan informasi tentang variabilitas data (Antomonov, 2024). Range atau rentang adalah selisih sederhana antara nilai terbesar dan terkecil dalam data. Deviasi standar (*standard deviation*) mengukur seberapa jauh rata-rata penyimpangan setiap nilai dari rata-rata keseluruhan, memberikan informasi yang lebih rinci tentang dispersi. Varians adalah kuadrat dari deviasi standar. Dalam kesehatan masyarakat, pemahaman mendalam tentang ukuran-ukuran ini sangat penting untuk menggambarkan dan membandingkan karakteristik populasi atau sampel penelitian dengan akurat.

3.3 Analisis Deskriptif dan Inferensial

3.3.1 Statistik Deskriptif dalam Kesehatan Masyarakat

Statistik deskriptif adalah teknik statistik untuk merangkum dan mendeskripsikan karakteristik utama dari data yang dikumpulkan, memberikan gambaran komprehensif tentang data tanpa melakukan inferensi (Olowe et al., 2024). Statistik deskriptif mencakup perhitungan frekuensi dan persentase setiap kategori, perhitungan rata-rata, median, modus, serta visualisasi data melalui tabel dan grafik yang informatif. Dalam kesehatan masyarakat, statistik deskriptif sangat berguna untuk memberikan gambaran awal yang jelas tentang profil kesehatan suatu populasi atau komunitas tertentu.

Contoh aplikasi praktis statistik deskriptif dalam kesehatan masyarakat termasuk menggambarkan karakteristik demografis lengkap dari populasi (umur, jenis kelamin, tingkat pendidikan, status pekerjaan), prevalensi berbagai penyakit dalam komunitas tertentu, distribusi jumlah kasus penyakit menurut wilayah geografis atau waktu tertentu, serta profil lengkap faktor risiko dalam populasi (Antomonov, 2024). Data deskriptif ini sering menjadi langkah pertama penting dalam analisis epidemiologis dan dapat memberikan wawasan penting tentang pola kesehatan masyarakat, memandu penelitian lebih lanjut.

3.3.2 Statistik Inferensial dan Uji Hipotesis

Statistik inferensial adalah teknik statistik untuk membuat kesimpulan tentang populasi berdasarkan informasi dan data yang

diperoleh dari sampel penelitian yang telah dipilih (Mishra & Khan, 2024). Uji hipotesis adalah prosedur statistik formal untuk menentukan apakah observasi yang ditemukan dalam sampel mencerminkan perbedaan atau hubungan yang benar-benar ada di populasi, atau hanya terjadi karena kebetulan semata (*chance*).

Dalam kesehatan masyarakat, uji hipotesis sering digunakan untuk membandingkan dua atau lebih kelompok, misalnya membandingkan angka kejadian penyakit antara dua daerah yang berbeda atau membandingkan efektivitas dua intervensi kesehatan yang berbeda (Junker et al., 2023). Uji-t (*t-test*) digunakan untuk membandingkan rata-rata dua kelompok independen. Uji *chi-square* digunakan untuk menguji asosiasi atau hubungan antara dua variabel kategorik. ANOVA digunakan untuk membandingkan rata-rata lebih dari dua kelompok secara simultan. Nilai p (*p-value*) mewakili probabilitas bahwa observasi yang ditemukan akan terjadi jika hipotesis nol benar-benar berlaku. Nilai p kurang dari 0.05 secara konvensional dianggap signifikan secara statistik dalam penelitian kesehatan masyarakat (Antomonov, 2024).

3.3.3 Penggunaan Software Statistik

Dalam era modern dan digital, pengolahan dan analisis data statistik dilakukan secara rutin menggunakan program komputer khusus yang dirancang untuk tujuan ini (Olowe et al., 2024). Program statistik yang umum digunakan dalam kesehatan masyarakat termasuk SPSS (*Statistical Package for Social Sciences*), R, Python, SAS, dan Stata. Setiap program memiliki

kekuatan, kelemahan, dan keunikan tersendiri dalam cara pengoperasian dan analisis.

SPSS adalah program yang terkenal *user-friendly* dengan antarmuka grafis yang intuitif, memudahkan pengguna untuk melakukan analisis tanpa harus menulis kode program yang kompleks. R adalah program *open-source* yang sangat fleksibel, powerful, dan mendukung analisis statistik lanjutan, tetapi memerlukan pengetahuan *programming* yang lebih mendalam. Python semakin populer untuk analisis data besar (*big data*) dan aplikasi *machine learning*. Pemilihan software tergantung pada kebutuhan spesifik penelitian, ketersediaan sumber daya finansial, dan tingkat keahlian peneliti dalam penggunaan teknologi (Mishra & Khan, 2024).

3.3.4 Interpretasi Hasil Analisis

Interpretasi hasil analisis statistik memerlukan pemahaman mendalam tentang metode statistik yang digunakan dan konteks penelitian secara keseluruhan (Antomonov, 2024). Peneliti harus mempertimbangkan dua aspek penting: signifikansi statistik (apakah hasil bermakna secara statistik berdasarkan nilai p) dan signifikansi klinis atau praktis (apakah hasil bermakna dalam praktik kesehatan masyarakat yang nyata).

Interpretasi yang baik juga memerlukan pertimbangan cermat terhadap faktor-faktor lain yang mungkin memengaruhi hasil, seperti bias potensial dalam desain, *confounding variables* yang tidak terukur, atau keterbatasan desain penelitian

yang digunakan (Junker et al., 2023). Seorang peneliti harus berhati-hati dalam membuat kesimpulan dan menghindari klaim kausal yang tidak didukung oleh bukti yang kuat dan metodologi yang tepat. Kesalahan interpretasi yang umum termasuk menginterpretasikan korelasi sebagai kausalitas, memberikan interpretasi yang terlalu luas dari hasil penelitian yang terbatas, atau mengabaikan nilai p yang marginal.

3.4 Praktik Terbaik dalam Pengolahan Data

3.4.1 Manajemen Data dan Dokumentasi

Manajemen data yang baik dan terstruktur adalah fondasi penting untuk penelitian berkualitas tinggi dalam kesehatan masyarakat (Olowe et al., 2024). Ini meliputi dokumentasi yang lengkap dan terinci tentang prosedur pengumpulan data, definisi operasional dari setiap variabel, kode data yang digunakan, dan prosedur pembersihan data yang dilakukan. Dokumentasi yang baik memastikan transparansi penuh dalam penelitian dan memudahkan verifikasi serta reproduksi penelitian oleh peneliti lain di masa depan.

Data harus disimpan dengan aman menggunakan sistem yang terpercaya, dengan kontrol akses yang sesuai dan backup reguler untuk mencegah kehilangan data yang tidak terduga. Dalam era digital, penting untuk menggunakan protokol keamanan data yang standar internasional, termasuk enkripsi data sensitif dan penggunaan password yang kuat dan unik (Antomonov, 2024).

Organisasi file dan *database* harus logis, konsisten, dan mudah dipahami, memudahkan peneliti menemukan data yang diperlukan dengan cepat tanpa kebingungan.

3.4.2 Etika Pengolahan Data Kesehatan

Pengolahan data kesehatan melibatkan informasi yang sangat sensitif dan pribadi dari responden, oleh karena itu etika penanganan data adalah prioritas utama yang tidak dapat diabaikan (Mishra & Khan, 2024). Prinsip-prinsip etika meliputi privasi (perlindungan informasi pribadi dari akses tidak sah), kerahasiaan (menjaga informasi tetap rahasia dan tidak dikirimkan tanpa otorisasi), transparansi (mengkomunikasikan bagaimana data akan digunakan), dan keadilan (memastikan data digunakan adil dan tidak diskriminatif).

Dalam kebanyakan yurisdiksi, ada regulasi ketat tentang penanganan data kesehatan pribadi yang harus dipatuhi. Peneliti harus memahami dan mematuhi peraturan ini, termasuk memperoleh persetujuan etik (*ethical approval*) dari komite penelitian setempat sebelum mengumpulkan data dari subyek manusia (Junker et al., 2023). Data yang diidentifikasi secara pribadi (*personally identifiable information*) harus dikodifikasi dengan cermat atau dianonimkan sepenuhnya sebelum disimpan dalam sistem atau dibagikan dengan pihak ketiga.

3.4.3 Transparansi dan Reprodusibilitas

Transparansi dalam pengolahan dan analisis data adalah prinsip penting dalam penelitian ilmiah modern yang mendukung

kredibilitas sains (Antomonov, 2024). Peneliti harus secara jelas mendokumentasikan setiap langkah dalam proses analisis, termasuk keputusan yang dibuat, alasan logis di balik keputusan tersebut, dan asumsi yang digunakan dalam analisis. Dengan melakukan ini, peneliti memungkinkan pembaca untuk mengevaluasi kredibilitas hasil secara kritis dan memungkinkan peneliti lain untuk mereplikasi analisis dengan tepat.

Reproduksibilitas berarti bahwa peneliti lain dengan menggunakan data dan prosedur yang sama akan mendapatkan hasil yang identik atau sangat mirip (Olowe et al., 2024). Ini penting untuk membangun kepercayaan dalam hasil penelitian dan mendukung kemajuan ilmiah. Dalam era digital modern, semakin banyak jurnal akademik berkualitas tinggi yang meminta peneliti untuk menyediakan kode analisis lengkap dan data (dalam format yang tersanitasi dan telah dianonimkan) sebagai bagian integral dari proses publikasi.

3.4.4 Komunikasi Hasil Analisis

Komunikasi hasil analisis statistik yang efektif adalah keterampilan penting dan praktis untuk peneliti kesehatan masyarakat (Mishra & Khan, 2024). Hasil penelitian harus dikomunikasikan dengan cara yang akurat namun mudah dipahami oleh audiens target, baik itu profesional kesehatan, pembuat kebijakan, atau masyarakat umum yang non-teknis.

Dalam komunikasi kepada profesional kesehatan atau peneliti lain, penyajian dapat lebih teknis dan detail dengan istilah

statistik yang spesifik. Dalam komunikasi kepada pembuat kebijakan atau masyarakat umum, penyajian harus lebih sederhana dan menghindari jargon teknis yang tidak perlu agar tidak membingungkan audiens (Antomonov, 2024). Grafik dan visualisasi yang jelas dan menarik sering lebih efektif daripada tabel angka yang kompleks dalam mengkomunikasikan temuan kepada audiens yang non-teknis dan membantu transfer pengetahuan yang lebih baik.

Latihan Soal

Berikut ini disajikan lima soal esai singkat tanpa panduan jawaban untuk menguji pemahaman Anda tentang pengolahan dan penyajian data statistik dalam kesehatan masyarakat.

Soal 1. Jelaskan secara detail perbedaan antara data primer dan data sekunder dalam konteks penelitian kesehatan masyarakat. Berikan masing-masing satu contoh konkret dari setiap jenis data dan diskusikan secara mendalam keuntungan serta keterbatasan masing-masing jenis data dalam praktik penelitian lapangan.

Soal 2. Anda sedang melakukan penelitian tentang prevalensi hipertensi di suatu komunitas perkotaan. Jelaskan secara terperinci langkah-langkah pengolahan data yang harus Anda lakukan, mulai dari pengumpulan data awal di lapangan hingga persiapan akhir data untuk analisis statistik lebih lanjut, dengan menekankan tantangan yang mungkin dihadapi.

Soal 3. Diskusikan secara kritis mengapa validasi dan verifikasi data sangat penting dalam penelitian kesehatan

masyarakat. Jelaskan potensi dampak negatif yang nyata dari data yang tidak valid atau tidak terverifikasi dengan baik terhadap hasil penelitian dan pengambilan keputusan kesehatan publik yang dapat memengaruhi ribuan orang.

Soal 4. Anda memiliki data tentang jumlah kasus penyakit menular di lima wilayah geografis berbeda selama periode 12 bulan berturut-turut. Jelaskan jenis grafik atau visualisasi data apa yang paling sesuai dan efektif untuk menyajikan data ini dan jelaskan alasan logis di balik pilihan Anda dengan merujuk pada prinsip-prinsip penyajian data yang efektif.

Soal 5. Jelaskan prinsip-prinsip etika fundamental yang harus dipertimbangkan saat menangani data kesehatan pribadi yang sensitif dalam penelitian kesehatan masyarakat. Bagaimana Anda memastikan bahwa data tetap rahasia dan privasi responden terlindungi dengan optimal sambil tetap memungkinkan analisis data yang bermakna untuk kepentingan publik kesehatan?

DAFTAR PUSTAKA

- Antomonov, M. (2024). Processing and analysis of the results questionnaires in the public health system. *Environment & Health*. <https://doi.org/10.32402/dovkil2024.03.033>
- Junker, S., Damerow, S., Walther, L., & Mauz, E. (2023). Development of a prototype for high-frequency mental health surveillance in Germany: data infrastructure and statistical methods. *Frontiers in Public Health*. <https://doi.org/10.3389/fpubh.2023.1208515>
- Mishra, A., & Khan, Mohd. M. (2024). The science and art of biostatistics: a comprehensive overview. *Asian Journal of Hospital Pharmacy*. <https://doi.org/10.38022/ajhp.v4i2.86>
- Olowe, K. J., Edoh, N. L., Zouo, S. J. C., & Olamijuwon, J. (2024). Theoretical perspectives on biostatistics and its multifaceted applications in global health studies. *International Journal of Applied Research in Social Sciences*. <https://doi.org/10.51594/ijarss.v6i11.1726>

BAB 4

UKURAN PEMUSATAN DAN PENYEBARAN DATA

Shiva Aflahiyah, S.Tr.Keb., M.K.M.

Sebelum memasuki ukuran pemusatan data dan penyebaran data, perlu dipahami terlebih dahulu konsep dasar mengenai statistika. Statistika adalah bidang ilmu yang mempelajari proses pengumpulan, pengolahan, peringkasan, dan analisis data, serta penarikan kesimpulan berdasarkan data sampel guna menggambarkan atau merepresentasikan karakteristik populasi secara keseluruhan berdasarkan hasil observasi (Ali et al., 2019; Islam & Al-Shiha, 2018). Statistika secara umum dibagi menjadi 2 jenis, yakni statistika deskriptif dan inferensial. Statistika deskriptif bertujuan untuk menjelaskan atau menggambarkan karakteristik masing-masing variabel yang diteliti, sedangkan statistik inferensial bertujuan untuk menarik kesimpulan mengenai populasi berdasarkan sampel acak yang diambil dari populasi tersebut (Ali et al., 2019; Manikandan, 2011).

Statistika deskriptif mencakup berbagai metode yang digunakan untuk menyederhanakan kumpulan data yang besar, yang

biasanya disajikan dalam bentuk tabel atau grafik, sehingga karakteristik distribusi data dapat dipahami dengan lebih mudah. Ukuran statistik deskriptif yang sering digunakan pada umumnya ada 2, yakni ukuran pemusatan data dan ukuran penyebaran data (Ali et al., 2019; Islam & Al-Shiha, 2018).

4.1 Konsep Dasar Ukuran Pemusatan Data

Ukuran pemusatan data merupakan bagian dari statistik deskriptif yang digunakan untuk menggambarkan nilai pusat atau nilai yang paling mewakili suatu variabel penelitian dan menjadi acuan dalam melihat sebaran data. Pada umumnya, nilai-nilai suatu variabel cenderung terkonsentrasi di sekitar pusat data, sehingga ukuran pemusatan data diperlukan untuk mengetahui gambaran umum dan karakteristik utama dari data yang diteliti.

Pada bidang kesehatan masyarakat, ukuran pemusatan data dapat memberikan rangkuman informasi kuantitatif yang berkaitan dengan kondisi kesehatan suatu populasi, seperti umur, tekanan darah, indeks massa tubuh, atau kadar zat gizi tertentu. Selain itu, ukuran pemusatan data berfungsi merangkum data kesehatan yang kompleks menjadi informasi yang lebih sederhana dan mudah dipahami, sehingga mendukung proses analisis, evaluasi, serta pengambilan keputusan di bidang kesehatan. Ukuran pemusatan data yang paling umum digunakan meliputi *mean*, *median*, dan *mode* (Hastono, 2024).

Pemilihan ukuran pemusatan yang tepat sangat bergantung pada karakteristik data dan tujuan analisis. Tidak ada satu ukuran pemusatan yang selalu paling benar untuk semua situasi. Oleh karena itu, pemahaman yang baik mengenai sifat distribusi data menjadi kunci dalam menentukan ukuran pemusatan yang paling representatif. Dalam praktik biostatistik kesehatan masyarakat, penggunaan ukuran pemusatan yang tepat akan meningkatkan ketepatan interpretasi data dan mendukung penyusunan kebijakan kesehatan yang lebih berbasis bukti. Dengan demikian, konsep dasar ukuran pemusatan data tidak hanya bersifat teknis, tetapi juga memiliki implikasi penting dalam perencanaan, evaluasi, dan pengambilan keputusan di bidang kesehatan masyarakat (Triola, 2020; Rosner, 2023).

4.2 Jenis-Jenis Ukuran Pemusatan

Tiga ukuran pemusatan yang paling umum digunakan adalah *mean*, *median*, dan *mode*. Masing-masing ukuran memiliki karakteristik, kelebihan, dan keterbatasan yang perlu dipertimbangkan sesuai dengan sifat data dan tujuan analisis.

4.2.1 Mean sebagai Nilai Rata-Rata

Mean atau nilai rata-rata diperoleh dengan menjumlahkan seluruh nilai data kemudian membaginya dengan jumlah observasi. Ukuran ini banyak digunakan karena mampu memberikan gambaran umum tentang tingkat atau kecenderungan nilai data secara keseluruhan. Dalam konteks data numerik yang berdistribusi

simetris, *mean* sering dianggap sebagai ukuran pemusatan yang paling representatif (Ali et al., 2019; Islam & Al-Shiha, 2018).

Mean dibedakan menjadi mean populasi dan mean sampel. Mean populasi dilambangkan dengan μ dan dihitung dari seluruh nilai yang terdapat dalam populasi. Sedangkan mean sampel dilambangkan dengan $\bar{x}$ dan dihitung dari data sampel yang tersedia. Mean sampel digunakan sebagai pendekatan untuk memperkirakan nilai mean populasi (Ali et al., 2019; Islam & Al-Shiha, 2018). Rumus mean populasi dipaparkan sebagai berikut:

$$\mu = \frac{X_1 + X_2 + \dots + X_n}{N} = \frac{\sum_{i=1}^N X_i}{N}$$

Keterangan

μ	:	Mean populasi
X_i	:	Nilai ke-i dalam populasi
N	:	Jumlah seluruh anggota populasi

Sedangkan rumus mean sampel dapat dihitung berdasarkan data tidak berkelompok maupun data yang disajikan dalam bentuk kelompok yakni dengan distribusi frekuensi. Rumus mean data tidak berkelompok dipaparkan sebagai berikut:

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

Keterangan

$\bar{x}$	:	Mean sampel
x_i	:	Nilai ke-i dalam sampel

n : Jumlah anggota sampel

Berbeda dengan data tidak berkelompok, pada data sampel yang disajikan dalam distribusi frekuensi setiap nilai data memiliki frekuensi tertentu, sehingga perhitungan mean sampel dilakukan dengan menggunakan nilai data dan frekuensinya. Berikut rumus dari mean sampel data berkelompok (Islam & Al-Shiha, 2018)

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + \dots + f_nx_n}{f_1 + f_2 + \dots + f_k} = \frac{\sum_{i=1}^n f_i x_i}{n}$$

Keterangan

x_i : Nilai data (misal berat badan, jumlah kunjungan, jumlah anak)

f_i : frekuensi kemunculan nilai tersebut

n : Jumlah seluruh data

Agar lebih mudah memahami penggunaan masing-masing mean, berikut contoh pengaplikasiannya:

Contoh Mean Populasi

Seorang perawat ingin mengetahui rata-rata suhu tubuh seluruh pasien yang berkunjung ke puskesmas pada pagi hari. Pada hari pengamatan, terdapat 10 pasien yang datang pada pagi hari dan seluruhnya dilakukan pengukuran suhu tubuh. Data suhu tubuh pasien tersebut adalah (36,5°C; 36,7°C; 36,6°C; 36,8°C; 36,9°C;

36,5°C; 36,7°C; 36,6°C; 36,8°C; 36,7°C). Hitunglah mean populasi (μ) suhu tubuh pasien tersebut.

$$\mu = \frac{36,5 + 36,7 + 36,6 + 36,8 + 36,9 + 36,5 + 36,7 + 36,6 + 36,8 + 36,7}{10}$$

$$\mu = 36,68$$

Jadi, rata-rata (mean populasi) suhu tubuh pasien adalah 36,68°C.

Contoh Mean Sampel Tidak Berkelompok

Seorang perawat ingin mengetahui rata-rata suhu tubuh pasien yang berobat ke puskesmas pada pagi hari. Karena jumlah pasien cukup banyak, perawat hanya mengukur suhu tubuh 4 pasien secara acak. Suhu tubuh keempat pasien tersebut adalah 36,5°C; 36,8°C; 37,0°C; dan 36,7°C. Hitunglah mean sampel ($\bar{x}$) suhu tubuh pasien tersebut.

$$\bar{x} = \frac{36,5 + 36,8 + 37,0 + 36,7}{4}$$

$$\bar{x} = 36,75$$

Berdasarkan contoh kasus di atas, perbedaan mean sampel dan mean populasi terletak pada cakupan data yang digunakan. Mean populasi dihitung dari seluruh data suhu tubuh pasien yang berkunjung ke puskesmas pada pagi hari, sehingga nilai yang diperoleh menggambarkan kondisi sebenarnya dari populasi tersebut. Sebaliknya, mean sampel dihitung hanya dari sebagian data pasien yang dipilih sebagai sampel, sehingga nilainya digunakan untuk mengestimasi rata-rata suhu tubuh seluruh pasien.

Contoh Mean Sampel Berkelompok

Seorang bidan ingin mengetahui rata-rata jumlah tablet zat besi (Fe) yang dikonsumsi ibu hamil selama satu minggu. Data yang diperoleh dari 10 ibu hamil disajikan sebagai berikut:

Jumlah Tablet Fe yang diminum	Frekuensi (f)
5	2
6	4
7	3
8	1
Total	10

Hitunglah mean ($\bar{x}$) berat badan balita tersebut!

$$\bar{x} = \frac{(5 \times 2) + (6 \times 4) + (7 \times 3) + (8 \times 1)}{10}$$
$$\bar{x} = \frac{63}{10} = 6,3 \cong 6$$

Sehingga, rata-rata jumlah tablet Fe yang dikonsumsi ibu hamil adalah 6 tablet per minggu.

Mean memiliki beberapa kelebihan yakni lebih mudah dipahami dan mudah dihitung, sehingga sering digunakan dalam analisis data. Selain itu, mean juga bersifat unik yaitu hanya memiliki satu nilai untuk setiap kumpulan data. Perhitungan mean melibatkan seluruh nilai data, sehingga setiap data ikut berperan dalam menentukan nilai rata-rata yang diperoleh (Ali et al., 2019).

Akan tetapi, *mean* memiliki kelemahan utama yaitu tidak dapat digunakan pada data kualitatif, karena penggunaan mean

hanya dapat dilakukan pada data kuantitatif. Selain itu, mean sangat peka terhadap nilai ekstrem atau outlier. Kehadiran satu atau beberapa nilai yang sangat tinggi atau sangat rendah dapat menggeser nilai *mean* secara signifikan, sehingga tidak lagi mencerminkan kondisi mayoritas data (Triola, 2020).

Contoh kasus pada mean yang dipengaruhi nilai outlier, seseorang ingin mengukur rata-rata lama hari rawat inap 6 pasien di sebuah bangsal. Lima pasien menjalani perawatan selama 3–4 hari, sedangkan satu pasien harus dirawat selama 20 hari. Meskipun sebagian besar pasien dirawat dalam waktu relatif singkat, keberadaan satu pasien dengan lama rawat inap yang jauh lebih panjang menyebabkan nilai mean meningkat secara signifikan dan tidak lagi mewakili kondisi sebagian besar pasien.

4.2.2 Median sebagai Nilai Tengah

Median merupakan nilai yang berada pada posisi tengah setelah seluruh data diurutkan dari nilai terkecil ke terbesar atau sebaliknya. Nilai median membagi distribusi frekuensi menjadi dua bagian yang sama, di mana 50% data berada pada atau di bawah median, dan 50% lainnya berada di atas median. Oleh karena itu, median sering disebut sebagai persentil ke-50 dan dikenal pula sebagai rata-rata posisi (positional average) (Islam & Al-Shiha, 2018; Manikandan, 2011; Rosner, 2016).

Cara menghitung median dapat dilakukan dengan mengikuti langkah-langkah sebagai berikut:

1. Data sampel terlebih dahulu diurutkan untuk memperoleh susunan data terurut ($y_1, y_2, y_3, \dots, y_n$).

2. Tentukan nilai mediannya dengan rumus

$$\text{Median} = \frac{n + 1}{2}$$

3. Jika jumlah data genap, maka tidak terdapat satu nilai yang tepat berada di tengah. Oleh karena itu, terdapat dua nilai yang berada di posisi tengah, yaitu nilai pada urutan ke- m dan ke- $(m+1)$ setelah data diurutkan. Untuk memperoleh satu nilai yang mewakili posisi tengah tersebut, median ditentukan dengan menghitung rata-rata dari kedua nilai tengah tersebut

$$\text{Median} = \frac{y_m + y_{m+1}}{2}$$

Contoh Kasus Penggunaan Median

1. Seorang guru ingin mengetahui nilai median dari hasil ujian lima siswa. Nilai ujian kelima siswa tersebut adalah sebagai berikut: 70, 85, 60, 90, dan 75. Tentukan nilai mediannya

Penyelesaian:

- a. Data diurut dari nilai terkecil ke terbesar: 60, 70, 75, 85, 90
- b. Tentukan nilai median:

$$\text{Median} = \frac{5 + 1}{2} = 3$$

- c. Nilai median yang ditemukan adalah 3, sehingga nilai median dari hasil ujian pada 5 siswa adalah 75

2. Seorang peneliti ingin mengetahui nilai median dari tinggi badan enam ibu hamil yang sedang melakukan kunjungan antenatal care (ANC). Data tinggi badan keenam ibu hamil tersebut 58

adalah sebagai berikut: 150 cm, 155 cm, 160 cm, 158 cm, 162 cm, dan 165 cm. Tentukan nilai mediannya

- a. Data diurutkan dari nilai terkecil ke terbesar: 150, 155, 158, 160, 162, 165
- b. Jumlah data adalah 6 (genap), sehingga median diperoleh dari rata-rata dua nilai tengah, yaitu data pada urutan ke-3 dan ke-4
- c. Tentukan nilai median:

$$\text{Median} = \frac{158 + 160}{2} = 159$$

- d. Median tinggi badan enam ibu hamil tersebut adalah 159 cm

Median memiliki beberapa kelebihan. Pertama, median bersifat sederhana, sehingga mudah dipahami dan mudah dihitung. Kedua, median bersifat unik, artinya hanya terdapat satu nilai median untuk setiap kumpulan data. Ketiga, median tidak terlalu dipengaruhi oleh nilai ekstrem, berbeda dengan mean yang sangat sensitif terhadap nilai yang sangat tinggi atau sangat rendah (Islam & Al-Shiha, 2018).

Jika data memiliki sebaran yang tidak simetris atau mengandung outlier, *median* sering kali lebih tepat digunakan dibandingkan *mean*. Oleh karena itu, *median* banyak dimanfaatkan dalam analisis data sosial, ekonomi, dan kesehatan yang cenderung memiliki distribusi miring. Nilai *median* memberikan gambaran posisi sentral yang lebih stabil dan realistis terhadap kondisi data secara umum (Islam & Al-Shiha, 2018).

Namun demikian, median juga memiliki beberapa keterbatasan. Median tidak mempertimbangkan seluruh nilai data secara numerik setelah data diurutkan, karena penentuannya hanya didasarkan pada posisi data. Selain itu, median pada umumnya hanya dapat digunakan untuk data kuantitatif.

4.2.3 Modus sebagai Nilai yang Paling Sering Muncul

Mode atau Modus adalah nilai yang paling sering muncul atau memiliki frekuensi tertinggi dalam suatu kumpulan data. Apabila seluruh nilai dalam data berbeda satu sama lain atau masing-masing memiliki frekuensi yang sama, maka data tersebut tidak memiliki modus. Selain itu, suatu kumpulan data dapat memiliki lebih dari satu modus, tergantung pada distribusi frekuensi nilainya (Ali et al., 2019; Hastono, 2024).

Contoh Kasus Penggunaan Modus

Seorang perawat ingin mengetahui berat badan balita yang paling sering muncul pada kegiatan penimbangan di Posyandu. Data berat badan 10 balita yang ditimbang adalah sebagai berikut:

Berat Badan Balita (Kg)	Frekuensi (f)
9	2
10	5
11	1
12	1
13	1

Berat Badan Balita (Kg)	Frekuensi (f)
Total	10

Berdasarkan data diatas berat badan 10 kg muncul paling banyak, sehingga modus dari data berat badan balita tersebut adalah 10 kg.

Modus memiliki beberapa kelebihan. Pertama, modus bersifat sederhana sehingga mudah dipahami dan mudah ditentukan. Kedua, nilai modus tidak terlalu dipengaruhi oleh adanya nilai yang sangat tinggi atau sangat rendah, berbeda dengan mean yang sensitif terhadap nilai ekstrem. Ketiga, modus dapat digunakan baik pada data kuantitatif (berupa angka) maupun data kualitatif (berupa kategori), sehingga penggunaannya cukup luas (Islam & Al-Shiha, 2018; Manikandan, 2011).

Namun modus juga memiliki beberapa keterbatasan. Pertama, modus kurang baik digunakan sebagai ukuran pemusatan data karena hanya bergantung pada beberapa nilai yang paling sering muncul. Kedua, perhitungan modus tidak mempertimbangkan seluruh nilai dalam kumpulan data. Ketiga, dalam beberapa kumpulan data, modus bisa saja tidak ada apabila semua nilai memiliki frekuensi yang sama. Keempat, suatu kumpulan data juga dapat memiliki lebih dari satu modus apabila terdapat lebih dari satu nilai dengan frekuensi tertinggi (De Veaux et al., 2021) .

4.3 Konsep Ukuran Penyebaran Data

Statistik merupakan ilmu yang mempelajari variasi dalam data. Apabila seluruh data bersifat seragam atau homogen, misalnya semua pasien sembuh dalam waktu yang sama yaitu 6 hari, maka analisis statistik menjadi tidak diperlukan. Permasalahan statistik justru muncul ketika terdapat variasi, seperti adanya pasien yang sembuh dalam 2 hari dan sebagian lainnya memerlukan hingga 10 hari. Pada konteks permasalahan tersebut, ukuran dispersi berperan sebagai indikator kuantitatif yang menggambarkan sejauh mana data menyebar dari nilai pusatnya, baik menggunakan mean maupun median. Ukuran dispersi memberikan informasi mengenai tingkat homogenitas data, yaitu seberapa mirip data satu dengan yang lain, serta menunjukkan reliabilitas nilai rata-rata. Nilai mean hanya dapat dikatakan representatif apabila penyebaran data relatif kecil, sebaliknya jika nilai mean berasal dari data dengan penyebaran yang sangat luas cenderung kurang mencerminkan kondisi populasi secara akurat.

4.3.1 Peran Ukuran Penyebaran dalam Analisis Kesehatan Masyarakat

Ukuran penyebaran dalam analisis kesehatan masyarakat berfungsi untuk menggambarkan ketimpangan atau variasi kondisi kesehatan antarindividu dalam suatu populasi. Penyebaran data yang sempit menunjukkan bahwa sebagian besar individu memiliki kondisi kesehatan yang relatif seragam, sedangkan penyebaran yang luas mengindikasikan adanya perbedaan yang mencolok. Informasi

ini penting dalam perencanaan intervensi kesehatan karena populasi dengan variasi tinggi sering kali membutuhkan pendekatan yang lebih spesifik dan tersegmentasi. Dengan demikian, ukuran penyebaran membantu pengambil keputusan memahami kompleksitas kondisi kesehatan secara lebih komprehensif, bukan hanya berdasarkan nilai rata-rata.

4.3.2 Interpretasi Penyebaran Data untuk Penilaian Populasi

Interpretasi ukuran penyebaran data harus selalu dikaitkan dengan konteks populasi yang dianalisis. Penyebaran yang besar tidak selalu bermakna negatif, tetapi dapat mencerminkan keberagaman karakteristik individu dalam suatu wilayah. Namun, dalam kasus tertentu, seperti variasi indeks massa tubuh atau kadar hemoglobin yang terlalu lebar, kondisi ini dapat menjadi indikator adanya masalah gizi atau ketimpangan akses layanan kesehatan. Oleh karena itu, ukuran penyebaran tidak hanya berfungsi sebagai alat statistik, tetapi juga sebagai dasar interpretatif untuk memahami kondisi kesehatan populasi dan merumuskan kebijakan yang lebih tepat sasaran.

4.4 Jenis-Jenis Ukuran Penyebaran

Ukuran penyebaran merupakan komponen penting dalam statistika deskriptif karena berfungsi untuk menggambarkan tingkat variasi atau keragaman data di sekitar nilai pemusatan. Dalam analisis biostatistik, ukuran penyebaran membantu memahami seberapa jauh data menyebar, sehingga interpretasi terhadap nilai

rata-rata menjadi lebih bermakna. Beberapa ukuran penyebaran yang umum digunakan meliputi rentang, varians, dan simpangan baku, koefisien variasi, kuartil dan interquartile (IQR) serta *box and whisker plot* (Islam & Al-Shiha, 2018; Rosner, 2016). Masing-masing ukuran memiliki karakteristik dan kegunaan yang berbeda sesuai dengan tujuan analisis data.

4.4.1 Rentang

Rentang merupakan ukuran penyebaran yang paling sederhana, yaitu selisih antara nilai maksimum dan nilai minimum dalam suatu kumpulan data (Islam & Al-Shiha, 2018). Ukuran ini memberikan gambaran awal mengenai luas penyebaran data, namun penggunaan rentang memiliki kelemahan yakni sangat sensitif dengan nilai ekstrem (*outlier*) sehingga kurang stabil untuk analisis yang mendalam. Interpretasi dalam penggunaan rentang sangat sederhana, semakin besar nilai rentang, semakin lebar sebaran data, yang menandakan variasi antarobservasi semakin tinggi. Berikut rumus dari penggunaan rentang (*range*):

$$Range (R) = Nilai_{maksimum} - Nilai_{minimum}$$

Contoh Kasus Penggunaan Rentang

Seorang petugas kesehatan ingin mengetahui variasi waktu penyembuhan pasien dari penyakit tertentu di sebuah rumah sakit. Dari 5 pasien, waktu penyembuhan yang tercatat adalah 3, 4, 5, 7, dan 10 hari.

Penyelesaian:

Berdasarkan kasus diatas nilai maksimum adalah 10 hari, sedangkan nilai minimum adalah 3 hari, sehingga rentangnya:

$$Range (R) = 10 - 3 = 7 \text{ Hari}$$

Interpretasi dari perhitungan di atas adalah bahwa waktu penyembuhan pasien di rumah sakit tersebut tersebar dalam kisaran 7 hari.

4.4.2 Varians sebagai Dasar Pengukuran Variabilitas

Varians merupakan ukuran penyebaran yang dihitung dari rata-rata kuadrat selisih setiap nilai data terhadap nilai rata-rata. Penggunaan kuadrat bertujuan untuk menghilangkan nilai negatif dan menekankan perbedaan yang lebih besar dari rata-rata (Triola, 2020). Varians banyak digunakan dalam pengembangan metode statistika inferensial, seperti analisis varians dan pemodelan statistik. Namun, kelemahan utama varians terletak pada satuannya yang merupakan kuadrat dari satuan data asli, sehingga kurang intuitif untuk interpretasi praktis, misalnya Kg^2 dan cm^2 . Varians pada konteks biostatistik lebih sering digunakan sebagai dasar perhitungan ukuran penyebaran lain yang lebih mudah diinterpretasikan, yaitu simpangan baku (Triola, 2020). Berikut rumus dari varians:

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

Keterangan

s^2 : varians

x_i : nilai data ke-i

x_i : rata-rata sampel
 n : jumlah data

Contoh Kasus Penggunaan Varians

Seorang peneliti ingin mengetahui variasi panjang badan bayi di sebuah Posyandu. Setelah mengukur panjang badan 5 bayi, didapatkan data (dalam cm): 48, 50, 52, 49, 51

Penyelesaian

1. Hitung rata-rata panjang badan bayi

$$\bar{x} = \frac{48 + 50 + 52 + 49 + 51}{5}$$

$$\bar{x} = 50$$

2. Masukkan ke rumus varians

$$s^2 = \frac{(48 - 50)^2 + (50 - 50)^2 + (52 - 50)^2 + (49 - 50)^2 + (51 - 50)^2}{5 - 1}$$

$$s^2 = \frac{(-2)^2 + (0)^2 + (2)^2 + (-1)^2 + (1)^2}{4}$$

$$s^2 = \frac{10}{4} = 2,5$$

Hasil varians diatas menunjukkan bahwa panjang bayi tidak tersebar terlalu jauh dengan nilai varians 2,5.

4.4.3 Simpangan Baku dalam Analisis Biostatistik

Simpangan baku merupakan akar kuadrat dari varians dan menjadi ukuran penyebaran yang paling sering digunakan dalam analisis biostatistik. Simpangan baku mengukur tingkat penyebaran atau seberapa “terpusat” data terhadap nilai rata-rata (Islam & Al-Shiha, 2018). Semakin kecil nilai simpangan baku, semakin rapat

data berkumpul di sekitar rata-rata; sebaliknya, semakin besar nilai simpangan baku, semakin luas sebaran data dari rata-ratanya. Keunggulan utama simpangan baku terletak pada satuannya yang sama dengan data asli, sehingga lebih mudah dipahami dan diinterpretasikan. Simpangan baku memberikan gambaran seberapa jauh data menyimpang dari nilai rata-rata secara umum.

Simpangan baku pada pengaplikasian biostatistik sering digunakan untuk membandingkan tingkat variasi antar kelompok data serta sebagai dasar dalam penentuan interval kepercayaan dan pengujian hipotesis. Ukuran ini sangat relevan dalam analisis data kesehatan karena mampu menggambarkan heterogenitas karakteristik biologis dan klinis suatu populasi secara lebih akurat (Montgomery & Runger, 2021). Berikut rumus dari simpangan baku:

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

Contoh Kasus Penggunaan Standar Deviasi

Seorang peneliti ingin mengetahui variasi panjang badan bayi di sebuah Posyandu. Setelah mengukur panjang badan 5 bayi, didapatkan data (dalam cm): 48, 50, 52, 49, 51 (contoh soal di varian)

Penyelesaian

1. Hitung rata-rata panjang badan bayi (sama seperti langkah pertama contoh kasus varian)

$$\bar{x} = \frac{48 + 50 + 52 + 49 + 51}{5}$$

$$\bar{x} = 50$$

2. Masukkan ke rumus varians (sama seperti langkah pertama contoh kasus varian)

$$s^2 = \frac{(48 - 50)^2 + (50 - 50)^2 + (52 - 50)^2 + (49 - 50)^2 + (51 - 50)^2}{5 - 1}$$

$$s^2 = \frac{(-2)^2 + (0)^2 + (2)^2 + (-1)^2 + (1)^2}{4}$$

$$s^2 = \frac{10}{4} = 2,5$$

3. Masukkan ke rumus standar deviasi (simpangan baku), yakni akar kuadrat dari rumus varian

$$\sqrt{2,5} = 1,58$$

Berdasarkan hasil simpangan baku diatas ditemukan bahwa persebaran panjang bayi menyebar sekitar $\pm 1,58$ cm

4.4.4 Koefisien Variasi

Penggunaan varians dan standar deviasi bertujuan untuk mengukur sebaran nilai dari satu variabel dalam satu populasi, akan tetapi terkadang peneliti perlu membandingkan sebaran antara dua variabel (Islam & Al-Shiha, 2018). Pada situasi seperti ini, peneliti tidak bisa langsung menggunakan varians atau standar deviasi karena dua alasan: (i) variabel mungkin memiliki satuan yang berbeda, dan (ii) variabel mungkin memiliki nilai rata-rata yang berbeda (Islam & Al-Shiha, 2018). Untuk membandingkan variasi secara adil, peneliti membutuhkan ukuran variasi relatif yang tidak

bergantung pada satuan maupun rata-rata. Ukuran ini disebut koefisien variasi (*Coefficient of Variation (C.V.)*).

Koefisien variasi memberikan cara yang proporsional untuk membandingkan sebaran dua set data yang berbeda, sehingga memudahkan penilaian mengenai variabilitas relatif masing-masing variabel secara adil dan konsisten. Berikut rumus dari koefisien variasi:

$$C.V. = \frac{S}{\bar{x}} \times 100\%$$

Contoh Kasus Penggunaan Koefisien Variasi

Sebuah rumah sakit ingin membandingkan variasi tekanan darah sistolik pasien di dua ruang rawat inap: Ruang Rawat Inap A dan Ruang Rawat Inap B. Berikut data tekanan darah pasien:

Ruang Rawat Inap	Rata-rata tekanan darah (mmHg)	Standar Deviasi
A	135	10
B	147	15

Penyelesaian:

1. Jika data belum memiliki nilai rata-rata (mean), hitunglah nilai mean terlebih dahulu
2. Setelah nilai mean diperoleh, lanjutkan dengan menghitung standar deviasi atau simpangan baku (SD)
3. Masukkan ke rumus koefisien variasi pada tiap kelompok
 - a. Ruang Rawat Inap A

$$C.V. = \frac{10}{135} \times 100\% = 7,41\%$$

b. Ruang Rawat Inap B

$$C.V. = \frac{15}{147} \times 100\% = 10,2\%$$

Berdasarkan hasil diatas ditemukan ruang rawat inap B memiliki variasi tekanan darah lebih tinggi dibandingkan dengan ruang rawat inap A (C.V. 7,41% dibandingkan dengan 10,2%).

4.4.5 Interquartile Range (IQR)

Rentang antar-kuartil (*interquartile range* atau IQR) merupakan ukuran variasi data yang menunjukkan selisih antara kuartil ketiga (Q3) dan kuartil pertama (Q1). Rentang antar-kuartil dihitung dengan mengambil selisih antara kuartil ketiga (Q3) dan kuartil pertama (Q1) (Hastono, 2024; Islam & Al-Shiha, 2018). Rentang antar-kuartil memberikan ukuran yang lebih bermakna karena merepresentasikan rentang 50% data tengah. IQR merupakan langkah yang digunakan untuk mengatasi keterbatasan dari penggunaan nilai rentang (*range*). *Interquartile Range* dapat dirumuskan sebagai berikut:

$$IQR = Q3 - Q1$$

Kuartil pertama (Q1) merupakan nilai di bawah 25% data, artinya 25% data berada di bawah Q1. Sedangkan kuartil ketiga merupakan nilai di bawah 75% data, artinya 75% data berada di bawah Q3. Sehingga hasil IQR menunjukkan rentang 50% data tengah (Q2), yaitu data yang berada di antara Q1 dan Q3. IQR bertujuan untuk melihat data tengah, ukuran ini tidak terpengaruh oleh nilai ekstrem (*outlier*), sehingga lebih representatif untuk

menggambarkan variasi data daripada range biasa (Hastono, 2024; Islam & Al-Shiha, 2018).

Contoh Kasus Penggunaan *Interquartile Range* (IQR)

Seorang peneliti ingin menilai rentang pengetahuan yang dimiliki responden terkait pemberian asi eksklusif, dari 12 responden ditemukan hasil sebagai berikut

55,70,40,60,85,50,90,65,75,45,80,35

Penyelesaian

1. Urutkan data terlebih dahulu

35,40,45,50,55,60,65,70,75,80,85,90

2. Tentukan nilai median (Q2), Q1, dan Q3 dengan total responden sebanyak 12 orang

- a. Nilai Q2 sama seperti penentuan median:

$$Q2 = \frac{60 + 65}{2} = 62,5$$

- b. Nilai Q1(dapat ditentukan berdasarkan nilai 25% bawah) yakni :

$$Q1 = \frac{45 + 50}{2} = 47,5$$

- c. Nilai Q3 (dapat ditentukan berdasarkan nilai 25% atas) yakni:

$$Q3 = \frac{75 + 80}{2} = 77,5$$

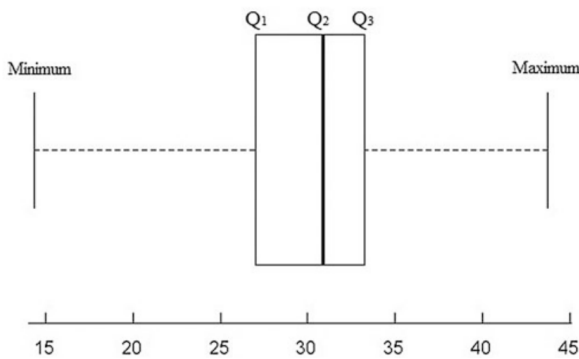
3. Hitung IQR

$$IQR = 77,5 - 47,5 = 30$$

Hasil di atas menunjukkan bahwa 50% responden memiliki skor pengetahuan antara 47,5 hingga 77,5, dengan rentang nilai data tengah (IQR) sebesar 30.

4.4.6 Box-and-Whisker Plot (Box Plot)

Box and whisker plot atau diagram kotak adalah representasi grafis dari ringkasan lima angka: nilai minimum, kuartil pertama (Q1), median (Q2), kuartil ketiga (Q3), dan nilai maksimum. Kotak menunjukkan *interquartile range* (IQR), yaitu rentang 50% data tengah, dengan garis di tengah kotak menandai median. Whiskers memanjang dari kotak hingga nilai minimum dan maksimum (tidak termasuk outlier), sedangkan outlier digambarkan sebagai titik terpisah (Islam & Al-Shiha, 2018).



Gambar 1 Box and Whisker Plot

Sumber: (Islam & Al-Shiha, 2018)

Box plot memungkinkan peneliti melihat pusat, sebaran, simetri, dan outlier secara cepat, serta memudahkan perbandingan

distribusi data antar kelompok. Dengan demikian, box plot menjadi alat visual yang efisien untuk menilai karakteristik utama dari data sekaligus mengidentifikasi nilai ekstrem yang mungkin signifikan.

Latihan Soal (Esai)

1. Jelaskan pengertian ukuran pemusatan data dan manfaatnya dalam kesehatan masyarakat.
2. Bandingkan mean, median, dan modus serta jelaskan kapan masing-masing digunakan.
3. Mengapa ukuran penyebaran penting dalam analisis data kesehatan?
4. Uraikan jenis-jenis ukuran penyebaran data beserta fungsinya.
5. Jelaskan perbedaan dari varians dan simpangan baku

DAFTAR PUSTAKA

- De Veaux, R. D., Velleman, P. F., & Bock, D. E. (2021). *Intro stats* (5th ed.). Pearson Education.
- Montgomery, D. C., & Runger, G. C. (2021). *Applied statistics and probability for engineers* (7th ed.). Wiley.
- Rosner, B. (2023). *Fundamentals of biostatistics*. Cengage Learning.
- Triola, M. F. (2020). *Elementary statistics* (13th ed.). Pearson Education.
- Ali, Z., Bhaskar, Sb., & Sudheesh, K. (2019). Descriptive statistics: Measures of central tendency, dispersion, correlation and regression. *Airway*, 2(3), 120. https://doi.org/10.4103/ARWY.ARWY_37_19
- Hastono, S. P. (2024). Analisis data pada bidang kesehatan (IV, Vol. 1). Rajagrafindo Persada.
- Islam, M. A., & Al-Shiha, A. (2018). Foundations of biostatistics. *Foundations of Biostatistics*, 1–459. <https://doi.org/10.1007/978-981-10-8627-4/COVER>
- Manikandan, S. (2011). Measures of central tendency: Median and mode. *Journal of Pharmacology & Pharmacotherapeutics*, 2(3), 214–215. <https://doi.org/10.4103/0976-500X.83300>
- Rosner, B. (2016). *Fundamentals of Biostatistics* (8th ed.). Cengage Learning.

BAB 5

DISTRIBUSI DATA DAN INTERPRETASI HASIL

Fitri Eka Sari S, M.Kes.

5.1 Konsep Distribusi Data

5.1.1 Pengertian Distribusi Data dalam Penelitian

Distribusi data adalah cara bagaimana nilai-nilai dalam suatu dataset tersebar di atas suatu rentang nilai tertentu dalam penelitian kuantitatif. Distribusi data merupakan aspek fundamental dalam analisis statistik yang menggambarkan pola penyebaran data terhadap nilai-nilai tertentu dan memengaruhi pemilihan metode statistik yang tepat untuk analisis lebih lanjut (Sadeesh et al., 2026). Pemahaman yang mendalam tentang distribusi data sangat penting karena banyak prosedur statistik memiliki asumsi tertentu tentang bagaimana data seharusnya terdistribusi sebelum dapat diterapkan secara valid. Distribusi data dapat berbentuk normal, skew, atau bentuk distribusi lainnya yang masing-masing memiliki karakteristik dan implikasi berbeda untuk analisis statistik.

Selain itu, distribusi data juga berkaitan erat dengan konsep frekuensi, yaitu seberapa sering suatu nilai muncul dalam suatu

dataset. Penyajian distribusi data biasanya dilakukan melalui tabel distribusi frekuensi, histogram, polygon frekuensi, maupun grafik kurva distribusi. Melalui penyajian ini, peneliti dapat dengan mudah mengidentifikasi pola umum data, seperti apakah data cenderung mengelompok pada nilai tertentu, menyebar secara merata, atau memiliki nilai ekstrem (outlier). Keberadaan outlier dalam distribusi data juga menjadi perhatian penting karena dapat memengaruhi nilai statistik seperti rata-rata dan simpangan baku, sehingga berpotensi menimbulkan bias dalam interpretasi hasil penelitian (Triola, 2018).

Lebih lanjut, distribusi data juga berperan dalam menentukan jenis analisis statistik yang akan digunakan, apakah menggunakan statistik parametrik atau nonparametrik. Statistik parametrik umumnya mensyaratkan data berdistribusi normal, sedangkan statistik nonparametrik lebih fleksibel karena tidak bergantung pada asumsi distribusi tertentu. Oleh karena itu, sebelum melakukan analisis lanjutan, peneliti biasanya melakukan uji normalitas seperti uji Kolmogorov-Smirnov atau Shapiro-Wilk untuk memastikan kesesuaian distribusi data dengan metode analisis yang akan digunakan (Ghasemi & Zahediasl, 2012).

Dengan demikian, pemahaman terhadap distribusi data tidak hanya membantu dalam menggambarkan karakteristik data secara deskriptif, tetapi juga menjadi dasar penting dalam pengambilan keputusan metodologis dalam penelitian kuantitatif. Hal ini memastikan bahwa hasil analisis yang diperoleh valid, reliabel, dan dapat dipertanggungjawabkan secara ilmiah.

5.1.2 Jenis-Jenis Distribusi Data

Ada beberapa jenis distribusi data yang sering dijumpai dalam penelitian, dengan distribusi normal atau *normal distribution* merupakan yang paling penting dan sering diasumsikan dalam banyak prosedur statistik parametrik. Distribusi normal memiliki bentuk lonceng atau *bell curve* yang simetris di mana rata-rata (*mean*), median, dan modus berada pada titik yang sama di tengah distribusi (Mishra et al., 2019). Distribusi normal ditandai oleh parameter mean dan standar deviasi yang menentukan lokasi pusat dan sebaran data. Selain distribusi normal, ada juga distribusi yang miring atau *skewed distribution* di mana data tidak simetris dan memiliki ekor yang panjang pada satu sisi. Distribusi yang miring ke kanan (*positively skewed*) memiliki ekor panjang ke arah nilai-nilai positif yang tinggi, sedangkan distribusi yang miring ke kiri (*negatively skewed*) memiliki ekor panjang ke arah nilai-nilai negatif yang rendah (Ruta et al., 2025). Distribusi normal dan distribusi menceng (*skewed*), terdapat beberapa jenis distribusi lain yang juga penting dalam analisis data, seperti distribusi uniform dan distribusi bimodal. Distribusi uniform adalah distribusi di mana semua nilai memiliki probabilitas yang relatif sama untuk muncul, sehingga tidak terdapat puncak yang dominan dalam distribusi tersebut. Sementara itu, distribusi bimodal memiliki dua puncak (*modus*) yang menunjukkan adanya dua kelompok atau subpopulasi dalam dataset, yang dapat mengindikasikan adanya perbedaan karakteristik dalam data yang diteliti (Walpole et al., 2012). Selain itu, terdapat

pula distribusi leptokurtik dan platikurtik yang menggambarkan tingkat keruncingan (kurtosis) suatu distribusi dibandingkan dengan distribusi normal. Distribusi leptokurtik memiliki puncak yang lebih tajam dan ekor yang lebih tebal, yang menunjukkan adanya kemungkinan nilai ekstrem yang lebih besar. Sebaliknya, distribusi platikurtik memiliki puncak yang lebih datar dan sebaran yang lebih merata. Analisis kurtosis ini membantu peneliti dalam memahami karakteristik penyebaran data secara lebih mendalam (Field, 2013).

5.1.3 Ukuran Kemiringan dan Keruncingan Distribusi

Skewness atau kemiringan adalah ukuran statistik yang menggambarkan asimetri dari distribusi data di sekitar mean. Nilai skewness yang sama dengan nol menunjukkan distribusi yang simetris sempurna atau normal, nilai skewness positif menunjukkan distribusi yang miring ke kanan dengan ekor panjang ke arah nilai-nilai yang tinggi, dan nilai skewness negatif menunjukkan distribusi yang miring ke kiri dengan ekor panjang ke arah nilai-nilai yang rendah (Sadeesh et al., 2026). Kurtosis adalah ukuran yang menggambarkan keruncingan atau ketajaman dari puncak distribusi data dibandingkan dengan distribusi normal. Distribusi dengan kurtosis tinggi (*leptokurtic*) memiliki puncak yang lebih tajam dan ekor yang lebih tebal dibanding distribusi normal, sedangkan distribusi dengan kurtosis rendah (*platykurtic*) memiliki puncak yang lebih datar dan ekor yang lebih tipis (Mishra et al., 2019). Interpretasi nilai skewness dan kurtosis juga memerlukan pemahaman terhadap batasan nilai yang dianggap masih mendekati

distribusi normal. Dalam praktik penelitian, nilai skewness dan kurtosis yang berada dalam rentang tertentu (misalnya antara -1 hingga +1 atau -2 hingga +2, tergantung pada literatur yang digunakan) sering dianggap masih dapat diterima sebagai indikasi bahwa data tidak menyimpang secara signifikan dari distribusi normal (George & Mallery, 2010). Oleh karena itu, peneliti perlu berhati-hati dalam menafsirkan nilai tersebut dan tidak hanya bergantung pada satu indikator saja. Nilai skewness dan kurtosis biasanya dianalisis bersamaan dengan uji normalitas dan visualisasi data seperti histogram atau Q-Q plot untuk memperoleh gambaran yang lebih komprehensif tentang distribusi data. Kombinasi pendekatan ini membantu peneliti dalam memastikan apakah asumsi normalitas terpenuhi atau tidak, terutama sebelum melakukan analisis statistik parametrik (Field, 2013).

5.1.4 Pentingnya Uji Normalitas dalam Penelitian

Uji normalitas adalah prosedur statistik yang digunakan untuk menentukan apakah data mengikuti distribusi normal atau tidak, yang merupakan asumsi penting bagi banyak prosedur statistik parametrik. Uji normalitas dapat dilakukan menggunakan berbagai metode baik metode numerik seperti *Shapiro-Wilk test*, *Kolmogorov-Smirnov test*, dan *Anderson-Darling test*, maupun metode visual seperti *histogram*, *Q-Q plot*, dan *box plot* (Mishra et al., 2019). Hasil uji normalitas akan menentukan apakah peneliti dapat menggunakan prosedur statistik parametrik yang mengasumsikan normalitas data atau harus menggunakan prosedur

non-parametrik yang tidak memerlukan asumsi normalitas. Pemahaman tentang hasil uji normalitas sangat penting untuk memastikan validitas kesimpulan statistik yang dihasilkan dari penelitian.

Selain metode yang telah disebutkan, penting untuk dipahami bahwa uji normalitas tidak hanya bergantung pada nilai statistik semata, tetapi juga perlu dikombinasikan dengan interpretasi visual. Dalam praktiknya, uji seperti Shapiro-Wilk dan Kolmogorov-Smirnov sangat sensitif terhadap ukuran sampel. Pada sampel yang besar, penyimpangan kecil dari distribusi normal dapat menghasilkan nilai p-value yang signifikan, sehingga data dianggap tidak normal, meskipun secara visual masih mendekati distribusi normal. Sebaliknya, pada sampel kecil, uji normalitas mungkin tidak cukup kuat untuk mendeteksi ketidaknormalan data (Ghasemi & Zahediasl, 2012).

Oleh karena itu, peneliti disarankan untuk tidak hanya bergantung pada satu jenis uji, melainkan mengombinasikan pendekatan numerik dan visual seperti histogram dan Q-Q plot untuk memperoleh gambaran yang lebih komprehensif mengenai distribusi data. Selain itu, jika data tidak berdistribusi normal, terdapat beberapa alternatif yang dapat dilakukan, seperti transformasi data (logaritma, akar kuadrat) atau penggunaan metode statistik non-parametrik seperti uji Mann-Whitney atau Kruskal-Wallis (Field, 2013).

Lebih lanjut, dalam banyak penelitian modern, asumsi normalitas juga sering dipertimbangkan secara fleksibel, terutama ketika ukuran sampel cukup besar, karena berdasarkan Teorema Limit Tengah (Central Limit Theorem), distribusi sampling dari rata-rata cenderung mendekati normal. Dengan demikian, pemahaman yang tepat mengenai uji normalitas akan membantu peneliti dalam memilih metode analisis yang sesuai serta meningkatkan keakuratan dan validitas hasil penelitian yang diperoleh.

5.2 Analisis Deskriptif Data

5.2.1 Pengertian dan Tujuan Statistik Deskriptif

Statistik deskriptif adalah metode statistik yang digunakan untuk meringkas dan menggambarkan karakteristik utama dari data dalam sampel penelitian dengan cara yang singkat dan mudah dipahami. Tujuan utama statistik deskriptif adalah memberikan gambaran yang jelas tentang distribusi dan pola data, mengidentifikasi nilai-nilai yang mewakili data, mengukur variabilitas data, dan mengidentifikasi pencilan atau *outliers* (Kotronoulas et al., 2023). Statistik deskriptif menghasilkan ringkasan numerik dan visual dari data yang memudahkan peneliti dan pembaca untuk memahami karakteristik utama dari sampel penelitian tanpa melakukan inferensi statistik yang lebih kompleks.

5.2.2 Ukuran Pemusatan Data

Ukuran pemusatan atau *measures of central tendency* adalah nilai-nilai yang mewakili pusat atau lokasi data dalam distribusi, yang paling umum digunakan adalah mean, median, dan modus. Mean atau rata-rata adalah jumlah semua nilai data dibagi dengan jumlah pengamatan, yang merupakan ukuran pemusatan yang paling sering digunakan dalam penelitian (Kotronoulas et al., 2023). Median adalah nilai tengah dari data yang telah diurutkan, dan lebih tahan terhadap pengaruh nilai-nilai ekstrem atau pencilan dibanding mean. Modus adalah nilai yang paling sering muncul dalam dataset dan dapat digunakan untuk data kategorikal maupun numerik. Pemilihan ukuran pemusatan yang tepat tergantung pada jenis data dan bentuk distribusi data. Ukuran Pemusatan data mean, median, dan modus, pemilihan ukuran pemusatan juga harus mempertimbangkan karakteristik distribusi data, khususnya apakah data berdistribusi normal atau tidak. Pada distribusi normal, nilai mean, median, dan modus cenderung berada pada posisi yang sama, sehingga mean sering digunakan sebagai representasi pusat data. Namun, pada distribusi yang menceng (*skewed*), mean dapat terpengaruh oleh nilai ekstrem sehingga tidak lagi merepresentasikan pusat data secara akurat. Dalam kondisi tersebut, median lebih disarankan karena lebih stabil terhadap pengaruh pencilan (Triola, 2018). Dalam beberapa analisis statistik, terutama pada data rasio atau interval, mean sering digunakan karena memiliki sifat matematis yang memungkinkan untuk dilakukan

operasi lanjutan seperti perhitungan variansi dan standar deviasi. Sementara itu, median lebih tepat digunakan pada data ordinal atau data numerik yang tidak memenuhi asumsi normalitas. Modus juga memiliki peran penting, khususnya dalam data kategorikal, karena dapat menunjukkan kategori yang paling dominan dalam suatu distribusi (Gravetter & Wallnau, 2014).

5.2.3 Ukuran Penyebaran Data

Ukuran penyebaran atau *measures of spread* adalah nilai-nilai yang menggambarkan variabilitas atau sebaran data di sekitar nilai pusat, yang paling umum digunakan adalah range, variansi, dan standar deviasi. *Range* adalah perbedaan antara nilai maksimum dan nilai minimum dalam dataset, yang memberikan gambaran sederhana tentang luas penyebaran data (Kotronoulas et al., 2023). Variansi adalah rata-rata dari kuadrat deviasi setiap nilai data dari mean, yang mengukur variabilitas data namun dalam unit kuadrat sehingga sulit diinterpretasi. Standar deviasi adalah akar kuadrat dari variansi dan memiliki unit yang sama dengan data asli, sehingga lebih mudah diinterpretasi dan lebih sering digunakan dalam penelitian. Ukuran penyebaran ini sangat penting untuk memahami seberapa jauh data tersebar dari pusat distribusi.

5.2.4 Penyajian Data Secara Visual

Penyajian data secara visual melalui grafik dan diagram adalah cara efektif untuk mengkomunikasikan pola dan karakteristik data kepada pembaca. Histogram adalah diagram batang yang menunjukkan frekuensi atau proporsi data dalam interval-interval

kelas yang berbeda, dan sangat berguna untuk menggambarkan bentuk distribusi data (Sadeesh et al., 2026). *Box plot* atau diagram kotak adalah grafik yang menampilkan median, kuartil, dan pencilan dalam dataset, dan berguna untuk membandingkan distribusi data antar kelompok. *Scatter plot* atau diagram pencar menunjukkan hubungan antara dua variabel numerik dan berguna untuk identifikasi pola korelasi atau asosiasi. Grafik garis (*line graph*) berguna untuk menampilkan perubahan data dari waktu ke waktu atau tren dalam dataset.

Selain range, variansi, dan standar deviasi, terdapat ukuran penyebaran lain yang juga penting, seperti interquartile range (IQR) dan simpangan rata-rata (mean absolute deviation). IQR merupakan selisih antara kuartil ketiga (Q3) dan kuartil pertama (Q1), yang menggambarkan sebaran 50% data tengah dan relatif tidak dipengaruhi oleh nilai pencilan (outlier). Oleh karena itu, IQR sering digunakan pada data yang tidak berdistribusi normal atau memiliki outlier yang ekstrem. Sementara itu, simpangan rata-rata mengukur rata-rata jarak absolut setiap data terhadap nilai mean, yang memberikan gambaran variasi data dengan pendekatan yang lebih sederhana (Triola, 2018). Pemilihan ukuran penyebaran yang tepat sangat bergantung pada karakteristik distribusi data dan tujuan analisis yang dilakukan oleh peneliti.

Dalam konteks penyajian data secara visual, pemilihan jenis grafik harus disesuaikan dengan jenis data dan tujuan penyajian agar informasi dapat tersampaikan secara efektif. Selain grafik yang telah

disebutkan, diagram batang (bar chart) juga sering digunakan untuk membandingkan nilai antar kategori, sedangkan diagram lingkaran (pie chart) digunakan untuk menunjukkan proporsi atau persentase suatu kategori terhadap keseluruhan data. Visualisasi data yang baik tidak hanya menampilkan informasi, tetapi juga harus memperhatikan aspek kejelasan, kesederhanaan, dan akurasi agar tidak menimbulkan misinterpretasi (Knafllic, 2015).

5.3 Analisis Inferensial Data

5.3.1 Pengertian dan Tujuan Statistik Inferensial

Statistik inferensial adalah metode statistik yang digunakan untuk membuat kesimpulan atau inferensi tentang populasi berdasarkan informasi dari sampel penelitian yang telah dianalisis. Tujuan utama statistik inferensial adalah menguji hipotesis tentang parameter populasi, memperkirakan parameter populasi berdasarkan data sampel, dan menentukan apakah perbedaan atau hubungan yang diamati dalam sampel juga berlaku untuk populasi secara keseluruhan (Ruta et al., 2025). Statistik inferensial sangat penting dalam penelitian karena memungkinkan peneliti untuk membuat generalisasi yang valid dari sampel ke populasi dengan mempertimbangkan ketidakpastian dan variabilitas sampling. Selain tujuan utama tersebut, statistik inferensial juga mencakup dua pendekatan utama, yaitu estimasi parameter dan pengujian hipotesis. Estimasi parameter dilakukan untuk memperkirakan nilai parameter populasi, baik dalam bentuk estimasi titik (*point estimate*) maupun

estimasi interval (*confidence interval*). Estimasi interval memberikan rentang nilai yang mungkin mengandung parameter populasi dengan tingkat kepercayaan tertentu, sehingga dianggap lebih informatif dibandingkan estimasi titik semata (Walpole et al., 2012). Di sisi lain, pengujian hipotesis digunakan untuk menilai kebenaran suatu dugaan atau asumsi tentang populasi berdasarkan data sampel. Dalam proses ini, peneliti menggunakan berbagai jenis uji statistik seperti uji *t*, uji *chi-square*, atau analisis varians (ANOVA), tergantung pada jenis data dan tujuan penelitian. Pemilihan metode inferensial yang tepat sangat bergantung pada skala pengukuran data, jumlah kelompok yang dibandingkan, serta terpenuhi atau tidaknya asumsi-asumsi statistik seperti normalitas dan homogenitas varians (Field, 2013). Statistik inferensial juga mempertimbangkan konsep kesalahan sampling (*sampling error*), yaitu perbedaan antara karakteristik sampel dan populasi yang terjadi secara alami karena penggunaan sebagian data saja. Oleh karena itu, hasil analisis inferensial selalu disertai dengan tingkat kepercayaan atau probabilitas tertentu untuk menunjukkan tingkat ketidakpastian dalam pengambilan keputusan.

5.3.2 Pengujian Hipotesis dan Signifikansi Statistik

Pengujian hipotesis adalah prosedur statistik yang digunakan untuk menentukan apakah hipotesis nol (*null hypothesis*) harus ditolak atau tidak berdasarkan data sampel. Hipotesis nol umumnya menyatakan bahwa tidak ada perbedaan atau hubungan antara variabel-variabel yang diteliti, sedangkan hipotesis alternatif

menyatakan bahwa ada perbedaan atau hubungan (Kotronoulas et al., 2023). Tingkat signifikansi atau *level of significance* yang umum digunakan adalah 0,05 atau 5 persen, yang berarti probabilitas menolak hipotesis nol padahal hipotesis nol benar adalah 5 persen. Nilai P atau *p-value* yang dihasilkan dari pengujian hipotesis menunjukkan probabilitas memperoleh hasil yang sama atau lebih ekstrem dari hasil yang diamati jika hipotesis nol benar, dan jika nilai P lebih kecil dari tingkat signifikansi maka hipotesis nol ditolak.

Selain konsep dasar tersebut, pengujian hipotesis juga melibatkan dua jenis kesalahan yang perlu diperhatikan, yaitu kesalahan tipe I dan kesalahan tipe II. Kesalahan tipe I terjadi ketika peneliti menolak hipotesis nol padahal sebenarnya benar, yang besarnya sama dengan tingkat signifikansi (α), misalnya 0,05. Sementara itu, kesalahan tipe II terjadi ketika peneliti gagal menolak hipotesis nol padahal sebenarnya salah, yang berkaitan dengan kekuatan uji (statistical power) (Wasserstein & Lazar, 2016). Oleh karena itu, dalam penelitian kuantitatif, penting untuk mempertimbangkan ukuran sampel yang memadai agar dapat meningkatkan kekuatan uji dan mengurangi kemungkinan kesalahan dalam pengambilan keputusan statistik. Selain nilai *p-value*, interpretasi hasil uji hipotesis juga sebaiknya dilengkapi dengan ukuran efek (effect size) dan interval kepercayaan (confidence interval). Ukuran efek memberikan informasi mengenai besar kecilnya pengaruh atau hubungan antar variabel, sehingga tidak hanya berfokus pada signifikansi statistik tetapi juga signifikansi

praktis. Sementara itu, interval kepercayaan memberikan rentang nilai yang diyakini mengandung parameter populasi dengan tingkat kepercayaan tertentu, biasanya 95% (Cumming, 2014).

5.3.3 Uji Statistik Parametrik dan Non-Parametrik

Uji statistik parametrik adalah uji yang mengasumsikan bahwa data mengikuti distribusi tertentu, biasanya distribusi normal, dan memiliki asumsi-asumsi lain seperti homogenitas variansi. Beberapa uji parametrik yang sering digunakan termasuk *independent samples t-test* untuk membandingkan mean antara dua kelompok independen, *paired t-test* untuk membandingkan mean sebelum dan sesudah intervensi pada sampel yang sama, dan *one-way ANOVA* untuk membandingkan mean antara lebih dari dua kelompok (Ruta et al., 2025). Uji statistik non-parametrik adalah uji yang tidak mengasumsikan bentuk distribusi tertentu dan digunakan ketika asumsi parametrik tidak terpenuhi atau data bersifat ordinal atau nominal. Uji non-parametrik yang sering digunakan termasuk *Mann-Whitney U test* sebagai alternatif non-parametrik untuk *independent samples t-test*, *Wilcoxon signed-rank test* untuk data berpasangan, dan *Kruskal-Wallis H test* sebagai alternatif non-parametrik untuk *one-way ANOVA*. Jenis-jenis ukuran efek yang telah disebutkan, penting juga untuk memahami interpretasi nilai ukuran efek dalam konteks penelitian. Sebagai contoh, Cohen (1988) memberikan pedoman umum untuk interpretasi *Cohen's d*, yaitu nilai sekitar 0,2 dianggap kecil, 0,5 sedang, dan 0,8 atau lebih dianggap besar. Namun demikian,

interpretasi ini tidak bersifat mutlak karena besarnya ukuran efek juga harus mempertimbangkan konteks bidang ilmu, desain penelitian, serta karakteristik populasi yang diteliti. Dalam penelitian kesehatan atau sosial, ukuran efek yang relatif kecil pun dapat memiliki implikasi penting jika berdampak pada populasi yang luas. Ukuran efek juga memiliki peran penting dalam perencanaan penelitian, khususnya dalam penentuan ukuran sampel (*sample size calculation*). Dengan mengetahui estimasi ukuran efek yang diharapkan, peneliti dapat menghitung jumlah sampel yang dibutuhkan agar penelitian memiliki kekuatan uji (*statistical power*) yang memadai. Hal ini membantu menghindari penelitian dengan sampel terlalu kecil yang berisiko tidak mampu mendeteksi efek yang sebenarnya ada (Sullivan & Feinn, 2012).

5.3.4 Estimasi Parameter dan Interval Kepercayaan

Estimasi parameter adalah proses memperkirakan nilai parameter populasi (*population parameter*) berdasarkan data sampel, dan dapat menggunakan titik estimasi (*point estimate*) yang merupakan nilai tunggal atau estimasi interval (*interval estimate*) yang merupakan rentang nilai. *Confidence interval* atau interval kepercayaan adalah rentang nilai yang diperkirakan mengandung parameter populasi dengan tingkat kepercayaan tertentu, biasanya 95 persen (Kotronoulas et al., 2023). Interval kepercayaan 95 persen berarti bahwa jika penelitian diulang berkali-kali dan interval kepercayaan dihitung untuk setiap pengulangan, maka 95 persen dari interval-interval tersebut akan mengandung parameter populasi yang

sebenarnya. Interval kepercayaan memberikan informasi yang lebih informatif daripada hanya nilai P karena menunjukkan tidak hanya apakah perbedaan signifikan tetapi juga besarnya perbedaan tersebut.

5.4 Interpretasi Hasil Statistik

5.4.1 Pemahaman Nilai P dan Signifikansi Klinis

Nilai P atau *p-value* menunjukkan probabilitas memperoleh hasil statistik yang sama atau lebih ekstrem dari hasil yang diamati jika hipotesis nol benar, namun sering kali nilai P disalahartikan atau diinterpretasi dengan salah dalam penelitian. Nilai P yang kecil seperti 0,01 atau 0,001 menunjukkan bahwa hasil yang diamati sangat tidak mungkin terjadi jika hipotesis nol benar, sehingga memberikan bukti kuat untuk menolak hipotesis nol (Zrelak, 2025). Penting untuk membedakan antara signifikansi statistik (*statistical significance*) yang diindikasikan oleh nilai P kecil dan signifikansi klinis atau praktis (*clinical or practical significance*) yang mengacu pada apakah perbedaan atau efek yang ditemukan memiliki dampak nyata dalam praktik kesehatan atau kehidupan sehari-hari.

5.4.2 Interpretasi Ukuran Efek dan Relevansi Klinis

Ukuran efek atau *effect size* adalah ukuran besarnya perbedaan atau hubungan antara variabel-variabel yang diteliti, dan merupakan informasi yang sangat penting untuk melengkapi nilai P dalam interpretasi hasil penelitian. Ukuran efek dapat diekspresikan dalam berbagai bentuk seperti perbedaan mean yang distandarisasi

(*standardized mean difference* atau Cohen's *d*), rasio odds (*odds ratio*), atau risiko relatif (*relative risk*) (Kotronoulas et al., 2023). Ukuran efek yang besar menunjukkan bahwa perbedaan atau hubungan yang ditemukan mungkin memiliki relevansi klinis yang signifikan, bahkan meskipun nilai *P* kecil. Sebaliknya, ukuran efek yang kecil menunjukkan bahwa meskipun perbedaan atau hubungan tersebut secara statistik signifikan namun mungkin tidak memiliki relevansi klinis yang berarti. Selain jenis-jenis ukuran efek yang telah disebutkan, penting juga untuk memahami interpretasi nilai ukuran efek dalam konteks penelitian. Sebagai contoh, Cohen (1988) memberikan pedoman umum untuk interpretasi *Cohen's d*, yaitu nilai sekitar 0,2 dianggap kecil, 0,5 sedang, dan 0,8 atau lebih dianggap besar. Namun demikian, interpretasi ini tidak bersifat mutlak karena besarnya ukuran efek juga harus mempertimbangkan konteks bidang ilmu, desain penelitian, serta karakteristik populasi yang diteliti. Dalam penelitian kesehatan atau sosial, ukuran efek yang relatif kecil pun dapat memiliki implikasi penting jika berdampak pada populasi yang luas. Efek juga memiliki peran penting dalam perencanaan penelitian, khususnya dalam penentuan ukuran sampel (*sample size calculation*). Dengan mengetahui estimasi ukuran efek yang diharapkan, peneliti dapat menghitung jumlah sampel yang dibutuhkan agar penelitian memiliki kekuatan uji (*statistical power*) yang memadai. Hal ini membantu menghindari penelitian dengan sampel terlalu kecil yang berisiko tidak mampu mendeteksi efek yang sebenarnya ada (Sullivan &

Feinn, 2012). Dengan demikian, ukuran efek tidak hanya berfungsi sebagai pelengkap nilai signifikansi statistik, tetapi juga sebagai indikator penting dalam menilai relevansi praktis dan kekuatan hubungan atau perbedaan dalam penelitian, sehingga sangat penting untuk selalu dilaporkan dan diinterpretasikan secara tepat.

5.4.3 Identifikasi dan Penanganan Pencilan dalam Data

Pencilan atau *outliers* adalah nilai-nilai yang jauh berbeda dari nilai-nilai lainnya dalam dataset dan dapat memengaruhi hasil analisis statistik secara signifikan. Pencilan dapat teridentifikasi melalui metode visual seperti *box plot* atau *scatter plot*, atau melalui metode numerik seperti perhitungan z-score atau interquartile range (Sadeesh et al., 2026). Pencilan dapat disebabkan oleh kesalahan pengukuran, kesalahan pengumpulan data, atau nilai-nilai yang benar-benar ekstrem dari populasi. Penanganan pencilan dapat dilakukan dengan menghilangkan pencilan jika diyakini hasil dari kesalahan, melakukan transformasi data, atau menggunakan metode analisis yang lebih tahan terhadap pencilan seperti median-based atau robust statistical methods. Selain metode identifikasi yang telah disebutkan, penting bagi peneliti untuk melakukan evaluasi kontekstual terhadap pencilan sebelum mengambil keputusan penanganan. Tidak semua pencilan harus dihapus, karena dalam beberapa kasus nilai tersebut justru merepresentasikan fenomena penting atau variasi alami dalam populasi. Oleh karena itu, peneliti perlu menelusuri sumber pencilan, misalnya dengan memeriksa kembali proses pengumpulan data, validitas instrumen, atau

kemungkinan kesalahan input data (entry error). Pendekatan ini membantu memastikan bahwa keputusan yang diambil tidak menghilangkan informasi yang relevan (Osborne & Overbay, 2004). Teknik lanjutan dalam menangani pencilan, seperti winsorizing, yaitu mengganti nilai ekstrem dengan nilai tertentu yang lebih mendekati distribusi data, serta trimming, yaitu menghapus sebagian kecil data ekstrem dari kedua ujung distribusi. Teknik ini sering digunakan untuk menjaga stabilitas analisis tanpa terlalu mengorbankan jumlah data. Dalam analisis statistik modern, penggunaan metode robust seperti regresi robust atau median absolut deviasi (MAD) juga semakin dianjurkan karena lebih tahan terhadap pengaruh pencilan dibandingkan metode konvensional berbasis mean (Wilcox, 2012).

5.4.4 Pelaporan Hasil Analisis Statistik yang Komprehensif

Pelaporan hasil analisis statistik harus dilakukan secara komprehensif dan jelas sehingga memudahkan pembaca untuk memahami dan mengevaluasi kevalidan kesimpulan penelitian. Pelaporan harus mencakup statistik deskriptif dari variabel-variabel utama seperti mean, standar deviasi, dan range untuk variabel numerik serta frekuensi dan proporsi untuk variabel kategorikal (Ruta et al., 2025). Pelaporan uji statistik harus mencakup jenis uji yang digunakan, nilai statistik uji, derajat kebebasan (*degrees of freedom*), nilai P, dan ukuran efek. Interpretasi hasil harus diberikan dengan jelas menjelaskan apa arti hasil statistik dalam konteks penelitian dan dampak praktis dari temuan penelitian.

pelaporan hasil analisis statistik juga harus mengikuti prinsip transparansi dan replikasi ilmiah. Peneliti perlu menyajikan informasi yang cukup rinci mengenai prosedur analisis yang digunakan, termasuk asumsi-asumsi yang diuji, langkah-langkah pengolahan data, serta perangkat lunak yang digunakan dalam analisis, seperti SPSS, Stata, atau R. Hal ini penting agar penelitian dapat direplikasi oleh peneliti lain dan meningkatkan kredibilitas hasil penelitian. Selain pelaporan angka-angka statistik, penyajian hasil juga sebaiknya dilengkapi dengan tabel dan grafik yang relevan untuk memperjelas temuan penelitian. Tabel digunakan untuk menyajikan data secara rinci dan sistematis, sedangkan grafik membantu pembaca memahami pola dan tren data secara lebih cepat. Namun demikian, penyajian visual harus tetap sederhana, tidak berlebihan, dan sesuai dengan prinsip kejelasan agar tidak menimbulkan bias interpretasi (Knafllic, 2015).

Latihan 5 Soal Essay

Soal 1: Jelaskan konsep distribusi data dan bagaimana bentuk distribusi data memengaruhi pemilihan metode analisis statistik yang tepat. Analisis perbedaan antara distribusi normal, distribusi yang miring ke kanan, dan distribusi yang miring ke kiri, serta bagaimana masing-masing bentuk distribusi ditunjukkan oleh nilai skewness dan kurtosis.

Soal 2: Diskusikan pentingnya uji normalitas dalam penelitian dan jelaskan berbagai metode yang dapat digunakan untuk

menguji normalitas data baik metode numerik maupun metode visual. Jelaskan implikasi hasil uji normalitas terhadap pemilihan antara prosedur statistik parametrik dan non-parametrik untuk analisis data.

Soal 3: Analisis komponen-komponen statistik deskriptif yang digunakan untuk meringkas dan menggambarkan data penelitian. Jelaskan ukuran pemusatan (*mean, median, modus*), ukuran penyebaran (*range, variansi, standar deviasi*), dan bagaimana penyajian data secara visual melalui grafik dan diagram dapat meningkatkan pemahaman tentang karakteristik data.

Soal 4: Diskusikan perbedaan antara statistik inferensial dan statistik deskriptif, serta jelaskan konsep pengujian hipotesis, nilai P, dan signifikansi statistik. Analisis pentingnya membedakan antara signifikansi statistik dan signifikansi klinis, serta peran ukuran efek dalam melengkapi nilai P untuk interpretasi hasil penelitian yang lebih lengkap.

Soal 5: Jelaskan prosedur identifikasi dan penanganan pencilan dalam dataset penelitian, serta bagaimana pencilan dapat memengaruhi hasil analisis statistik. Analisis standar pelaporan hasil analisis statistik yang komprehensif dan bagaimana peneliti dapat mengkomunikasikan temuan penelitian secara jelas dan akurat kepada pembaca untuk memastikan validitas dan kredibilitas kesimpulan penelitian.

DAFTAR PUSTAKA

Kocher, M. S., & Zurakowski, D. (2004). Clinical Epidemiology and Biostatistics. Journal of Bone and Joint Surgery. <https://doi.org/https://doi.org/10.2106/00004623-200403000-00024>

Cataldo, R., Arancibia, M., Stojanova, J., & Papuzinski, C. (2019). General concepts in biostatistics and clinical epidemiology: Observational studies with cross-sectional and ecological designs. Medwave.

<https://doi.org/https://doi.org/10.5867/medwave.2019.08.7698>

VanderWeele, T. J., Chen, Y., Long, K., Kim, E. S., Trudel-Fitzgerald, C., & Kubzansky, L. D. (2019). Positive Epidemiology? Epidemiology.

<https://doi.org/https://doi.org/10.1097/ede.0000000000001147>

Burgess, S., Foley, C. N., & Zuber, V. (2018). Inferring Causal Relationships Between Risk Factors and Outcomes from Genome-Wide Association Study Data. Annual Review of Genomics and Human Genetics (Print). <https://doi.org/10.1146/annurev-genom-083117-021731>

BAB 6

ANALISIS HUBUNGAN ANTAR VARIABEL DALAM KESEHATAN

Zakiyah, SST., M.K.M.

6.1 Konsep Dasar Analisis Hubungan Variabel Kesehatan

6.1.1 Pengertian Variabel dalam Penelitian Kesehatan

Dalam ilmu kesehatan, pemahaman tentang hubungan antar variabel merupakan fondasi penting untuk mengembangkan pengetahuan dan praktik klinis yang lebih baik (Kocher & Zurakowski, 2004). Variabel dapat didefinisikan sebagai karakteristik, atribut, atau pengukuran yang dapat berubah nilainya dan menjadi fokus penelitian. Dalam konteks kesehatan, variabel mencakup berbagai aspek mulai dari faktor biologis seperti tekanan darah, kadar kolesterol, hingga faktor sosial seperti pendidikan, penghasilan, dan akses ke pelayanan kesehatan. Setiap variabel memiliki sifat tersendiri dan dapat diukur dengan metode yang berbeda-beda.

Pentingnya memahami variabel terletak pada kemampuannya untuk menjelaskan pola penyakit, faktor risiko, dan hasil kesehatan dalam populasi. Ketika peneliti mempelajari hubungan antar variabel, mereka mencoba memahami bagaimana perubahan pada satu variabel dapat memengaruhi atau berhubungan dengan variabel lainnya. Misalnya, peneliti mungkin ingin mengetahui apakah ada hubungan antara pola makan, tingkat aktivitas fisik, dan berat badan pada sekelompok individu. Pemahaman ini memungkinkan para profesional kesehatan untuk mengidentifikasi determinan kesehatan dan mengembangkan intervensi yang lebih efektif.

6.1.2 Jenis-Jenis Variabel dalam Penelitian Epidemiologi

Dalam penelitian epidemiologi, variabel dapat diklasifikasikan menjadi beberapa kategori berdasarkan perannya dalam analisis (Cataldo et al., 2019). Variabel independen adalah variabel yang dianggap sebagai penyebab atau paparan dalam penelitian, seperti faktor risiko yang ingin dipelajari pengaruhnya terhadap kesehatan. Variabel dependen, sebaliknya, adalah hasil atau outcome yang ingin dijelaskan, misalnya terjadinya penyakit tertentu atau perubahan status kesehatan individu.

Selain itu, ada variabel perancu (confounding variables) yang dapat memengaruhi hubungan antara variabel independen dan dependen. Variabel perancu ini memiliki hubungan independen dengan kedua variabel utama dan dapat menyebabkan bias dalam hasil penelitian jika tidak dikendalikan dengan baik. Contohnya, ketika mempelajari hubungan antara konsumsi alkohol dan penyakit

jantung, usia dapat menjadi variabel perancu karena baik konsumsi alkohol maupun penyakit jantung sama-sama meningkat dengan bertambahnya usia.

6.1.3 Pentingnya Hubungan Antar Variabel dalam Kesehatan Masyarakat

Analisis hubungan antar variabel sangat penting dalam kesehatan masyarakat karena membantu mengidentifikasi faktor-faktor yang berkontribusi terhadap status kesehatan populasi (VanderWeele et al., 2019). Melalui pemahaman hubungan ini, profesional kesehatan dapat mengembangkan program pencegahan yang lebih tertarget dan intervensi yang lebih efektif. Sebagai contoh, jika penelitian menunjukkan hubungan kuat antara kemiskinan dan rendahnya akses layanan kesehatan dengan meningkatnya angka penyakit menular, upaya kesehatan masyarakat dapat diarahkan untuk memperbaiki kondisi sosial ekonomi dan meningkatkan akses layanan.

Pemahaman tentang hubungan ini juga penting untuk mengalokasikan sumber daya kesehatan secara lebih efisien, merencanakan program kesehatan yang sesuai dengan kebutuhan populasi spesifik, dan mengevaluasi efektivitas intervensi yang telah dilakukan. Selain itu, analisis hubungan variabel membantu peneliti membangun model prediktif yang dapat digunakan untuk mengidentifikasi individu dengan risiko tinggi terhadap penyakit tertentu, sehingga intervensi preventif dapat dilakukan lebih awal.

6.1.4 Tantangan dalam Menganalisis Hubungan Variabel Kesehatan

Meskipun penting, analisis hubungan antar variabel dalam kesehatan menghadapi berbagai tantangan yang perlu diatasi. Salah satu tantangan utama adalah membedakan antara korelasi dan kausalitas. Korelasi menunjukkan bahwa dua variabel bergerak bersama, namun ini tidak berarti satu variabel menyebabkan perubahan pada variabel lain. Untuk menentukan hubungan kausal, peneliti harus merancang penelitian dengan cermat dan mempertimbangkan berbagai faktor alternatif yang dapat menjelaskan hubungan yang diamati (Cataldo et al., 2019).

Tantangan lainnya meliputi kehadiran bias dalam pengumpulan data, keterbatasan sampel penelitian yang mungkin tidak representatif terhadap populasi yang lebih besar, serta adanya kesalahan pengukuran yang dapat memengaruhi akurasi hasil analisis. Variabilitas dalam cara pengukuran variabel antar penelitian juga dapat menyulitkan perbandingan hasil antar studi. Peneliti harus memiliki pemahaman mendalam tentang metodologi penelitian dan statistik untuk dapat mengatasi tantangan-tantangan ini dan menghasilkan kesimpulan yang valid dan dapat diandalkan.

6.2 Desain Penelitian untuk Menganalisis Hubungan Variabel

6.2.1 Studi Observasional dan Eksperimental

Dalam mempelajari hubungan antar variabel dalam kesehatan, peneliti dapat memilih antara desain penelitian observasional atau

eksperimental (Kocher & Zurakowski, 2004). Penelitian observasional melibatkan pengamatan terhadap kelompok subjek tanpa melakukan intervensi atau manipulasi terhadap variabel yang diteliti. Peneliti hanya mencatat dan menganalisis hubungan yang terjadi secara alami dalam populasi. Keuntungan dari desain ini adalah biaya yang lebih rendah dan kemungkinan untuk mempelajari dampak jangka panjang dari paparan tertentu.

Sebaliknya, penelitian eksperimental melibatkan manipulasi deliberat terhadap variabel independen dan penugasan acak subjek ke dalam kelompok perlakuan berbeda. Penelitian jenis ini dianggap lebih kuat dalam menentukan hubungan kausal karena randomisasi membantu meminimalkan pengaruh variabel perancu. Akan tetapi, penelitian eksperimental sering kali lebih mahal, memakan waktu, dan dalam beberapa kasus tidak etis dilakukan pada manusia, terutama ketika intervensi dapat menyebabkan kerugian.

6.2.2 Studi Potong Lintang (Cross-Sectional Study)

Studi potong lintang merupakan salah satu desain observasional yang paling sering digunakan dalam epidemiologi. Dalam desain ini, pengumpulan data mengenai paparan dan outcome dilakukan pada satu titik waktu, sehingga peneliti dapat menggambarkan prevalensi penyakit dan hubungannya dengan berbagai faktor pada populasi tertentu (Cataldo et al., 2019). Keuntungan dari studi ini adalah relatif murah, cepat, dan dapat memberikan gambaran umum tentang kondisi kesehatan suatu populasi pada waktu tertentu.

Namun, studi potong lintang memiliki keterbatasan dalam menentukan hubungan kausal karena tidak dapat memastikan urutan temporal antara paparan dan outcome. Ketika data dikumpulkan pada saat yang sama, sulit untuk menentukan apakah paparan mendahului outcome atau sebaliknya. Selain itu, studi ini juga rentan terhadap bias seleksi dan bias informasi, terutama jika tingkat responss rendah atau ada perbedaan dalam kualitas informasi antar kelompok.

6.2.3 Studi Kohort (Cohort Study)

Studi kohort adalah desain observasional prospektif di mana peneliti mengikuti kelompok individu yang memiliki paparan berbeda sepanjang waktu untuk melihat bagaimana mereka mengembangkan outcome yang diminati (Kocher & Zurakowski, 2004). Kelompok yang terpapar dan yang tidak terpapar diikuti ke depan dalam waktu untuk membandingkan tingkat terjadinya outcome di antara kedua kelompok. Desain ini memiliki keunggulan dalam menetapkan urutan temporal yang jelas antara paparan dan outcome, sehingga dapat memberikan bukti lebih kuat tentang hubungan kausal.

Studi kohort juga memungkinkan perhitungan risiko insidensi dan risiko relatif, yang merupakan ukuran asosiasi penting dalam epidemiologi. Meskipun demikian, studi kohort memerlukan waktu lama, biaya yang besar, dan sumber daya manusia yang signifikan. Selain itu, ada risiko kehilangan data (attrition) ketika subjek penelitian meninggalkan studi selama periode follow-up.

6.2.4 Studi Kasus-Kontrol dan Studi Ekologis

Studi kasus-kontrol merupakan desain observasional retrospektif yang mulai dengan mengidentifikasi individu dengan outcome tertentu (kasus) dan membandingkan mereka dengan individu tanpa outcome tersebut (kontrol) dalam hal paparan masa lalu (Cataldo et al., 2019). Desain ini sangat efisien untuk mempelajari penyakit yang jarang terjadi atau memiliki periode laten yang panjang, karena peneliti dapat secara deliberat memilih subjek berdasarkan status outcome. Studi kasus-kontrol lebih cepat dan murah dibandingkan studi kohort.

Sementara itu, studi ekologis melibatkan analisis data agregat pada tingkat populasi atau kelompok, bukan pada tingkat individu. Data mengenai paparan dan outcome diukur untuk sekelompok individu sebagai satu unit, dan hubungan dianalisis pada tingkat agregat. Meskipun studi ekologis berguna untuk mengidentifikasi pola umum dan menghasilkan hipotesis penelitian, mereka rentan terhadap kesalahan ekologis (ecological fallacy), di mana hubungan yang diamati pada tingkat kelompok tidak dapat begitu saja digeneralisasikan ke tingkat individu.

6.3 Mengidentifikasi dan Mengendalikan Bias dalam Analisis Hubungan Variabel

6.3.1 Definisi dan Jenis-Jenis Bias

Bias adalah kesalahan sistemik bukan acak yang dapat menyebabkan estimasi hubungan antar variabel menjadi tidak

akurat (Kocher & Zurakowski, 2004). Bias dapat muncul di berbagai tahap penelitian, mulai dari desain studi, pengumpulan data, pengolahan data, hingga analisis. Ada tiga kategori utama bias yang perlu diperhatikan: bias seleksi, bias informasi, dan bias perancu. Bias seleksi terjadi ketika cara subjek dipilih untuk penelitian berhubungan dengan paparan dan outcome secara bersamaan, sehingga menyebabkan distorsi pada estimasi asosiasi.

Bias informasi muncul ketika ada kesalahan dalam pengukuran paparan atau outcome, baik karena alat pengukur yang tidak akurat, kesalahan responden dalam mengingat informasi masa lalu, atau kesalahan pada proses pengumpulan data. Bias perancu (confounding bias) terjadi ketika ada variabel ketiga yang memiliki hubungan independen dengan kedua variabel utama dan tidak dikendalikan dalam analisis, sehingga mengaburkan atau mengubah hubungan yang sebenarnya antara variabel independen dan dependen.

6.3.2 Strategi Pencegahan Bias dalam Tahap Desain

Pencegahan bias dimulai dari tahap desain penelitian dengan memilih desain yang tepat untuk pertanyaan penelitian yang akan dijawab. Randomisasi adalah strategi paling kuat untuk meminimalkan bias seleksi dan bias perancu dalam penelitian eksperimental, karena penugasan acak ke dalam kelompok perlakuan dan kontrol membantu memastikan bahwa kedua kelompok setara dalam hal karakteristik lainnya (Cataldo et al., 2019). Blinding, yaitu penyembunyian informasi tentang penugasan

kelompok dari subjek penelitian dan atau peneliti, dapat membantu mengurangi bias informasi dan bias pengukuran.

Pembatasan sampel penelitian juga dapat membantu mengendalikan variabel perancu. Misalnya, jika peneliti ingin mempelajari hubungan antara kafein dan anemia, mereka dapat membatasi penelitian hanya pada wanita usia reproduksi untuk menghilangkan pengaruh perbedaan usia dan jenis kelamin. Matching adalah teknik lain di mana peneliti memasang subjek dalam kelompok berbeda berdasarkan karakteristik tertentu yang diduga menjadi perancu, sehingga memastikan bahwa kelompok setara dalam hal karakteristik tersebut.

6.3.3 Strategi Kontrol Bias dalam Tahap Analisis

Ketika bias tidak dapat sepenuhnya dicegah dalam tahap desain, peneliti dapat menggunakan berbagai teknik analisis untuk mengendalikan bias di tahap analisis data (Kocher & Zurakowski, 2004). Stratifikasi melibatkan pembagian data ke dalam strata berdasarkan variabel perancu yang diduga, dan analisis dilakukan secara terpisah dalam setiap stratum untuk melihat apakah asosiasi konsisten di semua strata. Jika asosiasi berbeda antar strata, hal ini menunjukkan adanya modifikasi efek, yaitu bahwa kekuatan hubungan bervariasi bergantung pada nilai variabel ketiga.

Analisis multivariabel seperti regresi logistik dan regresi linear memungkinkan peneliti untuk menyesuaikan atau mengontrol efek dari beberapa variabel perancu secara bersamaan dalam satu model. Teknik ini sangat berguna ketika ada banyak variabel perancu

potensial yang perlu dipertimbangkan. Analisis sensitivitas juga penting untuk dilakukan, yaitu mengulangi analisis dengan asumsi atau metode yang berbeda untuk melihat apakah kesimpulan tetap sama, sehingga menunjukkan robustness dari hasil penelitian.

6.3.4 Evaluasi Kualitas Bukti dan Tingkat Kesimpulan

Dalam mengevaluasi hubungan antar variabel, penting untuk mempertimbangkan tingkat bukti yang tersedia dan seberapa yakin kita dapat mengambil kesimpulan tentang hubungan kausal. Bukti dari studi desain yang berbeda memiliki kekuatan yang berbeda, dengan umumnya uji coba terkontrol acak dianggap memberikan bukti paling kuat, diikuti studi kohort prospektif, studi kasus-kontrol, dan studi potong lintang (Cataldo et al., 2019). Kriteria Bradford Hill, yang meliputi kekuatan asosiasi, konsistensi, spesifisitas, temporalitas, gradien biologis, dan plausibilitas biologis, dapat digunakan untuk menilai apakah hubungan yang diamati kemungkinan bersifat kausal.

Peneliti juga harus mempertimbangkan adanya bias publikasi, yaitu kecenderungan studi dengan hasil signifikan untuk dipublikasikan daripada studi dengan hasil negatif. Meta-analisis dari beberapa studi dapat memberikan estimasi efek yang lebih presisi dengan mengkombinasikan data dari studi individual, namun kualitas meta-analisis bergantung pada kualitas studi yang diinklusi. Transparansi dalam melaporkan desain studi, metode, dan pembatasan sangat penting untuk memungkinkan pembaca mengevaluasi keandalan dan relevansi hasil penelitian.

6.4 Metode Statistik untuk Menganalisis Hubungan Variabel

6.4.1 Ukuran Asosiasi dalam Epidemiologi

Untuk mengkuantifikasi hubungan antara variabel independen dan dependen, epidemiolog menggunakan berbagai ukuran asosiasi (Burgess et al., 2018). Risiko relatif (relative risk) adalah rasio dari risiko terjadinya outcome pada kelompok terpapar dibandingkan dengan risiko pada kelompok tidak terpapar. Ukuran ini sangat berguna dalam studi kohort dan memberikan interpretasi yang intuitif tentang berapa kali lebih besar risiko pada kelompok terpapar. Odds ratio adalah rasio dari odds terjadinya outcome pada kelompok terpapar dibandingkan dengan odds pada kelompok tidak terpapar, dan umumnya digunakan dalam studi kasus-kontrol.

Risiko atribut (attributable risk) adalah selisih antara risiko pada kelompok terpapar dan kelompok tidak terpapar, dan menunjukkan risiko absolut yang dapat dikaitkan dengan paparan. Fraksi risiko atribut adalah proporsi dari kasus penyakit dalam kelompok terpapar yang dapat dikaitkan dengan paparan, dan memberikan informasi tentang seberapa besar kontribusi paparan tertentu terhadap beban penyakit. Memahami perbedaan antara ukuran asosiasi ini penting untuk interpretasi yang akurat dari hasil penelitian dan untuk membuat keputusan kesehatan masyarakat yang tepat.

6.4.2 Analisis Univariabel dan Multivariabel

Analisis univariabel melibatkan pemeriksaan hubungan antara satu variabel independen dan satu variabel dependen tanpa

mempertimbangkan pengaruh variabel lain. Analisis ini memberikan gambaran pertama tentang apakah ada asosiasi antara dua variabel, namun hasilnya dapat menyesatkan jika ada variabel perancu yang memengaruhi hubungan tersebut. Analisis bivariabel, yang melibatkan dua variabel independen, dapat memberikan gambaran yang lebih lengkap, tetapi masih belum cukup untuk mengendalikan semua variabel perancu yang potensial.

Analisis multivariabel, sebaliknya, melibatkan pemodelan hubungan antara satu atau lebih variabel independen dan satu variabel dependen sambil mengontrol pengaruh dari variabel independen lain yang dianggap sebagai perancu atau modifikator efek. Teknik ini memungkinkan peneliti untuk memperkirakan asosiasi independen antara setiap variabel independen dan outcome, sementara menyesuaikan untuk pengaruh variabel lain dalam model. Pilihan teknik multivariabel bergantung pada sifat variabel dependen dan asumsi yang ingin dipenuhi dalam analisis.

6.4.3 Regresi Logistik dan Regresi Linear

Regresi logistik adalah teknik analisis multivariabel yang digunakan ketika variabel dependen bersifat biner atau kategoris, seperti ada atau tidak adanya penyakit (Burgess et al., 2018). Dalam regresi logistik, peneliti membangun model yang mengestimasi probabilitas atau odds dari outcome berdasarkan nilai variabel independen. Koefisien dalam model regresi logistik dapat diinterpretasikan sebagai perubahan dalam log odds dari outcome

untuk setiap unit perubahan dalam variabel independen, setelah menyesuaikan untuk variabel lain dalam model.

Regresi linear, di sisi lain, digunakan ketika variabel dependen bersifat kontinu, seperti tekanan darah atau tingkat kolesterol. Dalam regresi linear, peneliti membangun model yang mengestimasi nilai rata-rata variabel dependen berdasarkan nilai variabel independen. Koefisien regresi dapat diinterpretasikan sebagai perubahan rata-rata dalam variabel dependen untuk setiap unit perubahan dalam variabel independen, setelah menyesuaikan untuk variabel lain. Kedua teknik ini dapat diperluas untuk menangani kompleksitas data seperti interaksi antar variabel, hubungan nonlinear, dan struktur data berlapis.

6.4.4 Inferensi Kausal dan Mendelian Randomisasi

Mengingat tantangan dalam menentukan hubungan kausal dari data observasional, peneliti telah mengembangkan metode canggih untuk memperkuat inferensi kausal (Burgess et al., 2018). Mendelian randomisasi adalah pendekatan yang menggunakan varian genetik yang terbukti berhubungan dengan variabel independen untuk mengidentifikasi hubungan kausal dengan outcome, mengandalkan asumsi bahwa alokasi alel genetik adalah acak dan mirip dengan randomisasi dalam uji coba terkontrol. Metode ini dapat membantu mengatasi bias perancu yang tidak teramati dalam studi observasional.

Pendekatan kausal lainnya meliputi penggunaan diagram kausal (causal diagrams) untuk secara visual menggambarkan

hubungan hipotetis antar variabel dan mengidentifikasi set variabel yang perlu dikontrol untuk estimasi efek kausal yang tidak bias. Analisis sensitivitas dapat digunakan untuk menilai seberapa sensitif kesimpulan kausal terhadap pelanggaran asumsi atau adanya bias yang tidak teramati. Meskipun metode-metode ini memberikan alat yang lebih kuat untuk inferensi kausal, mereka memerlukan pemahaman mendalam tentang epidemiologi, statistik, dan biologi sistem untuk diterapkan dengan benar.

6.5 Aplikasi Praktis dalam Penelitian Kesehatan Populasi

6.5.1 Identifikasi Faktor Risiko dan Faktor Protektif

Salah satu aplikasi praktis analisis hubungan antar variabel adalah mengidentifikasi faktor risiko yang meningkatkan kemungkinan terjadinya penyakit dan faktor protektif yang menurunkan risiko tersebut (VanderWeele et al., 2019). Melalui penelitian epidemiologis yang terstruktur dengan baik, peneliti dapat menentukan paparan atau kondisi apa yang berkontribusi terhadap terjadinya penyakit dalam populasi. Pengetahuan ini memungkinkan pengembangan strategi pencegahan yang menargetkan faktor-faktor penting ini, baik melalui perubahan perilaku individu, modifikasi lingkungan, maupun kebijakan kesehatan masyarakat.

Identifikasi faktor protektif sama pentingnya dengan identifikasi faktor risiko, karena pemahaman tentang apa yang menjaga kesehatan individu dan populasi dapat membantu dalam mengembangkan program promosi kesehatan yang efektif.

Penelitian menunjukkan bahwa memahami aset kesehatan masyarakat, seperti dukungan sosial, kohesi komunitas, dan akses ke layanan kesehatan berkualitas, dapat memberikan gambaran yang lebih lengkap tentang determinan kesehatan daripada hanya fokus pada faktor negatif semata.

6.5.2 Stratifikasi Risiko dan Identifikasi Kelompok Rentan

Dengan menganalisis hubungan antar variabel, peneliti dapat mengidentifikasi kelompok populasi yang memiliki risiko lebih tinggi terhadap penyakit tertentu dan memerlukan perhatian kesehatan khusus. Stratifikasi risiko melibatkan pembagian populasi ke dalam kategori risiko berbeda berdasarkan kombinasi faktor-faktor yang relevan. Pendekatan ini memungkinkan alokasi sumber daya kesehatan yang lebih efisien dengan menargetkan intervensi preventif dan kuratif kepada kelompok dengan kebutuhan paling mendesak (Cataldo et al., 2019).

Identifikasi kelompok rentan juga penting untuk memastikan bahwa program kesehatan masyarakat tidak hanya efektif secara keseluruhan, tetapi juga adil dan dapat mengurangi kesenjangan kesehatan antar kelompok populasi. Penelitian sering kali menunjukkan bahwa pola risiko penyakit dan akses terhadap layanan kesehatan bervariasi menurut demografi, sosial ekonomi, dan faktor geografis, sehingga diperlukan pendekatan yang disesuaikan dengan konteks lokal.

6.5.3 Evaluasi Program Kesehatan dan Intervensi

Analisis hubungan antar variabel juga digunakan untuk mengevaluasi efektivitas program kesehatan dan intervensi yang telah dilaksanakan. Dengan mengukur perubahan dalam variabel hasil kesehatan sebelum dan sesudah intervensi, atau dengan membandingkan hasil pada kelompok yang menerima intervensi dengan kelompok kontrol, peneliti dapat menentukan apakah intervensi mencapai tujuannya (Kocher & Zurakowski, 2004). Evaluasi ini penting tidak hanya untuk menilai keberhasilan program, tetapi juga untuk mengidentifikasi area yang memerlukan perbaikan dan untuk memberikan justifikasi untuk pendanaan berkelanjutan.

Studi tentang mediator dan moderator dalam hubungan antara intervensi dan outcome dapat memberikan wawasan tentang mekanisme bagaimana intervensi bekerja dan untuk siapa intervensi paling efektif. Pemahaman ini dapat membantu dalam mengadaptasi dan mengoptimalkan program kesehatan untuk meningkatkan dampak kesehatannya terhadap populasi.

6.5.4 Pemodelan dan Prediksi Outcome Kesehatan

Teknik analisis multivariabel dapat digunakan untuk membangun model prediktif yang dapat memperkirakan risiko individu terhadap penyakit tertentu berdasarkan profil faktor risiko mereka (Burgess et al., 2018). Model-model ini, seperti model regresi logistik atau pohon keputusan, dapat digunakan dalam praktik klinis untuk membantu clinician dalam membuat keputusan

tentang screening, diagnosis, atau pengobatan. Model prediktif juga dapat digunakan dalam kesehatan masyarakat untuk mengidentifikasi area geografis dengan beban penyakit tinggi atau untuk meramalkan trend penyakit di masa depan.

Kualitas model prediktif bergantung pada keakuratan data yang digunakan untuk membangun model, relevansi variabel yang dimasukkan dalam model, dan seberapa baik model dapat digeneralisasikan ke populasi baru atau setting berbeda. Validasi eksternal dari model pada data independen sangat penting untuk menilai generalisabilitas dan aplikabilitas praktis dari model dalam setting klinis atau kesehatan masyarakat yang sebenarnya.

Latihan Soal Essay Singkat

Soal 1: Jelaskan perbedaan antara korelasi dan kausalitas dalam analisis hubungan antar variabel kesehatan. Berikan contoh konkret untuk mengilustrasikan perbedaan tersebut dan jelaskan mengapa perbedaan ini penting dalam penelitian epidemiologi.

Soal 2: Seorang peneliti melakukan studi potong lintang untuk mengkaji hubungan antara konsumsi gula harian dan kejadian diabetes tipe 2 pada populasi urban. Diskusikan kekuatan dan keterbatasan desain studi ini dalam menentukan hubungan kausal antara konsumsi gula dan diabetes, serta jelaskan bagaimana desain studi alternatif dapat mengatasi keterbatasan tersebut.

Soal 3: Dalam sebuah studi, ditemukan bahwa penggunaan obat antidepresi berhubungan dengan peningkatan risiko bunuh diri

pada remaja. Jelaskan bagaimana reverse causation dapat memengaruhi interpretasi hubungan ini dan diskusikan metodologi apa yang dapat digunakan untuk membedakan antara efek obat sebenarnya versus reverse causation.

Soal 4: Anda diminta menganalisis faktor-faktor yang memengaruhi kehadiran ibu untuk pemeriksaan kesehatan ibu hamil. Jelaskan variabel independen, variabel dependen, dan variabel perancu yang mungkin relevan dalam studi ini, serta diskusikan strategi analisis multivariabel yang dapat digunakan untuk mengidentifikasi faktor independen yang memengaruhi kehadiran pemeriksaan.

Soal 5: Sebuah penelitian menemukan bahwa orang yang berolahraga secara teratur memiliki tingkat kolesterol yang lebih rendah. Jelaskan bagaimana Anda akan merancang studi untuk menentukan apakah olahraga secara kausal menyebabkan penurunan kolesterol atau apakah hubungan ini disebabkan oleh variabel perancu lain. Diskusikan juga bagaimana Anda akan mengidentifikasi mekanisme melalui mana olahraga dapat memengaruhi kolesterol.

DAFTAR PUSTAKA

- Kocher, M. S., & Zurakowski, D. (2004). Clinical Epidemiology and Biostatistics. Journal of Bone and Joint Surgery. <https://doi.org/https://doi.org/10.2106/00004623-200403000-00024>
- Cataldo, R., Arancibia, M., Stojanova, J., & Papuzinski, C. (2019). General concepts in biostatistics and clinical epidemiology: Observational studies with cross-sectional and ecological designs. Medwave. <https://doi.org/https://doi.org/10.5867/medwave.2019.08.7698>
- VanderWeele, T. J., Chen, Y., Long, K., Kim, E. S., Trudel-Fitzgerald, C., & Kubzansky, L. D. (2019). Positive Epidemiology? Epidemiology. <https://doi.org/https://doi.org/10.1097/ede.0000000000001147>
- Burgess, S., Foley, C. N., & Zuber, V. (2018). Inferring Causal Relationships Between Risk Factors and Outcomes from Genome-Wide Association Study Data. Annual Review of Genomics and Human Genetics (Print). <https://doi.org/10.1146/annurev-genom-083117-021731>

BAB 7

PENGUNAAN

BIOSTATISTIK DALAM

PEMANTAUAN KESEHATAN

MASYARAKAT

Tri Budiarti, S.K.M, M.K.M.

7.1 Pengertian Biostatistik dalam Kesehatan Masyarakat

Biostatistik merupakan aplikasi dari ilmu statistik untuk data biologis dan kesehatan yang menjadi fondasi penting dalam pemantauan kesehatan masyarakat. Dalam konteks kesehatan masyarakat, biostatistik digunakan untuk mengumpulkan, menganalisis, dan menginterpretasi data kesehatan secara sistematis dan berkelanjutan guna mendukung pengambilan keputusan yang berbasis bukti. Pemahaman mendalam tentang biostatistik sangat penting bagi para profesional kesehatan, pemantau penyakit, dan pembuat kebijakan kesehatan karena memungkinkan mereka untuk mengidentifikasi masalah kesehatan, memantau tren penyakit, dan

mengevaluasi efektivitas program kesehatan yang telah dilaksanakan.

7.1.1 Definisi dan Ruang Lingkup Biostatistik

Biostatistik didefinisikan sebagai penggunaan metode statistik untuk data yang berasal dari ilmu biologi dan kesehatan dalam rangka mengambil kesimpulan yang valid tentang fenomena kesehatan dan penyakit. Ruang lingkup biostatistik dalam kesehatan masyarakat mencakup desain survei epidemiologi, analisis data penyakit, interpretasi hasil penelitian kesehatan, dan evaluasi program kesehatan. Biostatistik berbeda dari statistik umum karena fokusnya pada masalah spesifik dalam kesehatan, seperti estimasi prevalensi penyakit, perhitungan insidensi, analisis faktor risiko, dan pemodelan penyakit. Penggunaan biostatistik yang tepat memastikan bahwa data kesehatan yang dikumpulkan dapat memberikan informasi yang akurat dan dapat dipercaya untuk mengarahkan intervensi kesehatan.

7.1.2 Pentingnya Biostatistik dalam Kebijakan Kesehatan

Biostatistik memainkan peran krusial dalam merumuskan kebijakan kesehatan yang efektif dan efisien. Dengan menganalisis data epidemiologi menggunakan metode biostatistik yang tepat, pembuat kebijakan dapat mengidentifikasi masalah kesehatan yang paling mendesak, menentukan prioritas intervensi, dan mengalokasikan sumber daya kesehatan secara optimal. Biostatistik juga membantu dalam memantau pencapaian target kesehatan, mengevaluasi dampak program kesehatan yang telah berjalan, dan

menyesuaikan strategi kesehatan berdasarkan bukti empiris. Dalam era kesehatan masyarakat modern, pengambilan keputusan yang didasarkan pada data statistik telah menjadi standar dalam mengelola masalah kesehatan di tingkat nasional maupun lokal.

7.1.3 Jenis Data dalam Biostatistik Kesehatan

Data dalam biostatistik kesehatan dapat dikategorikan menjadi beberapa jenis tergantung pada sumbernya dan karakteristiknya. Data primer adalah data yang dikumpulkan langsung dari responden melalui survei, wawancara, atau observasi lapangan. Data sekunder adalah data yang sudah dikumpulkan sebelumnya dan tersedia dalam laporan kesehatan, sistem informasi, atau publikasi ilmiah. Data kuantitatif adalah data dalam bentuk angka yang dapat dianalisis menggunakan metode statistik, sedangkan data kualitatif adalah data dalam bentuk deskripsi yang memberikan konteks untuk memahami fenomena kesehatan. Pemilihan jenis data yang tepat sangat penting karena setiap jenis data memiliki keunggulan dan keterbatasan tersendiri dalam memberikan informasi tentang kesehatan masyarakat.

7.1.4 Aplikasi Metode Statistik dalam Analisis Kesehatan Masyarakat

Metode statistik yang digunakan dalam biostatistik kesehatan mencakup statistik deskriptif yang memberikan gambaran ringkas tentang data kesehatan, statistik inferensia yang memungkinkan generalisasi dari sampel ke populasi, dan analisis multivariat yang menganalisis hubungan kompleks antara berbagai

faktor kesehatan. Statistik deskriptif mencakup perhitungan proporsi, prevalensi, insidensi, dan visualisasi data melalui tabel dan grafik. Statistik inferensia mencakup uji hipotesis dan perhitungan interval kepercayaan yang membantu peneliti membuat kesimpulan tentang populasi berdasarkan data sampel. Analisis multivariat seperti regresi logistik dan regresi linear memungkinkan identifikasi faktor risiko yang paling berpengaruh terhadap terjadinya penyakit.

7.2 Surveilans Kesehatan Masyarakat

Surveilans kesehatan masyarakat merupakan sistem pengamatan yang sistematis dan berkelanjutan terhadap pola penyakit dan kondisi kesehatan yang memengaruhi terjadinya peningkatan penularan penyakit. Surveilans dirancang untuk mengumpulkan, menganalisis, dan menyebarluaskan informasi kesehatan guna mengarahkan tindakan penanggulangan penyakit secara efektif dan efisien. Dalam praktiknya, surveilans kesehatan masyarakat menggunakan prinsip dan metode biostatistik untuk mengolah data penyakit dan kesehatan menjadi informasi yang berguna bagi para pengambil keputusan.

7.2.1 Komponen Sistem Surveilans Kesehatan

Sistem surveilans kesehatan masyarakat terdiri dari beberapa komponen utama yang saling terkait. Pertama, pengumpulan data dilakukan melalui laporan dari fasilitas kesehatan, laboratorium, dan masyarakat tentang kasus penyakit yang terjadi. Kedua, pencatatan dan pengolahan data dilakukan untuk mengorganisir informasi yang

dikumpulkan menjadi bentuk yang dapat dianalisis. Ketiga, analisis data dilakukan menggunakan metode biostatistik untuk mengidentifikasi pola, tren, dan anomali dalam data penyakit. Keempat, interpretasi hasil analisis dilakukan untuk memahami makna epidemiologi dari data yang ada. Kelima, diseminasi informasi adalah penyebarluasan hasil analisis kepada pihak-pihak yang perlu untuk mengambil tindakan. Keenam, respons dan evaluasi dilakukan untuk mengevaluasi efektivitas tindakan yang telah diambil berdasarkan informasi surveilans (Iqbal, 2025).

7.2.2 Desain Surveilans dan Pengumpulan Data

Desain surveilans kesehatan dapat berbentuk surveilans rutin yang dilakukan secara berkelanjutan untuk penyakit-penyakit tertentu, atau surveilans sentinel yang dilakukan pada lokasi terpilih untuk memantau penyakit prioritas. Surveilans aktif melibatkan pencarian informasi secara proaktif dari fasilitas kesehatan, sedangkan surveilans pasif mengandalkan laporan spontan dari fasilitas kesehatan yang diintegrasikan dalam sistem informasi kesehatan. Pengumpulan data surveilans dapat dilakukan melalui berbagai cara, mulai dari wawancara langsung dengan penderita, pemeriksaan laboratorium, hingga analisis data elektronik dari sistem informasi kesehatan. Efektivitas surveilans sangat tergantung pada kualitas data yang dikumpulkan, ketepatan waktu pengumpulan, dan komitmen dari berbagai pihak dalam melaporkan kejadian penyakit (Wahyuni et al., 2026).

7.2.3 Analisis Data Surveilans dan Identifikasi Masalah

Analisis data surveilans menggunakan metode biostatistik untuk mengidentifikasi masalah kesehatan yang memerlukan tindakan. Analisis deskriptif dilakukan untuk menggambarkan distribusi penyakit menurut orang, tempat, dan waktu, membantu mengidentifikasi kelompok populasi yang paling berisiko, wilayah geografis dengan insidensi tertinggi, dan tren temporal penyakit. Pemetaan kasus dilakukan untuk visualisasi spasial distribusi penyakit, memudahkan identifikasi kluster kasus dan perencanaan intervensi yang ditargetkan. Analisis tren digunakan untuk memantau perubahan dalam frekuensi penyakit dari waktu ke waktu, mengidentifikasi peningkatan atau penurunan kasus. Ketika ditemukan peningkatan kasus yang tidak biasa, penyelidikan epidemiologi lebih lanjut dilakukan untuk mengidentifikasi sumber penularan dan faktor risiko (Arsyad et al., 2025).

7.2.4 Validitas dan Keandalan Data Surveilans

Kualitas data surveilans sangat penting untuk memastikan bahwa informasi yang dihasilkan dapat diandalkan dan berguna untuk pengambilan keputusan. Validitas data mengacu pada akurasi dalam pengumpulan dan pencatatan data, memastikan bahwa kasus yang dilaporkan memenuhi definisi kasus yang telah ditetapkan. Keandalan data mengacu pada konsistensi pengumpulan data dari waktu ke waktu dan di berbagai lokasi. Untuk meningkatkan validitas dan keandalan data surveilans, perlu dilakukan pelatihan berkala bagi petugas pencatat data, verifikasi data berkala untuk

mendeteksi kesalahan, dan audit sistem surveilans untuk mengevaluasi kualitas keseluruhan. Penggunaan sistem informasi berbasis elektronik juga dapat meningkatkan validitas dan keandalan data dengan mengurangi kesalahan pencatatan manual dan meningkatkan ketepatan waktu pelaporan.

7.3 Analisis Data Epidemiologi

Analisis data epidemiologi menggunakan metode biostatistik untuk memahami pola penyakit, mengidentifikasi faktor risiko, dan mengevaluasi efektivitas intervensi kesehatan. Analisis epidemiologi merupakan proses sistematis untuk mengubah data kesehatan mentah menjadi informasi yang berguna bagi pengambil keputusan dan pembuat kebijakan kesehatan.

7.3.1 Ukuran-Ukuran Epidemiologi

Ukuran epidemiologi merupakan alat penting dalam mengkuantifikasi beban penyakit dalam populasi. Prevalensi adalah proporsi dari jumlah kasus penyakit yang ada dalam populasi tertentu pada waktu tertentu, memberikan gambaran tentang seberapa luas penyakit tersebar dalam masyarakat. Insidensi adalah jumlah kasus baru penyakit yang terjadi dalam periode waktu tertentu dalam populasi yang berisiko, memberikan informasi tentang laju penularan penyakit. Mortalitas adalah proporsi dari kematian yang terjadi karena penyakit tertentu dalam populasi, menggambarkan tingkat keparahan penyakit. Morbiditas adalah proporsi dari orang yang menderita penyakit tertentu dalam

populasi, menunjukkan beban penyakit yang lebih luas. Risiko relatif dan odds ratio digunakan untuk menghitung kekuatan hubungan antara faktor risiko dan terjadinya penyakit. Pemahaman yang tepat tentang ukuran-ukuran ini penting untuk interpretasi yang benar tentang beban penyakit dan efektivitas program kesehatan.

7.3.2 Stratifikasi dan Perbandingan Data

Stratifikasi data epidemiologi dilakukan untuk membagi populasi menjadi subkelompok berdasarkan karakteristik demografis atau epidemiologi tertentu seperti usia, jenis kelamin, wilayah geografis, atau faktor risiko lainnya. Stratifikasi memungkinkan identifikasi perbedaan dalam beban penyakit antara berbagai kelompok populasi, yang penting untuk merancang intervensi yang ditargetkan. Perbandingan data antar waktu memungkinkan identifikasi tren penyakit, baik peningkatan maupun penurunan dari waktu ke waktu. Perbandingan data antar wilayah memungkinkan identifikasi disparitas geografis dalam beban penyakit, yang dapat mencerminkan perbedaan dalam akses terhadap layanan kesehatan, kondisi sosial ekonomi, atau faktor lingkungan. Dengan melakukan stratifikasi dan perbandingan yang sistematis, petugas kesehatan dapat mengidentifikasi kelompok populasi atau wilayah yang memerlukan perhatian khusus dan intervensi yang diprioritaskan.

7.3.3 Penyelidikan Wabah Berbasis Biostatistik

Penyelidikan wabah atau *outbreak investigation* adalah proses sistematis untuk mengidentifikasi sumber penularan dan

faktor risiko ketika terjadi peningkatan kasus penyakit yang tidak biasa. Langkah awal dalam penyelidikan wabah adalah verifikasi bahwa peningkatan kasus yang dilaporkan memang nyata dan bukan hasil dari kesalahan pencatatan data. Selanjutnya, dilakukan perhitungan ukuran epidemiologi seperti attack rate untuk menggambarkan keparahan wabah. Analisis epidemiologi dilakukan untuk mengidentifikasi pola distribusi kasus menurut orang, tempat, dan waktu. Studi kasus kontrol atau studi kohort dapat dilakukan untuk mengidentifikasi faktor risiko yang spesifik. Berdasarkan hasil penyelidikan, rekomendasi kontrol dilakukan untuk menghentikan penularan dan mencegah wabah di masa depan (Wignjadiputro et al., 2018).

7.3.4 Evaluasi Program Kesehatan Berbasis Data

Evaluasi program kesehatan menggunakan metode biostatistik untuk menilai pencapaian tujuan program, efektivitas intervensi, dan dampak program terhadap kesehatan masyarakat. Evaluasi dapat menggunakan metode kuantitatif seperti perhitungan perubahan dalam insidensi atau prevalensi penyakit sebelum dan sesudah program, atau metode kualitatif untuk memahami persepsi masyarakat tentang program. Indikator program kesehatan ditetapkan dengan jelas di awal program untuk memudahkan pengukuran hasil. Perbandingan antara pencapaian dan target program dilakukan untuk menilai seberapa baik program berjalan. Analisis efektivitas biaya dilakukan untuk menilai efisiensi penggunaan sumber daya program. Hasil evaluasi digunakan untuk

membuat rekomendasi untuk perbaikan atau perluasan program di masa depan.

7.4 Aplikasi Sistem Informasi dalam Pemantauan Kesehatan

Perkembangan teknologi informasi telah memberikan kontribusi yang signifikan dalam meningkatkan efisiensi dan efektivitas pemantauan kesehatan masyarakat. Sistem informasi kesehatan mengintegrasikan pengumpulan data, analisis, dan diseminasi informasi dalam satu platform yang dapat diakses oleh berbagai pihak yang berperan dalam kesehatan masyarakat.

7.4.1 Sistem Informasi Kesehatan dan Database

Sistem informasi kesehatan merupakan infrastruktur teknologi yang digunakan untuk mengumpulkan, menyimpan, memproses, dan menampilkan data kesehatan dalam format yang mudah diakses dan dianalisis. Database kesehatan dirancang untuk menyimpan informasi tentang kasus penyakit, demografi penduduk, hasil laboratorium, dan data kesehatan lainnya dengan struktur yang terorganisir. Sistem informasi yang baik memungkinkan petugas kesehatan untuk dengan cepat memasukkan data, mengambil data untuk analisis, dan menghasilkan laporan kesehatan secara otomatis. Integrasi antara berbagai sumber data kesehatan, seperti data dari puskesmas, rumah sakit, laboratorium, dan fasilitas kesehatan lainnya, dalam satu sistem informasi kesehatan memudahkan analisis data yang komprehensif dan pengambilan keputusan yang lebih baik. Keamanan dan privasi data menjadi pertimbangan

penting dalam desain dan implementasi sistem informasi kesehatan untuk melindungi informasi pribadi pasien.

7.4.2 Sistem Informasi Geografis untuk Pemetaan Penyakit

Sistem Informasi Geografis atau *Geographic Information System* (GIS) merupakan teknologi yang memungkinkan visualisasi data kesehatan dalam bentuk peta, membantu identifikasi pola spasial penyakit dan perencanaan intervensi yang lebih tertarget. GIS mengintegrasikan data epidemiologi dengan data geografis untuk menampilkan distribusi penyakit dalam ruang geografis. Pemetaan menggunakan GIS memungkinkan identifikasi *hotspot* atau area dengan konsentrasi kasus penyakit yang tinggi, yang memerlukan perhatian khusus dalam pengendalian penyakit. Visualisasi spasial juga membantu dalam memahami hubungan antara faktor lingkungan seperti sanitasi, kepadatan penduduk, dan distribusi layanan kesehatan dengan terjadinya penyakit. Pemetaan spasial penyakit juga dapat digunakan untuk monitoring dan evaluasi program kesehatan dengan membandingkan perubahan distribusi penyakit dari waktu ke waktu.

7.4.3 Manajemen Data Real-Time dan Pelaporan

Sistem informasi kesehatan modern memungkinkan pengumpulan dan analisis data secara real-time, memberikan informasi terkini tentang situasi kesehatan di masyarakat. Pelaporan otomatis dapat dikonfigurasi untuk mengirimkan peringatan ketika ditemukan peningkatan kasus yang melebihi ambang batas yang telah ditetapkan, memungkinkan responss cepat terhadap potensi

wabah. Dashboard interaktif dapat menampilkan indikator kesehatan utama dalam format visual yang mudah dipahami oleh pembuat keputusan. Pelaporan berbasis web memudahkan akses informasi kesehatan oleh berbagai pihak yang berperan dalam kesehatan masyarakat, mendukung koordinasi antar sektor dalam pengendalian penyakit. Integrasi sistem informasi kesehatan dengan sistem komunikasi modern memungkinkan pertukaran informasi yang cepat antar fasilitas kesehatan dan lembaga terkait.

7.4.4 Tantangan dan Solusi Implementasi Sistem Informasi Kesehatan

Implementasi sistem informasi kesehatan menghadapi berbagai tantangan, mulai dari keterbatasan anggaran untuk pengadaan perangkat keras dan lunak, kurangnya keahlian tenaga IT di fasilitas kesehatan tingkat bawah, hingga resistensi dari petugas kesehatan terhadap perubahan cara bekerja. Masalah interoperabilitas antara berbagai sistem informasi kesehatan yang ada di fasilitas kesehatan yang berbeda juga menjadi hambatan dalam integrasi data. Solusi terhadap tantangan ini mencakup pelatihan berkelanjutan bagi tenaga kesehatan tentang penggunaan sistem informasi, standarisasi format data antar fasilitas kesehatan untuk memudahkan integrasi, dan pengembangan sistem informasi yang user-friendly sehingga mudah digunakan oleh petugas kesehatan dengan berbagai tingkat keahlian. Dukungan dari tingkat pemerintahan yang lebih tinggi, baik dari segi anggaran maupun

kebijakan, sangat penting untuk mendukung implementasi sistem informasi kesehatan yang berkelanjutan.

7.5 Implementasi Program Kesehatan Berbasis Biostatistik

Implementasi program kesehatan masyarakat yang efektif memerlukan perencanaan yang cermat berbasis data epidemiologi dan biostatistik, pelaksanaan yang disertai dengan monitoring berkelanjutan, dan evaluasi yang sistematis untuk mengukur dampak program.

7.5.1 Deteksi Dini dan Skrining Penyakit

Deteksi dini penyakit merupakan salah satu strategi penting dalam pengendalian penyakit tidak menular melalui program screening terhadap faktor risiko penyakit. Program Pos Pembinaan Terpadu Penyakit Tidak Menular atau Posbindu PTM merupakan contoh program berbasis komunitas yang menggunakan pendekatan biostatistik dalam deteksi dini penyakit seperti hipertensi dan diabetes. Posbindu PTM melakukan pengukuran tekanan darah, kadar glukosa, dan analisis faktor risiko pada masyarakat untuk mengidentifikasi kasus-kasus baru penyakit tidak menular yang sebelumnya tidak terdiagnosis. Data dari hasil skrining di Posbindu PTM dikumpulkan, dianalisis menggunakan metode statistik deskriptif, dan digunakan untuk merancang intervensi yang lebih tertarget. Evaluasi berkala terhadap pencapaian program menggunakan data statistik membantu dalam mengidentifikasi

hambatan dalam pelaksanaan program dan merencanakan perbaikan (Arsyad et al., 2025).

7.5.2 Surveilans Berbasis Komunitas

Surveilans berbasis komunitas atau *Community Based Surveillance* (CBS) adalah pendekatan participatory yang melibatkan masyarakat secara langsung dalam deteksi dan pelaporan kejadian penyakit. CBS menggunakan jaringan kader kesehatan yang ada di tingkat komunitas untuk melakukan pengamatan berkelanjutan terhadap penyakit dan kondisi kesehatan yang memengaruhi masyarakat. Data yang dikumpulkan melalui CBS dianalisis menggunakan metode biostatistik untuk mengidentifikasi pola penyakit, faktor risiko, dan kebutuhan intervensi kesehatan. Keuntungan CBS adalah keterlibatan masyarakat lokal yang meningkatkan kepercayaan terhadap informasi kesehatan dan memudahkan implementasi intervensi yang disesuaikan dengan konteks lokal. Namun, CBS juga menghadapi tantangan dalam hal konsistensi dan akurasi data yang dikumpulkan oleh kader yang memiliki latar belakang pendidikan dan pelatihan yang beragam. Untuk meningkatkan kualitas data CBS, perlu dilakukan pelatihan berkala, verifikasi data, dan integrasi dengan sistem surveilans formal (Wahyuni et al., 2026).

7.5.3 Manajemen Penelitian Lapangan dan Validasi Data

Penelitian lapangan yang dilakukan untuk evaluasi program kesehatan memerlukan manajemen data yang ketat untuk memastikan kualitas data dan validitas hasil penelitian.

Pengumpulan data dilakukan melalui wawancara, pengukuran antropometri, pengambilan sampel biologis, atau analisis dokumen kesehatan dengan prosedur yang terstandarisasi. Pelatihan enumerator atau pengumpul data dilakukan untuk memastikan konsistensi dalam pengumpulan data di berbagai lokasi penelitian. Validasi data dilakukan dengan memeriksa kelengkapan data, mengidentifikasi data yang tidak masuk akal atau outlier, dan melakukan spot check atau verifikasi ulang pada sebagian sampel data. Pembersihan data atau *data cleaning* dilakukan untuk mengoreksi kesalahan, memisahkan atau menghapus data yang tidak valid. Dokumentasi lengkap tentang prosedur pengumpulan data dan penanganan data dikerjakan untuk transparansi dan memudahkan reproduibilitas penelitian.

7.5.4 Diseminasi Hasil dan Penggunaan Data untuk Aksi

Diseminasi hasil analisis data dan temuan penelitian dilakukan untuk memastikan bahwa informasi yang dihasilkan dapat digunakan oleh berbagai pihak yang berperan dalam kesehatan masyarakat. Hasil analisis dikomunikasikan dalam berbagai format yang sesuai dengan audiens, mulai dari laporan teknis untuk profesional kesehatan, ringkasan untuk pembuat kebijakan, hingga infografis dan media sosial untuk masyarakat umum. Forum diskusi dengan berbagai stakeholder kesehatan diadakan untuk mendiskusikan temuan penelitian, mengidentifikasi hambatan dalam implementasi rekomendasi, dan merencanakan tindak lanjut. Penggunaan data untuk aksi kesehatan dipastikan melalui

mekanisme feedback, dimana hasil analisis dikembalikan ke fasilitas kesehatan dan lembaga yang berperan sebagai dasar untuk penyesuaian program kesehatan. Pembelajaran dari pengalaman implementasi program kesehatan didokumentasikan dan dibagikan untuk meningkatkan efektivitas program di tempat lain.

Latihan Soal

Soal 1

Sebuah puskesmas melakukan surveilans penyakit hipertensi dan menemukan peningkatan jumlah kasus dari 120 kasus pada bulan Januari menjadi 280 kasus pada bulan Februari. Berdasarkan data surveilans ini, jelaskan langkah-langkah apa yang harus dilakukan oleh puskesmas menggunakan pendekatan biostatistik untuk memverifikasi peningkatan kasus tersebut dan melakukan penyelidikan epidemiologi. Apa saja ukuran epidemiologi yang perlu dihitung dan analisis apa yang perlu dilakukan?

Soal 2

Seorang analisis epidemiologi ditugaskan untuk menganalisis data distribusi kasus penyakit demam berdarah dengue di suatu kabupaten. Data menunjukkan bahwa kasus terkonsentrasi di beberapa kecamatan tertentu dengan kepadatan penduduk tinggi, sanitasi buruk, dan akses layanan kesehatan terbatas. Jelaskan bagaimana Sistem Informasi Geografis dapat digunakan untuk memvisualisasikan data ini, dan bagaimana visualisasi spasial dapat

membantu dalam perencanaan intervensi pencegahan yang lebih efektif dan tertarget.

Soal 3

Program Posbindu PTM di suatu kelurahan melakukan screening kesehatan terhadap 500 warga dan menemukan 180 orang dengan tekanan darah tinggi dan 95 orang dengan kadar glukosa darah tinggi, dengan 60 orang yang memiliki kedua kondisi tersebut. Menggunakan metode biostatistik, hitunglah prevalensi hipertensi, prevalensi diabetes, dan prevalensi komorbiditas hipertensi dan diabetes. Jelaskan implikasi temuan ini bagi perencanaan program kesehatan di kelurahan tersebut.

Soal 4

Sebuah fasilitas kesehatan ingin mengevaluasi efektivitas program pencegahan penyakit menular melalui pendekatan berbasis komunitas. Data yang dikumpulkan dari kader kesehatan menunjukkan adanya variasi dalam akurasi dan konsistensi pelaporan kasus dari berbagai wilayah komunitas. Jelaskan faktor-faktor apa yang dapat memengaruhi kualitas data surveilans berbasis komunitas, dan rekomendasikan solusi untuk meningkatkan validitas dan keandalan data yang dikumpulkan melalui surveilans ini.

Soal 5

Sebuah dinas kesehatan akan mengimplementasikan sistem informasi kesehatan terintegrasi di semua fasilitas kesehatan di kabupaten mereka. Jelaskan manfaat yang dapat diperoleh dari

implementasi sistem informasi kesehatan terintegrasi dalam meningkatkan pemantauan kesehatan masyarakat dan pengambilan keputusan berbasis data. Diskusikan juga tantangan teknis dan organisasi yang mungkin dihadapi dalam implementasi dan berikan rekomendasi untuk mengatasinya.

DAFTAR PUSTAKA

- Iqbal, M. F. (2025). ANALISIS KEJADIAN HIPERTENSI DENGAN PENDEKATAN EPIDEMIOLOGI DESKRIPTIF PADA DATA SURVEILANS DAERAH KHUSUS IBUKOTA JAKARTA. *Jurnal Ilmiah Keperawatan*. <https://doi.org/10.52236/ih.v13i1.697>
- Wahyuni, S., Ramadhan, S., & Rubiah, E. (2026). Surveilans Epidemiologi Penyakit Akibat Bencana Banjir Gampong Kubu Peusangan Siblah Krueng, Alue Kuta Jangka Bireuen, Cot Ara Kuta Blang, Lueng Daneun Kabupaten Bireuen. *Jurnal Pendidikan Tambusai*. <https://doi.org/10.31004/jptam.v10i1.36848>
- Arsyad, W., Amijaya, N. T. A., Febrianti, R., & Soesilawati, N. (2025). KONTRIBUSI PROGRAM POSBINDU PTM TERHADAP DETEKSI DINI DAN PENGENDALIAN PENYAKIT TIDAK MENULAR DI INDONESIA: TINJAUAN LITERATUR 2019–2024. *Indonesian Journal of Health Research Innovation*. <https://doi.org/10.64094/zvh3qn93>
- Wignjadiputro, I., Susilarini, N. K., Praptiningsih, C. Y., Sariwati, E., Setiawaty, V., & Samaan, G. (2018). Surveillance for Severe Acute Respiratory Infection as one approach to enhance Global Health Security in Indonesia. *Health Science Journal of Indonesia*. <https://doi.org/10.22435/HSJI.V9I1.473>

BAB 8

APLIKASI BIOSTATISTIK DALAM EVALUASI PROGRAM KESEHATAN

Agus Erwin Ashari, SKM, M.Kes.

8.1 Pengantar Biostatistik

Biostatistik adalah cabang ilmu statistik terapan yang mengkhususkan diri pada pengumpulan, pengolahan, analisis, interpretasi, dan penyajian data dalam bidang kesehatan dan biologi untuk mendukung penelitian serta evaluasi program secara ilmiah. Biostatistik berperan sebagai alat kuantitatif yang membantu peneliti dan praktisi kesehatan dalam menganalisis data sehingga temuan penelitian dapat diinterpretasikan dengan cara yang valid dan dapat dipertanggungjawabkan, serta menjadi dasar bagi pengambilan keputusan berbasis bukti dalam kebijakan dan praktik kesehatan . Dalam konteks penelitian kesehatan, penerapan metode statistik ini mencakup perancangan studi yang kuat, pengujian hipotesis, serta pengukuran hubungan antara variabel, sehingga hasilnya dapat digunakan untuk menilai efektivitas intervensi, mengidentifikasi

pola kejadian penyakit, dan mengevaluasi program kesehatan secara objektif serta sistematis.

Biostatistik sangat penting dalam penelitian kesehatan karena membantu menjawab pertanyaan ilmiah yang melibatkan ketidakpastian dan variasi dalam data. Misalnya, dalam studi klinis, biostatistik digunakan untuk menentukan ukuran sampel yang tepat, memilih metode analisis yang sesuai, serta mengendalikan bias, sehingga kesimpulan yang dihasilkan akurat dan dapat digeneralisasi. Selain itu, biostatistik menjadi tulang punggung dalam pengembangan dan praktik kesehatan masyarakat karena memungkinkan peneliti untuk menguji hipotesis tentang hubungan antara faktor risiko dan hasil kesehatan, serta menentukan signifikansi statistik dari temuan-temuan tersebut (turn0search1). Hal ini membantu profesional kesehatan dan pembuat kebijakan mengambil keputusan yang lebih informasional dan efektif dalam merancang strategi klinis maupun intervensi kesehatan masyarakat.

Dalam evaluasi program kesehatan, biostatistik memainkan peran kunci untuk menilai efektivitas dan dampak intervensi yang diterapkan. Melalui teknik statistik seperti analisis varians, regresi, dan uji hipotesis, biostatistik memungkinkan evaluasi perubahan dalam indikator kesehatan sebelum dan sesudah pelaksanaan suatu program. Misalnya, analisis data longitudinal dapat digunakan untuk melihat tren kesehatan populasi dari waktu ke waktu, sedangkan model statistik lainnya dapat menilai kontribusi masing-masing faktor terhadap hasil kesehatan tertentu. Dengan demikian,

biostatistik tidak hanya memberikan angka atau statistik semata, tetapi juga memungkinkan interpretasi ilmiah terhadap fenomena yang terjadi, sehingga hasil evaluasi program menjadi lebih kredibel dan relevan terhadap tujuan kesehatan masyarakat.

Lebih jauh lagi, biostatistik juga membantu dalam pemantauan kesehatan masyarakat dan pengambilan keputusan kebijakan yang lebih luas. Penerapan metode kuantitatif ini memungkinkan identifikasi pola atau tren dalam data epidemiologi, seperti angka kejadian penyakit, mortalitas, atau prevalensi faktor risiko tertentu dalam populasi. Informasi semacam ini sangat penting bagi perencanaan program kesehatan, penetapan prioritas sumber daya, serta respons terhadap wabah penyakit atau masalah kesehatan lainnya. Dengan kemampuan untuk menangkap hubungan kompleks antar variabel dan memperkirakan efek intervensi, biostatistik berkontribusi secara langsung terhadap pengembangan kebijakan yang berbasis bukti dan strategi pencegahan penyakit yang efektif.

Secara keseluruhan, biostatistik sebagai disiplin memberikan kerangka kerja metodologis yang kuat bagi penelitian dan evaluasi di bidang kesehatan. Melalui penerapan metode statistik yang tepat, biostatistik memungkinkan perumusan kesimpulan yang valid dari data penelitian, penilaian efektivitas intervensi kesehatan, serta dukungan bagi pengambilan keputusan berbasis bukti yang menjadi landasan perbaikan kualitas pelayanan kesehatan dan peningkatan derajat kesehatan masyarakat secara menyeluruh.

8.2 Desain dan Pengumpulan Data

Desain dan pengumpulan data merupakan tahap krusial dalam setiap penelitian, karena kualitas desain dan metode pengumpulan menentukan validitas, reliabilitas, dan interpretabilitas hasil penelitian. Dalam konteks penelitian ilmiah, pemilihan desain studi dan teknik pengumpulan data harus disesuaikan dengan tujuan penelitian, sifat variabel, serta sumber daya yang tersedia. Pendekatan yang sistematis memastikan bahwa data yang diperoleh akurat, representatif, dan mendukung analisis statistik yang valid.

8.2.1 Desain Studi

Pemilihan desain studi harus mempertimbangkan jenis pertanyaan penelitian dan karakteristik populasi yang diteliti. Beberapa desain yang umum digunakan meliputi survei, kohort, eksperimental, dan observasional. Desain survei memungkinkan pengumpulan data dari sampel besar untuk menilai persepsi, perilaku, atau kondisi tertentu pada populasi target. Desain kohort atau longitudinal digunakan untuk memantau perubahan atau efek suatu intervensi dari waktu ke waktu. Desain eksperimental menekankan manipulasi variabel independen dan pengendalian variabel lain untuk menentukan hubungan sebab-akibat, sementara desain observasional memfokuskan pada pengamatan fenomena secara alami tanpa intervensi. Pemilihan desain yang tepat membantu memastikan bahwa data yang diperoleh relevan, akurat, dan mampu menjawab pertanyaan penelitian secara sistematis (Li, 2022).

8.2.2 Pengumpulan Data

Pengumpulan data dilakukan dengan menggunakan teknik yang sesuai untuk memperoleh informasi yang diperlukan. Metode yang umum digunakan mencakup kuesioner untuk mengukur persepsi dan sikap, wawancara untuk mendapatkan informasi mendalam dan kontekstual, pengukuran laboratorium untuk data fisiologis atau biomedis, serta catatan medis untuk data klinis historis. Setiap teknik pengumpulan memiliki kelebihan dan keterbatasan, sehingga pemilihan metode harus disesuaikan dengan tujuan dan jenis data yang diinginkan. Kombinasi metode, seperti *mixed-method*, sering digunakan untuk meningkatkan validitas data dan memberikan perspektif yang lebih komprehensif. Data yang diperoleh kemudian dapat dianalisis menggunakan metode biostatistik atau teknik analisis kuantitatif dan kualitatif lainnya untuk menghasilkan kesimpulan yang valid dan dapat diandalkan.

Beberapa teknik utama yang umum digunakan dalam pengumpulan data antara lain ;

(a) Observasi: Pengamatan langsung terhadap fenomena, perilaku, atau objek di lapangan secara sistematis. Kelebihan: Data lebih akurat karena mencatat fenomena sebagaimana adanya tanpa bergantung pada laporan subjek.

Kekurangan: Rentan terhadap subjektivitas pengamat (*halo effect*) jika tidak dilakukan secara terstruktur.

Alat: Lembar observasi (checklist), kamera, atau catatan lapangan

(b) Wawancara: Tanya jawab langsung untuk mendapatkan informasi mendalam dari responden. Kelebihan dari wawancara : Mendapatkan data mendalam, perspektif personal, dan memungkinkan klarifikasi langsung. Kekurangan dari wawancara: Memerlukan waktu lama dan keterampilan komunikasi yang baik.

Alat dari wawancara: Panduan wawancara (interview guide), alat perekam suara, atau aplikasi video conference seperti Zoom.

(c) Kuesioner/Angket: Instrumen berupa daftar pertanyaan tertulis yang diberikan kepada responden secara terstruktur untuk mengumpulkan data kuantitatif dari sampel besar. Kelebihan dari Kuesioner/Angket; Efisien untuk menjaring data dari banyak orang dalam waktu singkat. Kekurangan dari Kuesioner/Angket: Peneliti tidak bisa melihat reaksi langsung responden dan tingkat pengembalian terkadang rendah. Alat dari Kuesioner/Angket: Formulir kertas, Google Forms, atau Typeform

(d) Diskusi Kelompok Terfokus (FGD): Diskusi terarah untuk menggali pandangan kelompok tentang topik spesifik. □ Tujuan: Memperoleh pemahaman komprehensif melalui dinamika interaksi antar peserta. Alat: Panduan diskusi, moderator, dan alat perekam

(e) Studi Dokumentasi: Mengumpulkan data dari sumber yang sudah ada sebelumnya, seperti arsip, jurnal, atau laporan pemerintah. Contoh Sumber: Data Sensus BPS, Google Scholar untuk publikasi ilmiah, dan laporan tahunan perusahaan. Alat: Skema klasifikasi data atau format ringkasan dokumen

Perbandingan Teknik pengumpulan data

Teknik	Jenis Data	Karakteristik Utama
Kuesioner	Kuantitatif	Skala besar, statistik, cepat
Wawancara	Kualitatif	Mendalam, eksploratif, personal
Observasi	Kualitatif/Kuantitatif	Real-time, perilaku nyata
Dokumentasi	Sekunder	Data historis, efisien biaya

Dengan desain studi yang tepat dan pengumpulan data yang sistematis, penelitian dapat menghasilkan temuan yang kredibel, mendukung pengambilan keputusan berbasis bukti, serta memberikan kontribusi signifikan terhadap pengembangan ilmu pengetahuan dan praktik profesional.

8.3 Analisis Statistik

Analisis statistik merupakan bagian penting dalam penelitian kuantitatif untuk memahami data, mengidentifikasi pola, dan menguji hipotesis. Dalam konteks penelitian organisasi dan manajemen sumber daya manusia, analisis statistik digunakan untuk memberikan pemahaman mendalam tentang karakteristik populasi serta menarik kesimpulan yang dapat digeneralisasi dari sampel ke

populasi. Analisis ini umumnya dibagi menjadi dua kategori utama, yaitu statistik deskriptif dan statistik inferensial.

8.3.1 Statistik Deskriptif

Statistik deskriptif digunakan untuk menyajikan data secara sistematis dan ringkas, sehingga memudahkan pemahaman karakteristik populasi atau sampel. Teknik ini mencakup ukuran pemusatan, seperti rata-rata (*mean*), median, dan modus, yang menggambarkan nilai pusat dari data. Selain itu, ukuran sebaran, seperti simpangan baku (*standard deviation*), rentang, dan varians, memberikan informasi mengenai penyebaran data di sekitar nilai pusat. Visualisasi data, seperti tabel, grafik batang, histogram, dan diagram lingkaran, digunakan untuk menyajikan data secara visual sehingga pola, tren, atau anomali dalam data dapat lebih mudah dikenali. Statistik deskriptif memberikan fondasi untuk interpretasi awal sebelum dilakukan analisis lebih lanjut (Field, 2020).

8.3.2 Statistik Inferensial

Statistik inferensial digunakan untuk membuat generalisasi atau kesimpulan dari sampel penelitian ke populasi yang lebih luas. Metode ini melibatkan pengujian hipotesis, seperti uji *t*, ANOVA, atau chi-square, untuk menentukan apakah perbedaan atau hubungan yang diamati signifikan secara statistik. Analisis regresi digunakan untuk menilai pengaruh variabel independen terhadap variabel dependen, sedangkan analisis multivariat, termasuk regresi berganda, analisis faktor, dan *structural equation modeling (SEM)*, memungkinkan pengujian hubungan kompleks antarvariabel.

Statistik inferensial membantu peneliti mengidentifikasi pola sebab-akibat, menguji model teoritis, dan memberikan dasar untuk pengambilan keputusan berbasis bukti dalam konteks organisasi (Hair et al., 2021).

8.3.3 Integrasi Statistik Deskriptif dan Inferensial

Penerapan kedua jenis statistik ini secara terpadu memungkinkan peneliti memahami data secara menyeluruh. Statistik deskriptif memberikan gambaran awal yang jelas tentang karakteristik sampel, sementara statistik inferensial memungkinkan peneliti menarik kesimpulan yang berlaku untuk populasi dan menguji hubungan antarvariabel. Kombinasi ini memastikan analisis data yang valid, akurat, dan relevan dengan tujuan penelitian, sehingga mendukung pengambilan keputusan dan pengembangan strategi organisasi yang efektif.

8.4 Interpretasi Hasil dan Evaluasi Program

Evaluasi program kesehatan memerlukan interpretasi hasil yang cermat untuk memastikan efektivitas intervensi dan perumusan kebijakan berbasis bukti. Analisis biostatistik menjadi alat utama dalam menilai dampak program, mengidentifikasi faktor risiko, serta memberikan dasar pengambilan keputusan yang akurat dan terukur.

8.4.1 Penggunaan Analisis Biostatistik

Analisis biostatistik digunakan untuk mengolah data kuantitatif yang diperoleh dari program kesehatan, seperti parameter fisiologis, indikator perilaku, atau hasil survei kesehatan masyarakat.

Dengan teknik statistik, peneliti dapat menilai perbedaan sebelum dan sesudah intervensi, menguji signifikansi perubahan, serta menentukan hubungan antara faktor-faktor risiko dengan hasil kesehatan. Beberapa contoh ; Analisis Deskriptif (Gambaran Status) digunakan untuk meringkas data agar mudah dipahami, misalnya dalam pemetaan akses sanitasi. Contoh: Menghitung persentase rumah tangga yang memiliki jamban sehat di suatu desa sebelum dan sesudah program pemucuan. Indikator: Mean (rata-rata), median, dan proporsi (%) cakupan sanitasi layak.

Analisis Korelasi dan Regresi (Mencari Hubungan) Digunakan untuk menilai sejauh mana faktor lingkungan memengaruhi kesehatan masyarakat. Regresi Linear: Mengukur seberapa besar penurunan angka kasus diare jika cakupan air bersih ditingkatkan sebesar 10%. Korelasi: Menilai hubungan antara tingkat polusi udara (PM2.5) dengan frekuensi kekambuhan asma pada anak-anak di wilayah industri.

Analisis Spasial (Pemetaan Risiko), Menganalisis distribusi geografis untuk mengidentifikasi "hotspots" atau wilayah yang paling rentan. Contoh: Memetakan persebaran kasus DBD berdasarkan titik koordinat genangan air untuk menentukan wilayah prioritas pengasapan (*fogging*) atau pemberantasan sarang nyamuk

8.4.2 Evaluasi Efektivitas Program

Interpretasi hasil analisis biostatistik memungkinkan evaluasi efektivitas program secara objektif. Program yang menunjukkan perbaikan signifikan dalam indikator kesehatan dapat

dianggap berhasil, sedangkan program yang tidak memberikan perubahan yang berarti memerlukan peninjauan ulang. Evaluasi ini juga membantu mengidentifikasi komponen program yang efektif dan komponen yang memerlukan perbaikan, sehingga sumber daya dapat dialokasikan secara lebih efisien. Selain itu, hasil evaluasi dapat digunakan untuk mengukur pencapaian tujuan program jangka pendek maupun jangka panjang, termasuk peningkatan kualitas hidup dan pengurangan risiko penyakit di masyarakat.

Contoh Evaluasi efektivitas program STBM (Sanitasi Total Berbasis Masyarakat) bertujuan untuk memastikan keberhasilan 5 pilar sanitasi dalam memutus rantai penularan penyakit, maka rancangannya sebagai berikut :

1. Indikator Keberhasilan Berdasarkan 5 Pilar STBM

Anda dapat menilai efektivitas dengan melihat pencapaian pada setiap pilar.

Pilar STBM	Indikator (Output/Outcome)	Evaluasi
1. Stop BABS	100% masyarakat memiliki akses & menggunakan jamban sehat (Status ODF/Open Defecation Free).	
2. Cuci Tangan (CTPS)	Tersedianya sarana CTPS dengan air mengalir & sabun di setiap rumah.	
3. PAM RT	Masyarakat mengelola air minum dan makanan dengan aman (perebusan/klorinasi).	

4. Sampah RT	Tidak ada sampah berserakan di sekitar rumah; adanya sistem pembuangan yang benar.
5. Limbah Cair RT	Limbah cair domestik (bekas mandi/cuci) tidak tergenang dan dialirkan ke SPAL.

2. Teknik Pengumpulan Data & Instrumen

Untuk mendapatkan data yang akurat, gunakan kombinasi teknik berikut:

- Observasi Lapangan (Inpeksi Sanitasi): Gunakan *checklist* untuk melihat fisik jamban atau tempat sampah.
- Wawancara Mendalam: Menanyakan hambatan warga dalam membangun sarana sanitasi secara mandiri.
- Analisis Data Sekunder: Mengambil data dari Aplikasi E-Monev STBM milik Kemenkes untuk melihat tren kenaikan akses.

3. Contoh Rubrik Penilaian Efektivitas (Skoring)

Gunakan skor 1-4 untuk menentukan status efektivitas program di suatu desa:

Skor	Kategori	Keterangan
4	Sangat Efektif	Desa sudah deklarasi ODF dan 5 pilar STBM diterapkan > 90% warga.

3	Efektif	Desa sudah ODF, namun pilar 2-5 masih dalam proses peningkatan (60-89%).
2	Kurang Efektif	Belum ODF, partisipasi masyarakat rendah, sarana fisik banyak yang rusak.
1	Tidak Efektif	Tidak ada perubahan perilaku, warga tetap BABS meskipun sudah dipicu

4. Analisis Biostatistik yang Digunakan

Untuk membuktikan efektivitas secara ilmiah, Anda bisa melakukan:

- Uji McNemar: Untuk melihat perubahan perilaku (Misal: "BABS" vs "Tidak BABS") pada subjek yang sama sebelum dan sesudah intervensi program.
- Uji Regresi Logistik: Jika Anda ingin melihat faktor apa yang paling memengaruhi keberhasilan STBM (misal: apakah faktor ekonomi atau tingkat pendidikan petugas pemicu).

8.4.3 Rekomendasi Kebijakan Berbasis Bukti

Hasil evaluasi program menjadi dasar untuk merumuskan rekomendasi kebijakan berbasis bukti (*evidence-based policy*). Data yang diperoleh melalui analisis biostatistik memberikan informasi yang jelas mengenai faktor risiko, efektivitas intervensi, dan

kebutuhan prioritas masyarakat. Rekomendasi ini dapat berupa penyesuaian prosedur, pengembangan program tambahan, atau alokasi sumber daya yang lebih optimal. Pendekatan berbasis bukti memastikan bahwa kebijakan yang diterapkan relevan, efisien, dan berdampak positif terhadap kesehatan masyarakat secara luas (Higgins et al., 2021).

Secara keseluruhan, interpretasi hasil dan evaluasi program berbasis biostatistik merupakan elemen krusial dalam pengelolaan program kesehatan. Pendekatan ini memungkinkan pengambilan keputusan yang akurat, peningkatan efektivitas intervensi, dan perumusan kebijakan yang mendukung kesejahteraan masyarakat secara berkelanjutan.

Latihan Soal (Esai)

1. Jelaskan peran biostatistik dalam evaluasi program kesehatan.
2. Apa perbedaan utama antara statistik deskriptif dan inferensial?
3. Sebutkan beberapa desain penelitian yang umum digunakan dalam evaluasi program kesehatan.
4. Bagaimana hasil analisis biostatistik dapat memengaruhi kebijakan kesehatan masyarakat?
5. Diskusikan tantangan pengumpulan data yang valid dan reliabel dalam penelitian kesehatan.

DAFTAR PUSTAKA

- Besral, S. (2023). Peran biostatistik dalam kesehatan masyarakat dan kebijakan berbasis bukti. Fakultas Kesehatan Masyarakat Universitas Indonesia.
- Bland, J. M., & Altman, D. G. (2020). Statistics notes: Biostatistical methods in healthcare research. *BMJ*, 369, m1589.
- Creswell, J. W., & Creswell, J. D. (2018). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Thousand Oaks, CA: Sage Publications.
- Dahlan, M. S. (2014). *Statistik untuk Kedokteran dan Kesehatan: Deskriptif, Bivariat, dan Multivariat*. Jakarta: Epidemiologi Indonesia.
- Daniel, W. W., & Cross, C. L. (2018). *Biostatistics: A Foundation for Analysis in the Health Sciences*. Wiley.
- Field, A. (2020). *Discovering statistics using IBM SPSS statistics* (5th ed.). Sage Publications.
- Gore, A. D. (2024). The evolving role of biostatistics in medical research. *Journal of Health Sciences Research*.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2021). *Multivariate data analysis* (8th ed.). Cengage Learning.
- Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2021). *Cochrane handbook for systematic reviews of interventions*. John Wiley & Sons.

- Indriani, D., dkk. (2019). "Analisis Faktor yang Berhubungan dengan Efektivitas Program STBM Pilar 1 di Wilayah Perdesaan". Jurnal Kesehatan Lingkungan Indonesia (JKLI). JKLI Undip.
- Li, X. (2022). Integrating digital tools in science education: Enhancing student engagement and understanding. *Journal of Science Education and Technology*, 31(4), 487–499.
- Kementerian Kesehatan RI. (2014). Peraturan Menteri Kesehatan Nomor 3 Tahun 2014 tentang Sanitasi Total Berbasis Masyarakat
- Kementerian Kesehatan RI. (2020). Pedoman Monitoring dan Evaluasi STBM. Jakarta: Direktorat Jenderal Pencegahan dan Pengendalian Penyakit (P2P).
- Moleong, L. J. (2017). *Metodologi Penelitian Kualitatif*. Bandung: Remaja Rosdakarya.
- Sugiyono. (2019). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- Prahasta, E. (2009). *Sistem Informasi Geografis: Konsep-Konsep Dasar*. Bandung: Informatika.
- Prawesti, D. (2021). "Evaluasi Dampak Program STBM terhadap Penurunan Angka Penyakit Diare". *Journal of Public Health Research*.

BAB 9

ETIKA DAN VALIDITAS DATA DALAM BIOSTATISTIK

Intan Pujilestari, SST., M.K.M.

9.1 Fondasi Etika dalam Penelitian Biostatistik

Etika dalam penelitian biostatistik merupakan prinsip fundamental yang memastikan integritas ilmiah, melindungi hak partisipan, dan menjaga kepercayaan publik terhadap komunitas peneliti dan ilmuwan. Dalam konteks biostatistik, etika mencakup tidak hanya dimensi moral pribadi, tetapi juga tanggung jawab profesional untuk mengumpulkan, menganalisis, dan melaporkan data dengan kejujuran dan akuntabilitas penuh. Pemahaman mendalam tentang etika penelitian sangat penting bagi para peneliti, biostatistisien, dan profesional kesehatan karena keputusan etis dalam menangani data dapat berdampak signifikan pada kualitas temuan penelitian dan kesejahteraan masyarakat yang menjadi subjek penelitian.

9.1.1 Prinsip-Prinsip Dasar Etika Biomedis

Etika biomedis dibangun atas empat prinsip fundamental yang menjadi pedoman dalam melaksanakan penelitian kesehatan. Pertama, prinsip otonomi atau *autonomy* mengakui hak setiap individu untuk membuat keputusan tentang partisipasi mereka dalam penelitian tanpa paksaan atau ancaman. Prinsip ini memerlukan informed consent yang jelas dan transparan sebelum seseorang berpartisipasi dalam penelitian. Kedua, prinsip keuntungan atau *beneficence* mendorong peneliti untuk merancang dan melaksanakan penelitian yang dapat memberikan manfaat nyata bagi partisipan atau masyarakat luas. Ketiga, prinsip tidak merugikan atau *non-maleficence* memastikan bahwa penelitian dirancang dengan sebaik mungkin untuk meminimalkan risiko dan potensi kerugian bagi partisipan. Keempat, prinsip keadilan atau *justice* mensyaratkan bahwa beban dan manfaat penelitian didistribusikan secara adil di antara berbagai kelompok populasi, menghindari eksploitasi kelompok rentan atau minoritas (Khairi et al., 2025).

9.1.2 Integritas Peneliti dan Kepercayaan Ilmiah

Integritas peneliti merupakan fondasi dari kepercayaan publik terhadap sains dan penelitian. Integritas ini mencerminkan komitmen peneliti untuk melaporkan hasil penelitian secara jujur, mengakui keterbatasan penelitian, dan tidak memanipulasi atau memalsukan data untuk mencapai hasil yang diinginkan. Peneliti yang memiliki integritas tinggi akan transparan dalam

mengomunikasikan metodologi penelitian, sumber pendanaan, dan potensi konflik kepentingan. Kepercayaan ilmiah yang dibangun melalui integritas peneliti sangat penting karena kebijakan publik, praktik klinis, dan keputusan penting lainnya sering didasarkan pada temuan penelitian. Ketika integritas peneliti diragukan atau terbukti telah melakukan pelanggaran, dampaknya melampaui penelitian individual tersebut dan dapat mengguncang kepercayaan publik terhadap institusi penelitian dan komunitas ilmuwan secara keseluruhan (Putra, 2025).

9.1.3 Peran Komite Etika Penelitian

Komite etika penelitian atau *Institutional Review Board* (IRB) memiliki peran penting dalam menjaga standar etika penelitian di institusi pendidikan dan penelitian. Komite ini bertugas untuk meninjau protokol penelitian sebelum dimulai, memastikan bahwa penelitian dirancang sesuai dengan prinsip etika, risiko-manfaat seimbang, dan prosedur perlindungan partisipan memadai. Proses review oleh komite etika mencakup evaluasi informed consent, identifikasi dan mitigasi risiko penelitian, perlindungan privasi dan kerahasiaan data, serta memastikan perlakuan yang adil terhadap semua partisipan. Keberadaan dan efektivitas komite etika penelitian yang independen dan berkompeten merupakan indikator penting dari komitmen institusi terhadap standar etika penelitian (Pramudiani et al., 2025).

9.1.4 Perlindungan Data Partisipan dan Kerahasiaan

Perlindungan data pribadi partisipan merupakan elemen kritis dalam penelitian yang melibatkan manusia. Peneliti berkewajiban untuk menjaga kerahasiaan informasi pribadi yang dikumpulkan selama penelitian, menggunakan identifier yang tidak dapat diidentifikasi, dan membatasi akses ke data hanya kepada anggota tim penelitian yang berhak. Perlindungan data mencakup penggunaan sandi yang kuat untuk database elektronik, penyimpanan fisik yang aman untuk dokumen berbasis kertas, dan protokol penghapusan data yang jelas setelah penelitian selesai. Di era digital, isu keamanan data menjadi semakin kompleks dengan risiko pelanggaran dan akses tidak sah terhadap data yang sensitif. Peneliti harus menerapkan standar keamanan data yang ketat dan memahami regulasi perlindungan data yang berlaku, seperti Undang-Undang Perlindungan Data Pribadi di berbagai negara.

9.2 Validitas Data dalam Konteks Biostatistik

Validitas data mengacu pada akurasi dan ketepatan dari pengukuran yang dilakukan, memastikan bahwa instrumen pengumpulan data benar-benar mengukur apa yang ingin diukur. Dalam biostatistik, validitas data sangat penting karena data yang tidak valid akan menghasilkan kesimpulan yang menyesatkan dan rekomendasi yang tidak tepat untuk praktik klinis atau kebijakan kesehatan. Validitas data dibangun melalui kombinasi dari validitas

konstruk, validitas isi, dan validitas kriteria yang semuanya berkontribusi pada keandalan dan keterandalan temuan penelitian.

9.2.1 Validitas Isi dan Validitas Konstruk

Validitas isi mengacu pada sejauh mana instrumen pengumpulan data mencakup semua elemen penting dari konstruk yang diukur dan tidak mencakup elemen yang tidak relevan. Pengujian validitas isi melibatkan review oleh para ahli di bidang terkait untuk menilai apakah pertanyaan atau item dalam instrumen secara akurat mencerminkan domain yang diukur. Validitas konstruk lebih luas dan mencakup pengujian apakah instrumen benar-benar mengukur konstruk teoritis yang ingin diukur, bukan sekadar atribut lain yang terkait. Pengujian validitas konstruk dapat dilakukan melalui analisis faktor, perbandingan dengan instrumen lain yang sudah tervalidasi, atau melalui studi multitrait-multimethod (Jurnal et al., 2025).

9.2.2 Validitas Kriteria dan Pengujian Empiris

Validitas kriteria mengevaluasi sejauh mana skor instrumen dapat memprediksi hasil pada kriteria eksternal tertentu. Validitas kriteria dapat berupa validitas prediktif, yang menilai apakah instrumen dapat memprediksi hasil di masa depan, atau validitas konkuren, yang menilai apakah instrumen berkorelasi dengan hasil yang diukur pada waktu yang sama. Pengujian empiris terhadap validitas kriteria melibatkan analisis korelasi antara skor instrumen dengan variabel kriteria, serta analisis regresi untuk menentukan kekuatan prediksi instrumen. Hasil pengujian validitas memberikan

bukti tentang kualitas instrumen dan dapat menginformasikan keputusan apakah instrumen layak digunakan dalam penelitian atau praktik klinis (Yarsasi et al., 2025).

9.2.3 Strategi Peningkatan Validitas Data

Peningkatan validitas data dapat dilakukan melalui berbagai strategi metodologis yang dirancang untuk mengurangi kesalahan pengukuran dan bias. Pertama, penggunaan instrumen yang sudah tervalidasi sebelumnya atau adaptasi instrumen yang sudah terbukti valid dapat menghemat waktu dan meningkatkan kepercayaan pada validitas instrumen. Kedua, pelatihan intensif kepada pengumpul data atau enumerator memastikan konsistensi dalam prosedur pengumpulan data di berbagai lokasi dan waktu. Ketiga, verifikasi data dan *data cleaning* yang sistematis mengidentifikasi dan mengoreksi kesalahan pencatatan atau data yang implausibel. Keempat, triangulasi data dari berbagai sumber atau metode dapat meningkatkan validitas hasil dengan memungkinkan cross-checking informasi dari perspektif yang berbeda (Audina & Albina, 2025).

9.2.4 Validitas dalam Konteks Teknologi Digital dan Big Data

Era digital dan perkembangan *big data* menghadirkan tantangan baru terhadap validitas data yang memerlukan pendekatan etika dan metodologis yang baru. Data yang dikumpulkan secara otomatis dari sistem digital atau media sosial mungkin tidak sepenuhnya akurat atau representatif, dapat mengandung *bias* algoritma, dan mungkin mengalami masalah dalam integritas data. Penggunaan kecerdasan buatan atau *Artificial*

Intelligence (AI) dalam analisis data menimbulkan pertanyaan tentang transparansi, akuntabilitas, dan validitas hasil yang dihasilkan oleh sistem komputasional yang kompleks. Peneliti yang menggunakan teknologi digital dan *big data* harus tetap kritis terhadap kualitas data, memahami keterbatasan dari sumber data, dan memastikan bahwa temuan yang dihasilkan tetap valid dan bermakna dalam konteks praktis (Putra, 2025).

9.3 Reliabilitas Data dan Konsistensi Pengukuran

Reliabilitas mengacu pada konsistensi atau stabilitas dari pengukuran, artinya instrumen yang reliabel akan menghasilkan hasil yang serupa ketika digunakan berulang kali pada kondisi yang sama. Reliabilitas data sangat penting dalam penelitian karena data yang tidak reliabel akan menghasilkan kesimpulan yang tidak stabil dan tidak dapat direplikasi oleh peneliti lain. Reliabilitas data diuji melalui berbagai metode yang disesuaikan dengan jenis instrumen dan desain penelitian.

9.3.1 Konsistensi Internal dan Reliabilitas Uji Ulang

Konsistensi internal mengukur sejauh mana item-item dalam instrumen berkorelasi satu sama lain dan mengukur konstruk yang sama. Pengujian konsistensi internal dilakukan dengan menghitung koefisien Cronbach's Alpha, yang berkisar dari 0 hingga 1, dengan nilai di atas 0,70 umumnya dianggap menunjukkan konsistensi internal yang baik. Reliabilitas uji ulang atau *test-retest reliability* mengevaluasi apakah instrumen menghasilkan hasil yang

serupa ketika digunakan pada waktu yang berbeda pada partisipan yang sama. Reliabilitas uji ulang penting untuk instrumen yang dirancang untuk mengukur karakteristik yang relatif stabil dari waktu ke waktu, seperti kepribadian atau sikap (Khairi et al., 2025).

9.3.2 Reliabilitas Antar-Penilai dan Standardisasi Pengukuran

Reliabilitas antar-penilai atau *inter-rater reliability* menjadi penting ketika data dikumpulkan oleh lebih dari satu pengumpul data atau ketika ada keputusan subjektif dalam penilaian hasil pengukuran. Koefisien Cohen's Kappa atau intraclass correlation coefficient (ICC) digunakan untuk mengevaluasi keselarasan antara berbagai penilai. Untuk meningkatkan reliabilitas antar-penilai, dilakukan standardisasi dalam pelatihan pengumpul data, pengembangan panduan penilaian yang jelas, dan periodic inter-rater agreement checks selama pengumpulan data berlangsung (Yarsasi et al., 2025).

9.3.3 Analisis Item dan Identifikasi Item Bermasalah

Analisis item dilakukan untuk mengevaluasi kualitas setiap item dalam instrumen pengumpulan data, termasuk mengidentifikasi item yang memiliki tingkat kesulitan ekstrem, daya pembeda rendah, atau kontribusi minimal terhadap reliabilitas instrumen secara keseluruhan. Item-item bermasalah dapat diidentifikasi melalui analisis item-total correlation, analisis faktor, atau analisis Rasch. Item yang bermasalah dapat direvisi, dihapus, atau diganti dengan item baru yang lebih baik (ANALISIS KUALITAS BUTIR SOAL UJIAN AKHIR SEMESTER MATA PELAJARAN AKIDAH

AKHLAK KELAS V MI RADHIATUL ADAWIYAH MAKASSAR, 2023).

9.3.4 Dokumentasi Prosedur Pengumpulan Data

Dokumentasi yang komprehensif tentang prosedur pengumpulan data, protokol standar, dan instruksi kepada pengumpul data sangat penting untuk memastikan reliabilitas dan validitas data. Dokumentasi harus mencakup deskripsi detail tentang setiap langkah pengumpulan data, kriteria inklusi dan eksklusi partisipan, definisi operasional dari setiap variabel yang diukur, dan prosedur handling dan penyimpanan data. Dokumentasi yang baik memudahkan audit dan memungkinkan peneliti lain untuk mereplikasi penelitian atau memahami potensi sumber kesalahan dalam pengumpulan data.

9.4 Manajemen Integritas Data dan Pelaporan Etis

Manajemen integritas data mencakup semua aspek dari pengumpulan data hingga pelaporan hasil, dengan tujuan memastikan bahwa data tetap akurat, lengkap, dan aman dari ancaman eksternal atau kesalahan internal. Integritas data juga mencerminkan komitmen peneliti untuk melaporkan hasil penelitian secara etis tanpa manipulasi, pemalsuan, atau menyembunyian temuan yang tidak menguntungkan.

9.4.1 Pelanggaran Integritas Data dan Tanda-Tandanya

Pelanggaran integritas data dapat berupa manipulasi data, fabrikasi atau penciptaan data palsu, plagiarisme dalam pelaporan

hasil, atau penyembunyian konflik kepentingan. Tanda-tanda pelanggaran integritas data dapat mencakup: nilai data yang tidak wajar atau tidak konsisten dengan distribusi yang diharapkan, missing data pattern yang mencurigakan, atau hasil yang terlalu sempurna untuk menjadi kenyataan. Sistem deteksi anomali berbasis statistik dapat membantu mengidentifikasi data yang mencurigakan untuk penyelidikan lebih lanjut (Pramudiani et al., 2025).

9.4.2 Audit Data dan Verifikasi Independen

Audit data dilakukan untuk memverifikasi akurasi, kelengkapan, dan konsistensi data yang telah dikumpulkan. Audit dapat dilakukan oleh pihak independen yang tidak terlibat langsung dalam pengumpulan data, yang memberikan perspektif objektif. Proses audit mencakup spot check pada sampel data asli, verifikasi terhadap sumber data atau dokumen pendukung, dan check konsistensi antar data dalam database. Verifikasi independen dapat meningkatkan kepercayaan pada kualitas data dan mendeteksi kesalahan atau anomali yang mungkin terlewatkan dalam proses pengumpulan data awal (Khairi et al., 2025).

9.4.3 Transparansi dalam Pelaporan Hasil dan Analisis

Transparansi dalam pelaporan hasil penelitian memerlukan peneliti untuk jujur mengomunikasikan temuan, termasuk hasil yang tidak signifikan atau berlawanan dengan hipotesis awal. Pelaporan yang transparan harus mencakup deskripsi detail tentang metodologi penelitian, termasuk jumlah partisipan yang diundang versus yang benar-benar berpartisipasi, tingkat dropout atau data missing, dan

analisis sensitivitas yang menunjukkan robustness dari temuan. Peneliti harus menghindari selective reporting atau cherry-picking hasil yang menguntungkan sambil menyembunyikan hasil yang tidak menguntungkan (Putra, 2025).

9.4.4 Penanganan Konflik Kepentingan dan Transparansi

Pendanaan

Konflik kepentingan dapat terjadi ketika peneliti memiliki kepentingan finansial atau personal yang dapat memengaruhi keputusan penelitian atau interpretasi hasil. Peneliti berkewajiban untuk mengungkapkan semua potensi konflik kepentingan, termasuk hubungan finansial dengan perusahaan farmasi, pemilik kepentingan dalam produk yang sedang diteliti, atau tekanan untuk menghasilkan hasil tertentu. Transparansi tentang sumber pendanaan penelitian, apakah dari pemerintah, industri, atau fondasi, penting untuk memungkinkan pembaca mengevaluasi potensi bias dalam penelitian (Muhammad et al., 2024).

9.5 Isu Etika Kontemporer dalam Biostatistik dan Penelitian

Kesehatan

Perkembangan teknologi dan perubahan konteks sosial menghadirkan isu etika baru yang memerlukan perhatian dari peneliti dan komunitas biostatistik modern.

9.5.1 Etika Penggunaan Data Genomik dan Privasi Genetik

Penelitian genomik menghadirkan tantangan etika unik terkait privasi genetik dan potensi diskriminasi berdasarkan

informasi genetik. Data genomik tidak hanya memberikan informasi tentang individu yang berpartisipasi dalam penelitian, tetapi juga mengandung informasi tentang keluarga dan keturunan mereka. Oleh karena itu, informed consent untuk penelitian genomik harus mencakup diskusi tentang penggunaan data di masa depan, berbagi data dengan peneliti lain, dan potensi implikasi bagi keluarga. Peneliti harus mengimplementasikan safeguards yang ketat untuk melindungi privasi data genomik dan memastikan bahwa data tidak digunakan untuk tujuan yang tidak disetujui atau untuk diskriminasi (Silaban, 2025).

9.5.2 Etika Penelitian yang Melibatkan Kelompok Rentan

Kelompok rentan seperti anak-anak, individu dengan gangguan kognitif, kelompok minoritas, dan populasi yang kurang beruntung secara ekonomi memerlukan perlindungan etika khusus dalam penelitian. Untuk kelompok ini, informed consent mungkin perlu melibatkan wali atau representative yang bertindak atas nama partisipan, dan ada kebutuhan tambahan untuk memastikan bahwa partisipasi dalam penelitian tidak akan mengakibatkan eksploitasi atau peningkatan beban yang tidak seimbang. Peneliti harus secara aktif mengevaluasi apakah manfaat penelitian akan tersebar secara adil di antara berbagai kelompok populasi (Pramudiani et al., 2025).

9.5.3 Etika Penggunaan Artificial Intelligence dalam Analisis Data Biostatistik

Penggunaan AI dan machine learning dalam analisis data biostatistik menghadirkan isu etika tentang transparansi,

akuntabilitas, dan keadilan algoritmik. Model AI yang kompleks sering bertindak sebagai *black box* di mana sulit untuk memahami bagaimana keputusan atau prediksi dibuat, menciptakan tantangan dalam mendokumentasikan alasan di balik keputusan klinis atau administratif berdasarkan model tersebut. Bias algoritma dapat memperkuat diskriminasi yang ada jika data pelatihan berisi bias atau tidak representatif dari populasi yang lebih luas. Peneliti menggunakan AI harus transparan tentang keterbatasan model, melakukan analisis yang ketat terhadap potensi bias, dan memastikan bahwa hasil model dapat dijelaskan dan diinterpretasi oleh pengguna akhir (Putra, 2025).

9.5.4 Etika Pembagian Data dan Open Science

Gerakan *open science* mendorong peneliti untuk berbagi data dan hasil penelitian secara terbuka untuk mempercepat kemajuan ilmiah dan meningkatkan reproducibility. Namun, pembagian data menghadirkan tantangan etika tentang privasi partisipan, informed consent untuk penggunaan data sekunder, dan potensi misuse data. Peneliti harus menyelimbungi mekanisme untuk berbagi data secara etis, seperti deidentifikasi data yang ketat, restricted access untuk data sensitif, dan pengembangan data sharing agreements yang jelas tentang penggunaan data yang diperbolehkan (Pramudiani et al., 2025).

Latihan Soal

Soal 1

Seorang peneliti merencanakan studi longitudinal untuk mengukur hubungan antara faktor risiko lingkungan dan perkembangan penyakit respiratori pada anak-anak. Peneliti akan mengumpulkan data respiratori dari 500 anak selama 5 tahun dengan pengukuran berulang setiap 6 bulan. Jelaskan pertimbangan etika apa yang harus dipertimbangkan dalam desain penelitian ini, terutama berkaitan dengan informed consent berkelanjutan, perlindungan data sensitif anak-anak, dan manajemen konflik kepentingan jika penelitian didanai oleh perusahaan yang memproduksi sumber potensial polusi.

Soal 2

Sebuah rumah sakit berencana untuk menggunakan model machine learning yang dilatih pada data pasien historis untuk memprediksi risiko infeksi nosokomial dan mengidentifikasi pasien yang memerlukan intervensi pencegahan tambahan. Diskusikan isu-isu etika terkait penggunaan AI dalam prediksi medis, termasuk transparansi model, potensi bias algoritma, informed consent untuk penggunaan data dalam machine learning, dan bagaimana memastikan bahwa algoritma tidak menimbulkan diskriminasi terhadap kelompok pasien tertentu.

Soal 3

Tim peneliti mengumpulkan data dari 1000 responden melalui survei online tentang perilaku kesehatan dan riwayat

penyakit. Setelah pengumpulan data, peneliti menemukan bahwa 15 persen data menunjukkan pola yang tidak wajar dan tidak konsisten dengan distribusi yang diharapkan. Jelaskan prosedur validasi dan verifikasi data apa yang harus dilakukan untuk mengidentifikasi penyebab anomali ini, bagaimana membedakan antara data yang benar-benar valid tetapi unusual dengan data yang manipulasi atau cacat, dan keputusan apa yang harus dibuat tentang data yang bermasalah ini.

Soal 4

Sebuah organisasi penelitian yang bekerja sama dengan universitas dan industri farmasi merencanakan uji klinis untuk obat baru. Jelaskan bagaimana transparansi tentang sumber pendanaan, konflik kepentingan, dan manajemen integritas data harus diterapkan dalam penelitian ini. Apa peran yang harus dimainkan oleh Komite Etika Penelitian Independen, dan bagaimana memastikan bahwa pelaporan hasil akan etis terlepas dari apakah obat terbukti efektif atau tidak?

Soal 5

Sebuah database kesehatan nasional berencana untuk menjadikan data health records yang deidentifikasi tersedia untuk peneliti akademik dan perusahaan teknologi kesehatan untuk keperluan penelitian dan pengembangan. Analisis pertimbangan etika terkait sharing data dalam konteks ini, termasuk: (a) bagaimana memastikan deidentifikasi yang cukup untuk melindungi privasi partisipan sambil mempertahankan utilitas data untuk penelitian, (b)

bagaimana mengelola informed consent retrospektif untuk penggunaan data sekunder, (c) bagaimana menyeimbangkan antara berbagi data terbuka dengan perlindungan privasi, dan (d) mekanisme governance apa yang perlu ditetapkan untuk mengawasi penggunaan data.

DAFTAR PUSTAKA

- Khairi, M. F., Hidayat, Mhd. N., Sismita, N., Nurhayati, N., & Herlinda, H. (2025). Evaluasi Keabsahan Dan Keandalan Instrumen Penelitian. *Jurnal Ilmiah Dikdaya*. <https://doi.org/10.33087/dikdaya.v15i2.891>
- Putra, R. (2025). Big Data dan Artificial Intelligence dalam Produksi Pengetahuan Ilmiah atas Tantangan Metodologis. *Jurnal Riset Manajemen Komunikasi*. <https://doi.org/10.29313/jrmk.v5i2.8721>
- Pramudiani, D., Hanifah, A., Astria, R., & Saputra, E. (2025). Refleksi Kritis dan Analisis Etika Penelitian Pendidikan. *Jurnal ISO: Jurnal Ilmu Sosial, Politik Dan Humaniora*. <https://doi.org/10.53697/iso.v6i1.3210>
- Jurnal, J. P. Q., Pendidikan, S. S. H., Zayrin, A. A., Nupus, H., Maizia³, K. K., Marsela, S., Hidayatullah, R., & Harmonedi. (2025). Analisis Instrumen Penelitian Pendidikan (Uji Validitas Dan Relibilitas Instrumen Penelitian). *Jurnal QOSIM Jurnal Pendidikan Sosial & Humaniora*. <https://doi.org/10.61104/jq.v3i2.1070>
- Yarsasi, S., Tahyudin, I., & Hariguna, T. (2025). Analisis Validitas dan Reliabilitas Kuesioner dengan Metode Partial Least Squares Structural Equation Modeling pada Aplikasi SMARTPLS. *Jurnal Pendidikan Dan Teknologi Indonesia*. <https://doi.org/10.52436/1.jpti.885>

ANALISIS KUALITAS BUTIR SOAL UJIAN AKHIR SEMESTER MATA PELAJARAN AKIDAH AKHLAK KELAS V MI RADHIATUL ADAWIYAH MAKASSAR. (2023). ANALISIS KUALITAS BUTIR SOAL UJIAN AKHIR SEMESTER MATA PELAJARAN AKIDAH AKHLAK KELAS V MI RADHIATUL ADAWIYAH MAKASSAR. Inspiratif Pendidikan. <https://doi.org/10.24252/ip.v12i1.32040>

Muhammad, F., Satria, I., & Astuti, T. (2024). Pengaruh Kompetensi, Integritas dan Independensi Auditor terhadap Kualitas Audit. Relevan : Jurnal Riset Akuntansi. <https://doi.org/10.35814/relevan.v5i1.7703>

Silaban, S. (2025). Hubungan Pemahaman Etika dengan Praktik Profesional: Studi Empiris pada Mahasiswa D3 Optometri. Jurnal Kesehatan Ilmiah Indonesia (Indonesian Health Scientific Journal). <https://doi.org/10.51933/health.v10i2.2448>

BAB 10

PEMANFAATAN

TEKNOLOGI DIGITAL

DALAM ANALISIS

BIOSTATISTIK KESEHATAN

Izal Puji Santoso, SKM., M.Kes.

10.1 Pengenalan Teknologi Digital dalam Analisis Biostatistik Kesehatan

10.1.1 Evolusi Teknologi Digital dalam Kesehatan

Teknologi digital telah mengubah secara fundamental cara data biostatistik dikumpulkan, dianalisis, dan dimanfaatkan dalam sistem kesehatan masyarakat modern. Perkembangan revolusioner dalam teknologi informasi dan komunikasi, termasuk internet, komputasi cloud (*cloud computing*), dan kecerdasan buatan (*artificial intelligence*), telah membuka peluang baru yang sebelumnya tidak terbayangkan untuk meningkatkan kualitas analisis data kesehatan (Arifian, 2025). Sebelum era digital, analisis biostatistik kesehatan masyarakat dilakukan secara manual dengan

menggunakan alat-alat sederhana seperti kalkulator mekanik dan lembar kerja kertas, proses yang sangat memakan waktu, rentan terhadap kesalahan, dan terbatas dalam kapasitasnya untuk menangani dataset besar.

Evolusi teknologi digital dimulai dari penggunaan komputer personal (*personal computers*) pada tahun 1980-an dan 1990-an, yang memungkinkan peneliti untuk menggunakan software statistik khusus seperti SPSS, SAS, dan Stata untuk analisis data yang lebih cepat dan akurat (Kotlovskiy et al., 2024). Kemudian, perkembangan internet dan teknologi web pada tahun 2000-an memungkinkan berbagi data dan kolaborasi antar peneliti secara real-time melintasi batas geografis. Era modern yang dimulai sekitar tahun 2010 ditandai dengan munculnya *big data analytics*, *cloud computing*, dan aplikasi kecerdasan buatan yang memungkinkan analisis data pada skala yang jauh lebih besar dan dengan kecepatan yang luar biasa cepat.

10.1.2 Peran dan Manfaat Teknologi Digital

Peran teknologi digital dalam analisis biostatistik kesehatan sangat komprehensif dan mencakup setiap aspek dari siklus hidup data. Pertama, teknologi digital memfasilitasi pengumpulan data yang lebih efisien melalui sistem informasi kesehatan (*health information systems*) yang terintegrasi, aplikasi mobile (*mobile apps*), dan perangkat penginderaan (*sensors*) yang dapat menangkap data secara otomatis dan real-time (Ahmed, 2025). Manfaat utama dalam pengumpulan data mencakup peningkatan akurasi data

dengan mengurangi kesalahan penulisan manual, peningkatan kecepatan pengumpulan data, serta kemampuan untuk mengumpulkan volume data yang jauh lebih besar dan lebih beragam dibandingkan metode tradisional.

Manfaat kedua adalah peningkatan signifikan dalam kemampuan pemrosesan dan analisis data. Software analisis modern dapat memproses jutaan baris data dalam hitungan detik, melakukan analisis statistik kompleks yang sebelumnya tidak mungkin dilakukan, dan menghasilkan visualisasi data yang canggih dan intuitif untuk memfasilitasi interpretasi hasil. Manfaat ketiga adalah peningkatan aksesibilitas informasi melalui dashboard interaktif (*interactive dashboards*), portal web, dan aplikasi mobile yang memungkinkan pengambil keputusan kesehatan untuk mengakses informasi penting kapan saja dan di mana saja (Kuriakose et al., 2024). Manfaat keempat adalah peningkatan kolaborasi dan berbagi pengetahuan melalui platform digital yang memfasilitasi komunikasi antar tim peneliti, berbagi hasil penelitian, dan pembelajaran bersama dalam komunitas penelitian global.

10.1.3 Tantangan Implementasi Teknologi

Meskipun teknologi digital menawarkan manfaat yang sangat besar, implementasinya dalam praktik biostatistik kesehatan menghadapi berbagai tantangan yang signifikan. Tantangan pertama adalah biaya investasi yang tinggi untuk membeli hardware, software, infrastruktur jaringan, dan pelatihan personel yang diperlukan untuk menerapkan sistem teknologi digital yang

komprehensif (Arifian, 2025). Biaya awal ini seringkali menjadi hambatan utama terutama bagi institusi kesehatan di negara-negara berkembang dengan anggaran terbatas.

Tantangan kedua adalah kurangnya keahlian teknis dan kompetensi digital di antara tenaga kesehatan dan peneliti yang masih terbiasa dengan metode analisis tradisional. Banyak profesional kesehatan memerlukan pelatihan komprehensif untuk memahami dan menggunakan software statistik modern dan alat analisis data yang canggih. Tantangan ketiga adalah masalah interoperabilitas antar sistem, di mana sistem informasi kesehatan dari berbagai vendor seringkali tidak dapat berkomunikasi dengan efisien satu sama lain, sehingga menyulitkan integrasi data dari berbagai sumber (Kotlovskiy et al., 2024). Tantangan keempat adalah kekhawatiran terhadap keamanan data dan privasi, terutama dengan meningkatnya insiden pelanggaran data dan serangan siber terhadap sistem kesehatan digital.

10.1.4 Kompetensi yang Diperlukan

Untuk memanfaatkan teknologi digital secara optimal dalam analisis biostatistik kesehatan, profesional kesehatan dan peneliti memerlukan seperangkat kompetensi yang komprehensif dan terus berkembang. Kompetensi pertama adalah pemahaman dasar tentang konsep-konsep teknis termasuk pengetahuan tentang struktur data, database, infrastruktur cloud, dan jaringan komputer. Kompetensi kedua adalah keterampilan dalam menggunakan software statistik modern seperti R, Python, SPSS, SAS, atau Stata, termasuk

kemampuan untuk menulis script dan melakukan automasi analisis data (Arifian, 2025). Kompetensi ketiga adalah kemampuan dalam menganalisis dan menginterpretasikan hasil dari analisis data kompleks, memahami asumsi-asumsi yang mendasari algoritma analisis, dan kemampuan untuk mengkomunikasikan temuan kepada berbagai audiens dengan tingkat keahlian yang berbeda-beda.

Kompetensi keempat adalah pemahaman tentang etika penelitian data, privasi data, keamanan informasi, dan kepatuhan terhadap regulasi seperti GDPR (*General Data Protection Regulation*) atau regulasi privasi data lokal. Kompetensi kelima adalah keterampilan dalam visualisasi data dan komunikasi visual yang efektif untuk membuat hasil analisis lebih mudah dipahami oleh pengambil keputusan. Terakhir, kompetensi keenam adalah kemampuan untuk terus belajar dan beradaptasi dengan perkembangan teknologi yang sangat cepat, karena teknologi digital terus berkembang dan inovasi baru selalu muncul secara konsisten (Kuriakose et al., 2024).

10.2 Software dan Platform Analisis Data Biostatistik

10.2.1 Software Statistik Tradisional dan Modern

Landscape software statistik untuk analisis biostatistik kesehatan telah berkembang pesat dengan munculnya berbagai pilihan yang memiliki kekuatan dan kelemahan yang berbeda-beda. Software tradisional seperti SPSS (*Statistical Package for the Social Sciences*) tetap populer karena antarmukanya yang user-friendly dan

ketersediaan berbagai fitur analisis statistik yang komprehensif. Namun, software ini memiliki keterbatasan dalam menangani dataset yang sangat besar (*big data*) dan dalam fleksibilitas untuk customisasi analisis yang kompleks (Kotlovskiy et al., 2024).

Software berbasis bahasa pemrograman seperti R dan Python telah menjadi pilihan semakin populer di kalangan peneliti dan data scientist kesehatan karena fleksibilitas, kekuatan komputasi, dan komunitas pengguna yang sangat aktif dan saling mendukung. R menawarkan ribuan paket (*packages*) yang dapat diunduh secara gratis dan digunakan untuk berbagai jenis analisis statistik, visualisasi data, dan *machine learning*. Python juga menawarkan kemampuan serupa dengan paket-paket seperti Pandas, NumPy, Scikit-learn, dan Matplotlib yang sangat powerful untuk analisis data kesehatan dalam skala besar (Arifian, 2025).

Software proprietary lainnya seperti SAS dan Stata juga tetap relevan khususnya dalam institusi besar dan organisasi yang memiliki anggaran substantial untuk lisensi software. Pemilihan software yang tepat bergantung pada kebutuhan spesifik analisis, keahlian tim, ketersediaan anggaran, dan infrastruktur teknologi yang ada di institusi.

10.2.2 Big Data Analytics dan Machine Learning

Perkembangan *big data* dalam kesehatan telah menciptakan peluang baru sekaligus tantangan dalam analisis biostatistik kesehatan. *Big data* dalam kesehatan mencakup data dari berbagai sumber seperti rekam medis elektronik (*electronic health*

records/EHR), data laboratorium, data imaging diagnostik, data dari perangkat wearable, data dari media sosial, dan data genetik yang semuanya menghasilkan volume data yang sangat besar dengan variasi dan kecepatan akses yang tinggi (Kuriakose et al., 2024). Analisis *big data* memerlukan algoritma dan teknik yang berbeda dari analisis statistik tradisional karena data yang bersifat heterogen, tidak terstruktur, dan seringkali mengandung banyak data yang hilang atau data yang berkualitas rendah.

Machine learning menawarkan solusi powerful untuk menganalisis *big data* ini dan mengekstrak insight yang berharga untuk kesehatan masyarakat. *Machine learning* dapat digunakan untuk berbagai aplikasi kesehatan termasuk prediksi risiko penyakit, diagnosis penyakit, personalisasi pengobatan (*personalized medicine*), peramalan epidemi, dan identifikasi faktor risiko yang sebelumnya tidak diketahui (Arifian, 2025). Teknik-teknik *machine learning* seperti *random forest*, *support vector machine*, *neural networks*, dan *gradient boosting* memiliki kemampuan untuk mengidentifikasi pola kompleks dalam data yang seringkali tidak terlihat oleh analisis statistik tradisional.

10.2.3 Cloud Computing dalam Analisis Biostatistik

Komputasi cloud (*cloud computing*) telah merevolusi cara analisis biostatistik dilakukan dengan menyediakan infrastruktur komputasi yang scalable, fleksibel, dan cost-effective. Platform cloud seperti AWS (*Amazon Web Services*), Microsoft Azure, dan Google Cloud Platform menyediakan akses ke kekuatan komputasi

yang sangat besar tanpa memerlukan investasi infrastruktur fisik yang mahal di institusi lokal. Keuntungan cloud computing untuk analisis biostatistik kesehatan mencakup skalabilitas yang tidak terbatas untuk menangani dataset yang terus berkembang, keandalan dan redundansi yang tinggi untuk memastikan kontinuitas layanan, dan keamanan data yang lebih baik dengan enkripsi dan backup otomatis.

Platform cloud juga memfasilitasi kolaborasi antar peneliti dari berbagai institusi dan lokasi geografis yang berbeda untuk bekerja pada proyek penelitian bersama secara real-time. Selain itu, cloud computing memungkinkan akses ke tools analisis yang sophisticated tanpa perlu menginstal software di setiap komputer individual, sehingga mengurangi kompleksitas manajemen IT dan meningkatkan efisiensi operasional (Ahmed, 2025). Namun, adopsi cloud computing juga menghadapi tantangan termasuk kekhawatiran tentang keamanan data sensitif, ketergantungan pada koneksi internet yang stabil, dan biaya yang dapat meningkat jika tidak dikelola dengan bijaksana.

10.2.4 Integrasi Platform Data Kesehatan

Integrasi berbagai sumber data kesehatan menjadi platform terpadu merupakan kebutuhan kritis untuk mendapatkan gambaran komprehensif tentang kesehatan masyarakat dan melakukan analisis yang holistik. Platform seperti DHIS2 (*District Health Information System 2*) telah dikembangkan khusus untuk memfasilitasi integrasi data dari berbagai program kesehatan dan tingkat organisasi dalam

satu sistem terpusat (Ahmed, 2025). DHIS2 menyediakan kemampuan untuk mengumpulkan data dari berbagai sumber, validasi data secara real-time, analisis data dengan berbagai metode statistik, visualisasi data melalui dashboard dan peta, serta fasilitas untuk berbagi informasi dengan stakeholder yang berbeda.

Integrasi platform data juga memungkinkan linking data dari berbagai sumber untuk menciptakan dataset yang lebih kaya dan memfasilitasi analisis yang lebih mendalam. Misalnya, menghubungkan data rekam medis elektronik dengan data sosial ekonomi, data lingkungan, dan data perilaku dapat memberikan pemahaman yang lebih holistik tentang faktor-faktor yang memengaruhi kesehatan masyarakat. Namun, integrasi data juga menghadapi tantangan teknis dan organisasi termasuk perbedaan format data, definisi variabel yang berbeda antar sumber, masalah kualitas data, dan kekhawatiran tentang privasi ketika menghubungkan data dari berbagai sumber (Kotlovskiy et al., 2024).

10.3 Artificial Intelligence dan Advanced Analytics

10.3.1 Pendahuluan AI dalam Biostatistik Kesehatan

Kecerdasan buatan (*Artificial Intelligence/AI*) telah menjadi salah satu teknologi paling transformatif dalam kesehatan masyarakat modern dan membuka kemungkinan baru yang sangat menarik untuk analisis biostatistik kesehatan. AI mencakup berbagai teknik komputasi yang memungkinkan mesin untuk mempelajari pola dari data, membuat keputusan, dan melakukan tugas-tugas yang

sebelumnya memerlukan kecerdasan manusia (Arifian, 2025). Dalam konteks biostatistik kesehatan, AI dapat digunakan untuk automasi analisis data yang kompleks, prediksi kesehatan yang lebih akurat, personalisasi intervensi kesehatan, dan penemuan pengetahuan baru dari data kesehatan yang ada.

Aplikasi AI dalam kesehatan masyarakat telah menunjukkan hasil yang sangat menjanjikan dalam berbagai domain termasuk diagnosis penyakit, prognosis penyakit, identifikasi pasien berisiko tinggi, deteksi anomali dalam data kesehatan, dan optimalisasi alokasi sumber daya kesehatan. Namun, implementasi AI dalam praktik kesehatan memerlukan infrastruktur teknologi yang solid, dataset yang besar dan berkualitas tinggi, keahlian teknis yang mendalam, dan pemahaman tentang konteks klinis dan epidemiologis yang relevan.

10.3.2 Machine Learning untuk Prediksi Kesehatan

Machine learning adalah subset dari AI yang fokus pada pengembangan algoritma yang dapat belajar dari data dan membuat prediksi atau keputusan tanpa diprogram secara eksplisit. Dalam kesehatan masyarakat, *machine learning* memiliki aplikasi yang sangat luas untuk prediksi kesehatan menjadi salah satu yang paling penting dan impactful. Model *machine learning* dapat dilatih menggunakan dataset historis yang besar berisi informasi tentang karakteristik pasien, faktor risiko, hasil kesehatan, dan faktor lingkungan lainnya untuk mempelajari pola-pola yang

menghubungkan faktor-faktor tersebut dengan hasil kesehatan yang spesifik (Kuriakose et al., 2024).

Aplikasi praktis prediksi kesehatan menggunakan *machine learning* termasuk memprediksi onset penyakit kronis seperti diabetes atau penyakit jantung pada individu yang sehat dengan akurasi yang tinggi berdasarkan data demografis, faktor risiko, dan riwayat kesehatan mereka. Model dapat juga digunakan untuk memprediksi prognosis penyakit atau risiko mortalitas pada pasien dengan kondisi kesehatan tertentu, sehingga memungkinkan intervensi klinis yang lebih tepat waktu dan efektif. Selain itu, prediksi epidemi menggunakan *machine learning* juga menjadi sangat penting untuk kesehatan publik, memungkinkan petugas kesehatan untuk mengantisipasi ledakan penyakit menular dan mempersiapkan responss yang sesuai (Arifian, 2025).

10.3.3 Digital Twin dalam Personalized Medicine

Digital twin adalah konsep teknologi yang relatively baru namun sangat powerful dalam konteks kesehatan masyarakat modern, di mana digital twin mewakili replika virtual yang akurat dari individu, populasi, atau sistem kesehatan fisik (Kuriakose et al., 2024). Digital twin kesehatan dapat dibangun dengan mengintegrasikan berbagai jenis data kesehatan real-time dari berbagai sumber termasuk rekam medis elektronik, data dari perangkat monitoring kesehatan (*wearable devices*), data genetik, data lingkungan, dan data perilaku untuk menciptakan model virtual yang komprehensif dari individu atau populasi tersebut.

Manfaat utama digital twin dalam kesehatan adalah kemampuannya untuk melakukan simulasi dan prediksi tentang bagaimana intervensi kesehatan yang berbeda akan memengaruhi kesehatan individu, memungkinkan personalisasi perawatan kesehatan pada tingkat yang sangat detail. Digital twin juga dapat digunakan untuk mengidentifikasi kombinasi faktor risiko yang unik pada setiap individu dan merekomendasikan strategi pencegahan atau intervensi yang paling efektif untuk mereka. Dalam penelitian klinis, digital twin dapat digunakan untuk melakukan simulasi trial virtual sebelum melakukan trial klinis yang sesungguhnya, sehingga meningkatkan efisiensi dan mengurangi risiko trial.

10.3.4 Etika dan Privasi Data

Sementara AI dan teknologi digital menawarkan manfaat yang sangat besar untuk kesehatan, implementasinya juga menghadirkan tantangan etika dan privasi data yang signifikan yang harus ditangani dengan serius. Isu privasi data adalah salah satu yang paling kritis karena analisis biostatistik kesehatan seringkali melibatkan data sensitif dan pribadi dari individu-individu, dan ada risiko bahwa data ini dapat disalahgunakan atau diungkapkan tanpa persetujuan (Ahmed, 2025). Regulasi seperti GDPR (*General Data Protection Regulation*) di Eropa dan regulasi privasi data lokal di berbagai negara telah ditetapkan untuk melindungi privasi individu dan memastikan bahwa institusi kesehatan menangani data dengan tanggung jawab.

Isu etika lainnya termasuk bias dalam algoritma AI yang dapat menghasilkan prediksi atau rekomendasi yang tidak adil atau diskriminatif terhadap kelompok-kelompok tertentu. Jika algoritma AI dilatih menggunakan data historis yang mengandung bias sosial atau kesenjangan kesehatan yang sudah ada, algoritma tersebut mungkin mereproduksi atau bahkan memperkuat bias-bias tersebut dalam keputusan atau prediksi yang dibuat (Arifian, 2025). Oleh karena itu, penting untuk melakukan audit bias (*bias audit*) pada algoritma AI dan memastikan bahwa algoritma tersebut adil dan tidak diskriminatif terhadap semua kelompok populasi.

10.4 Implementasi Praktis dan Studi Kasus

10.4.1 Sistem Informasi Kesehatan Terintegrasi (DHIS2)

DHIS2 adalah platform open-source yang telah dikembangkan dan diimplementasikan di lebih dari 100 negara untuk mengumpulkan, menganalisis, dan menyebarkan data kesehatan pada skala yang besar. DHIS2 menyediakan sistem manajemen data kesehatan yang komprehensif yang dapat menghubungkan berbagai program kesehatan dan tingkat organisasi dari fasilitas kesehatan lokal hingga tingkat nasional dalam satu platform terpadu (Ahmed, 2025). Platform ini memungkinkan pengumpulan data secara real-time dari berbagai sumber, validasi data otomatis untuk mendeteksi kesalahan atau anomali, analisis data menggunakan berbagai metode statistik, dan visualisasi data melalui dashboard interaktif dan peta geografis.

Implementasi DHIS2 telah menunjukkan hasil yang sangat signifikan dalam meningkatkan kualitas data kesehatan, meningkatkan aksesibilitas informasi kesehatan bagi pengambil keputusan, dan mendukung pengambilan keputusan berbasis data yang lebih baik. Namun, implementasi DHIS2 juga menghadapi tantangan praktis termasuk kebutuhan untuk training yang komprehensif bagi pengguna di berbagai level, integrasi dengan sistem legacy yang mungkin sudah ada, dan kebutuhan untuk customisasi sistem sesuai dengan kebutuhan lokal spesifik (Ahmed, 2025).

10.4.2 Mobile Health dan Digital Monitoring

Mobile health (mHealth) mengacu pada penggunaan perangkat mobile seperti smartphone dan tablet untuk memberikan layanan kesehatan, mengumpulkan data kesehatan, dan monitoring kesehatan populasi. Aplikasi mHealth telah berkembang pesat dengan penawaran yang beragam mulai dari aplikasi untuk edukasi kesehatan, aplikasi untuk konsultasi kesehatan jarak jauh, aplikasi untuk monitoring penyakit kronis, hingga aplikasi untuk surveilans penyakit menular (Arifian, 2025). Perangkat wearable seperti smartwatch, fitness tracker, dan sensor biomedis lainnya dapat mengumpulkan data kesehatan secara kontinyu dan real-time, memberikan informasi yang kaya tentang aktivitas fisik, pola tidur, detak jantung, dan berbagai parameter kesehatan lainnya.

Data yang dikumpulkan melalui aplikasi mHealth dan perangkat wearable dapat diintegrasikan dengan sistem informasi

kesehatan untuk analisis yang lebih mendalam dan personalisasi intervensi kesehatan. Namun, mHealth juga menghadapi tantangan termasuk variasi kualitas aplikasi, masalah keamanan data dan privasi, ketergantungan pada koneksi internet yang mungkin tidak selalu tersedia di daerah terpencil, dan kebutuhan untuk education dan pelatihan bagi pengguna agar dapat menggunakan aplikasi secara efektif (Kuriakose et al., 2024).

10.4.3 Visualisasi Data dan Dashboard

Visualisasi data merupakan aspek krusial dari analisis biostatistik kesehatan karena memungkinkan komunikasi hasil analisis yang kompleks dalam format yang mudah dipahami oleh audiens dengan tingkat keahlian yang berbeda-beda. Dashboard interaktif (*interactive dashboards*) telah menjadi alat penting untuk presentasi data kesehatan karena memungkinkan pengguna untuk mengeksplorasi data secara mandiri, melakukan filtering berdasarkan berbagai dimensi seperti wilayah geografis atau jenis penyakit, dan melihat tren dari waktu ke waktu (Ahmed, 2025).

Teknologi visualisasi data modern menggunakan berbagai teknik grafis yang sophisticated termasuk peta tematik (*choropleth maps*) untuk menampilkan variasi spasial dalam indikator kesehatan, grafik garis dan batang untuk menunjukkan tren temporal, scatterplot untuk mengeksplorasi hubungan antara variabel, dan heat map untuk menampilkan intensitas fenomena kesehatan di berbagai dimensi. Tools seperti Tableau, Power BI, dan Google Data Studio telah membuat pembuatan dashboard yang sophisticated menjadi

lebih mudah diakses bahkan oleh pengguna yang tidak memiliki keahlian pemrograman tingkat lanjut (Arifian, 2025).

10.4.4 Evaluasi Kinerja dan ROI

Evaluasi kinerja dari implementasi teknologi digital dalam analisis biostatistik kesehatan sangat penting untuk memastikan bahwa investasi teknologi memberikan nilai yang nyata dan terukur. Metrik kinerja yang dapat diukur termasuk peningkatan kualitas data dalam hal akurasi dan kelengkapan, peningkatan kecepatan analisis data dan waktu respons dalam pengambilan keputusan, peningkatan aksesibilitas informasi kesehatan bagi para pengambil keputusan, dan peningkatan dalam kualitas keputusan kesehatan yang dibuat (Kotlovskiy et al., 2024). ROI (*Return on Investment*) dari investasi teknologi dapat dievaluasi dengan membandingkan biaya total investasi termasuk hardware, software, infrastruktur, dan training dengan manfaat yang diperoleh seperti pengurangan kesalahan, peningkatan produktivitas, peningkatan revenue dari layanan kesehatan yang lebih baik, dan penghematan biaya dari pencegahan penyakit yang lebih efektif.

Studi evaluasi pada berbagai institusi kesehatan telah menunjukkan bahwa implementasi teknologi digital dalam analisis biostatistik kesehatan umumnya menghasilkan ROI yang positif dalam jangka menengah hingga panjang, meskipun periode payback mungkin berbeda-beda tergantung pada skala implementasi, kondisi baseline sebelum implementasi, dan efektivitas manajemen implementasi (Ahmed, 2025).

Latihan Soal

Berikut ini disajikan lima soal esai singkat tanpa panduan jawaban untuk menguji pemahaman Anda tentang pemanfaatan teknologi digital dalam analisis biostatistik kesehatan.

Soal 1. Jelaskan secara detail bagaimana evolusi teknologi digital telah mengubah cara analisis biostatistik kesehatan dilakukan dari era manual hingga era *big data* dan *artificial intelligence*. Diskusikan tantangan dan peluang yang muncul di setiap tahap evolusi tersebut.

Soal 2. Bandingkan dan kontraskan berbagai software statistik yang tersedia untuk analisis biostatistik kesehatan, mulai dari SPSS tradisional hingga R dan Python modern. Jelaskan situasi atau konteks di mana masing-masing software paling sesuai digunakan.

Soal 3. Jelaskan konsep *digital twin* dalam konteks personalized medicine dan kesehatan masyarakat. Bagaimana *digital twin* dapat digunakan untuk meningkatkan kualitas keputusan klinis dan pencegahan penyakit pada populasi?

Soal 4. Diskusikan tantangan etika dan privasi data yang muncul dengan penggunaan teknologi AI dan *machine learning* dalam analisis biostatistik kesehatan. Bagaimana cara kita dapat memastikan bahwa algoritma AI adil, tidak bias, dan menghormati privasi individu?

Soal 5. Anda diminta untuk mengimplementasikan sistem DHIS2 di suatu distrik kesehatan yang belum memiliki sistem informasi kesehatan digital yang terintegrasi. Jelaskan langkah-langkah implementasi yang diperlukan, tantangan yang mungkin dihadapi, dan strategi untuk memastikan keberhasilan implementasi jangka panjang.

DAFTAR PUSTAKA

- Ahmed, D. A. Y. (2025). Health Systems Governance in Somalia: An Examination of Validity, Digital Accountability, and Community Health Workforce through Mixed-Methods Research. *Advances in Social Sciences and Management*. <https://doi.org/10.63002/assm.304.1056>
- Arifian, R. (2025). Transformasi Artificial Intelligence dalam Sistem Pelayanan Kesehatan Primer-Sekunder Daerah. *Nusantara Innovation Journal*. <https://doi.org/10.70260/nij.v3i2.66>
- Kotlovskiy, M., Armashevskaya, O., Sokolovskaya, T., & Gusev, A. (2024). DEVELOPMENT AND APPLICATION OF DIGITAL TECHNOLOGIES IN SCIENTIFIC RESEARCH IN THE SPECIALTY «HEALTH CARE ORGANIZATION AND PUBLIC HEALTH». *Vrach i Informacionnye Tehnologii*. https://doi.org/10.25881/18110193_2024_2_40
- Kuriakose, S. M., Joseph, J., A, R., & Kollinal, R. K. (2024). The Rise of Digital Twins in Healthcare: A Mapping of the Research Landscape. *Cureus*. <https://doi.org/10.7759/cureus.65358>

PROFIL PENULIS



Nama : Bangkit Ary Pratama, SKM., M.Kes

Tempat/Tanggal Lahir: Sukoharjo, 20 Mei 1989

Alamat Domisili : Sukoharjo, Jawa Tengah

Hobi : Travelling

Pesan untuk pembaca:

Rekam medis bukan sekadar dokumen administratif, melainkan fondasi pelayanan kesehatan yang bermutu. Pemahaman atas makna di balik setiap catatan akan mendukung terciptanya layanan yang lebih profesional, manusiawi, dan berkesinambungan.



Nama : Hardian Bimanto, S.KM., M.Kes
Tempat/tanggal lahir : Surabaya, 25 oktober 1993
alamat : jagir sidoresmo 10/27
Riwayat Pendidikan : S2
Pekerjaan : Dosen
Pesan untuk pembaca : semangat dan tidak pantang menyerah



Nama : Desie Rahmawati, S.KM., M.Kes

Tempat/Tanggal Lahir : Lamongan, 18 Desember 1994

Alamat Domisili : Samarinda, Kalimantan Timur

Hobi : -

Pesan untuk pembaca : Melalui pemahaman epidemiologi, penulis berharap kita semua semakin tanggap terhadap ancaman penyakit dan semakin siap berperan aktif dalam menciptakan masyarakat yang lebih sehat.



Nama : Shiva Aflahiyah, S.Tr.Keb., M.K.M
Tempat/Tanggal Lahir : Gowa/ 01 April 1996
Alamat Domisili : Jalan Belimbing Gang G No. 7 Kota
Denpasar-Bali
Hobi : Penulis memiliki hobi membaca, menulis,
dan melakukan penelitian di bidang kesehatan.
Pesan untuk pembaca : Kesehatan merupakan aset terpenting
dalam kehidupan. Pencegahan dan penanganan masalah kesehatan
umum tidak hanya menjadi tanggung jawab tenaga kesehatan, tetapi
juga membutuhkan kesadaran dan partisipasi aktif seluruh
masyarakat. Melalui pemahaman yang tepat mengenai faktor risiko,
gejala, serta langkah penanganan dini, setiap individu dapat berperan
dalam menjaga kesehatan diri, keluarga, dan lingkungannya.
Kami berharap tulisan ini dapat memberikan wawasan serta menjadi
sumber informasi yang bermanfaat bagi pembaca dalam menerapkan

perilaku hidup bersih dan sehat, serta mendorong peningkatan derajat kesehatan di Masyarakat.



Nama : Fitri Eka Sari S, M.Kes

Tempat/Tanggal Lahir : Jakarta, 19 february 1975

Alamat Domisili : Bandarlampung, Lampung

Hobi : Menulis

Pesan untuk pembaca : Semoga buku ini dapat menambah wawasan keilmuan pembaca dan sebagai keterbaruan literature yang di gunakan.



Nama : Zakiyah, SST., M.K.M

Tempat/Tanggal Lahir : Susoh, 12 Juli 1991

Hobi : Membaca, menulis, dan berbagi ilmu di bidang kesehatan

Pesan untuk pembaca : Penulis berharap buku ini dapat menjadi panduan dalam mengolah, menganalisis, dan menginterpretasikan data kesehatan secara tepat sehingga dapat mendukung pengambilan keputusan yang berbasis bukti



Nama : Tri Budiarti, S.K.M, M.K.M
Tempat/tanggal lahir : Sidomulyo, 7 Juli 1996
Alamat : Jl. Bakti Dusun VII Desa Sidomulyo Kec.
Binjai Kab. Langkat Sumut
Riwayat Pendidikan : Magister Kesehatan Masyarakat, InKes
Delihusada Delitua
Pekerjaan : Dosen
Pesan untuk pembaca : Melalui buku ini, penulis menulis dengan penuh arti agar para pembaca dapat memahami dengan mudah dan bisa memberikan tambahan pengetahuan kepada para pembacanya khususnya yang berkaitan dengan bidang ini yang menjadi fokus pembahasan. Maka dari itu, kritik dan saran sangat diperlukan bagi penulis agar setiap tulisan nya dapat diperbaiki dan menyempurkannya dimasa yang akan datang.



Nama : Agus Erwin Ashari, SKM, M.Kes

Tempat/Tanggal Lahir : Ujung Pandang / 17 Agustus 1971

Alamat : BTN Axuri Blok P No. 54 Mamuju
Sulawesi Barat

Riwayat Pendidikan :

- SD Inpres Antang, Ujung Pandang 1984
- SMP Negeri 8 Ujung Pandang 1987
- SMA Negeri 5 Ujung Pandang 1990
- FKM UNHAS Ujung Pandang 1995
- Pasca Sarjana IKM UGM Jogjakarta 2003

Pekerjaan : ASN Dosen

Pesan untuk pembaca : Suatu proses tidak akan mengkhianati hasil,
proses yang baik akan menghasilkan sesuatu yang bermutu



Nama : Intan Pujilestari, SST., M.K.M

Tempat/tanggal lahir : Ciamis, 17 Januari 1990

Alamat : Gg. Slamet 1 RT/RW 006/003

Kiaracondong Kota Bandung

Riwayat Pendidikan :

- D4 Stikes Jenderal Achmad Yani
- S2 Kesehatan Masyarakat Universitas Muhammadiyah Jakarta

Pekerjaan : Dosen

Pesan untuk pembaca : Semoga apa yang Anda baca dapat membuka wawasan baru, menghadirkan sudut pandang berbeda, atau bahkan menyalakan semangat untuk terus belajar dan berkembang.



Nama : Izal Puji Santoso, SKM., M.Kes
Tempat/tanggal lahir : Blitar, 27 September 1995
Alamat : Jl. Pondok Bambu Komplek Green Tasbih
Balitan Banjarbaru
Riwayat Pendidikan : S1 dan S2 Kesehatan Masyarakat,
Universitas Lambung Mangkurat
Pekerjaan : Dosen
Pesan untuk pembaca : Semoga tulisan ini dapat memberikan
kontribusi positif serta menjadi referensi yang bermanfaat bagi
pengembangan ilmu pengetahuan dan praktik di bidang kesehatan
masyarakat.

Buku **Dasar-Dasar Biostatistik untuk Kesehatan Masyarakat** memberikan pemahaman mengenai peran biostatistik dalam mengolah, menganalisis, dan menginterpretasikan data kesehatan untuk mendukung pengambilan keputusan yang tepat. Di era informasi saat ini, kemampuan memahami data menjadi semakin penting dalam upaya meningkatkan kualitas kesehatan masyarakat.

Disajikan dengan bahasa yang mudah dipahami, buku ini membahas berbagai konsep dasar biostatistik, mulai dari pengumpulan data, penyajian data, ukuran pemusatan dan penyebaran, hingga teknik analisis yang digunakan dalam bidang kesehatan. Pembahasan juga dilengkapi dengan contoh-contoh yang dekat dengan permasalahan kesehatan sehari-hari sehingga memudahkan pembaca dalam memahami penerapannya.

Buku ini sangat cocok untuk masyarakat umum yang ingin memahami cara membaca dan memanfaatkan data kesehatan secara lebih efektif. Dengan penyajian yang sistematis dan praktis, buku ini menjadi bacaan yang bermanfaat untuk meningkatkan literasi data dan pemahaman terhadap berbagai informasi kesehatan yang berkembang di masyarakat.