

BUKU AJAR
BIOSTATISTIK DASAR

deepublish / publisher

UU No 28 tahun 2014 tentang Hak Cipta

Fungsi dan sifat hak cipta Pasal 4

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

Pembatasan Pelindungan Pasal 26

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

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

Sanksi Pelanggaran Pasal 113

1. Setiap Orang yang dengan tanpa hak melakukan pelanggaran hak ekonomi sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf i untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 1 (satu) tahun dan/atau pidana denda paling banyak Rp100.000.000 (seratus juta rupiah).
2. Setiap Orang yang dengan tanpa hak dan/atau tanpa izin Pencipta atau pemegang Hak Cipta melakukan pelanggaran hak ekonomi Pencipta sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf c, huruf d, huruf f, dan/atau huruf h untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 3 (tiga) tahun dan/atau pidana denda paling banyak Rp500.000.000,00 (lima ratus juta rupiah).

BUKU AJAR BIOSTATISTIK DASAR

IKE ANGGRAENI G
MASITHAH
RAHMI SUSANTI



deepublish

Cerdas, Bahagia, Mulia, Lintas Generasi.

BUKU AJAR BIOSTATISTIK DASAR

Ike Anggraeni G, dkk

Desain Cover :

Sumber :

Tata Letak :
Ajuk

Proofreader :
A. Timor Eldian

Ukuran :
x, 164 hlm, Uk: 15.5x23 cm

ISBN :
No ISBN

Cetakan Pertama :
Bulan 2022

Hak Cipta 2022, Pada Penulis

Isi diluar tanggung jawab percetakan

Copyright © 2022 by Deepublish Publisher
All Right Reserved

Hak cipta dilindungi undang-undang.
Dilarang keras menerjemahkan, memfotokopi, atau
memperbanyak sebagian atau seluruh isi buku ini
tanpa izin tertulis dari Penerbit.

PENERBIT DEEPUBLISH
(Grup Penerbitan CV BUDI UTAMA)
Anggota IKAPI (076/DIY/2012)
Jl.Rajawali, G. Elang 6, No 3, Drono, Sardonoharjo, Ngaglik, Sleman
Jl.Kaliurang Km.9,3 – Yogyakarta 55581
Telp/Faks: (0274) 4533427
Website: www.deepublish.co.id
www.penerbitdeepublish.com
E-mail: cs@deepublish.co.id

KATA PENGANTAR PENERBIT

Assalamualaikum, w.r. w.b.

Segala puji kami haturkan ke hadirat Allah Swt., Tuhan Yang Maha Esa, atas limpahan rahmat dan karunia-Nya. Tak lupa, lantunan selawat serta salam senantiasa tercurahkan kepada Nabi Besar Muhammad saw.

Dalam rangka mencerdaskan dan memuliakan umat manusia dengan penyediaan serta pemanfaatan ilmu pengetahuan dan teknologi untuk menciptakan industri *processing* berbasis sumber daya alam (SDA) Indonesia, Penerbit Deepublish dengan bangga menerbitkan buku dengan judul ***Buku Ajar Biostatistik Dasar***.

Terima kasih dan penghargaan terbesar kami sampaikan kepada penulis yang telah memberikan kepercayaan, perhatian, dan kontribusi penuh demi kesempurnaan buku ini. Semoga buku ini bermanfaat bagi semua pembaca, mampu berkontribusi dalam mencerdaskan dan memuliakan umat manusia, serta mengoptimalkan pemanfaatan ilmu pengetahuan dan teknologi di tanah air.

Wassalamualaikum, w.r. w.b.

Hormat Kami,

Penerbit Deepublish

KATA PENGANTAR

deeppublish / publisher

DAFTAR ISI

KATA PENGANTAR PENERBIT	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
PROLOG	x
BAB 1 KONSEP DASAR BIOSTATISTIK.....	1
A. Tujuan Pembelajaran	1
B. Konsep Umum Biostatistik dan Peranannya dalam Kesehatan Masyarakat.....	1
C. Data dan Variabel.....	2
D. Statistik Deskriptif dan Inferensial	9
Rangkuman.....	11
E. Soal Latihan.....	11
BAB 2 PENYAJIAN DATA.....	13
A. Tujuan Pembelajaran	13
B. Tekstual.....	13
C. Distribusi Frekuensi	14
D. Grafik.....	21
E. Rangkuman.....	33
F. Soal Latihan.....	33
BAB 3 UKURAN SENTRAL DAN TENDENSI.....	34
A. Tujuan Pembelajaran	34
B. Ukuran Tendensi Sentral	34
C. Ukuran Variabilitas.....	40
D. Normalitas Data.....	42
F. Rangkuman.....	44
G. Soal Latihan.....	44
BAB 4 TABULASI SILANG	45
A. Tujuan Pembelajaran	45
B. Pengertian Tabulasi	45
C. Persentase Baris.....	45
D. Persentase Kolom	46

E. Persentase Total	47
F. Rangkuman.....	47
G. Soal Latihan.....	48
BAB 5 TEORI PROBABILITAS.....	49
A. Tujuan Pembelajaran	49
B. Konsep Probabilitas.....	49
C. Permutasi dan Kombinasi	49
D. Distribusi Binomial dan Poisson	51
F. Rangkuman.....	55
G. Soal Latihan.....	55
BAB 6 CENTRAL LIMIT THEOREM	57
A. Tujuan Pembelajaran	57
B. Central Limit Theorem	57
C. Distribusi Sampling.....	57
D. Distribusi Probabilitas Normal.....	61
E. Distribusi Normal Baku (Standar).....	61
F. Rangkuman.....	62
G. Soal Latihan.....	62
BAB 7 TEORI ESTIMASI.....	63
A. Tujuan Pembelajaran	63
B. Estimasi.....	63
C. Estimasi Interval Kepercayaan dari Nilai Rerata Populasi.....	63
D. Estimasi Interval Kepercayaan dari Nilai Proporsi Populasi.....	64
E. Rangkuman.....	65
F. Soal Latihan.....	65
BAB 8 UJI HIPOTESIS.....	66
A. Tujuan Pembelajaran	66
B. Hipotesis.....	66
C. Area Penolakan	68
D. Kesalahan Pengambilan Keputusan.....	69
E. Langkah Uji.....	70
F. Penentuan Uji Statistik	71

G. Menguji Beda Mean 1 Sampel.....	71
H. Soal Latihan.....	79
BAB 9 UJI BEDA DUA RERATA.....	80
A. Tujuan Pembelajaran	80
B. Uji Beda Dua Rerata Independen.....	80
C. Uji Beda Dua Rerata Dependen.....	83
D. Rangkuman.....	90
E. Soal Latihan.....	90
BAB 10 UJI BEDA LEBIH DARI DUA RERATA.....	92
A. Tujuan Pembelajaran	92
B. Analisis Varians.....	92
C. Prosedur Analisis Varians	93
D. Uji Hipotesis Analisis Varians	95
E. Rangkuman.....	95
F. Soal Latihan.....	96
BAGIAN 11 UJI KORELASI DAN REGRESI.....	97
BAGIAN 12 UJI CHI-SQUARE	109
BAGIAN 13 UJI HIPOTESIS BEDA RERATA INDEPENDENT (NON-PARAMETRIK)	119
BAGIAN 14 BESAR SAMPEL UNTUK ESTIMASI.....	144

PROLOG

Buku ajar ***Biostatistik Dasar*** ini membahas tentang variabel (definisi variabel yang bersifat data kontinu dan diskret, variabel numerik vs kategorik), skala data, berbagai jenis sumber data (rutin vs *ad hoc*/survei vs sensus), penyajian data, ringkasan data numerik. Dibahas pula tentang probabilitas, central limit theorem dan estimasi. Teknik statistik deskriptif dan statistik inferensi, serta berbagai prinsip uji statistik dan menginterpretasi hasil statistik parametrik serta metode dan besar sampel.

BAB 1

KONSEP DASAR BIOSTATISTIK

A. Tujuan Pembelajaran

Setelah mempelajari bab ini mahasiswa diharapkan mampu menunjukkan peran biostatistika dalam kesehatan masyarakat. Selanjutnya mahasiswa diharapkan mampu membedakan: konsep data dan variabel, mampu menentukan skala data, membedakan sumber data, dan jenis statistika.

B. Konsep Umum Biostatistik dan Peranannya dalam Kesehatan Masyarakat

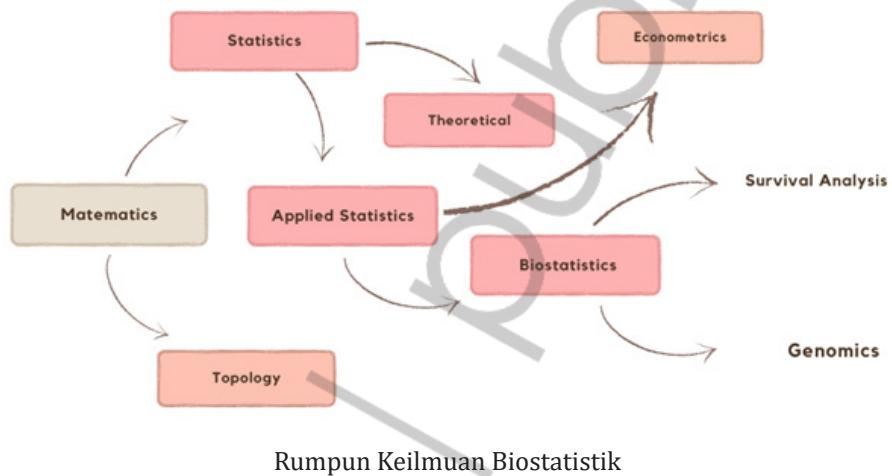
Apakah biostatistik, dan apa yang bisa dilakukan biostatistik?

Biostatistik berasal dari asal kata statistik secara umum sangat bervariasi, paling sering merujuk pada angka, atau data. "Statistics" berasal dari huruf Latin *statista*, yaitu orang yang berhubungan dengan masalah kenegaraan. Awalnya disebut *aritmatika negara*, yaitu tabulasi informasi tentang warga negara untuk data sensus, perpajakan dan perencanaan perang.

Seiring perkembangan ilmu pengetahuan statistik diartikan sebagai konsep "aktif". Melakukan statistik adalah cara berpikir tentang angka (mulai dari pengumpulan, analisis, dan penyajian) dengan penekanan pada menghubungkan interpretasi dan maknanya (Le, 2003).



Biostatistics when math and health combine, merupakan penerapan dan pengembangan metode statistik untuk masalah yang berhubungan dengan kehidupan, contoh: ekonomi, pertanian, kedokteran, biologi, dan kesehatan masyarakat. Berikut gambaran akar rumpun keilmuan biostatistik.

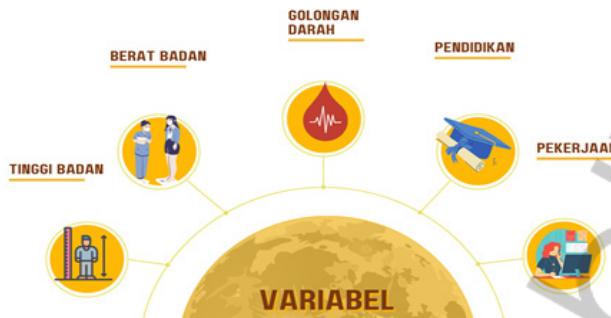


Contoh peran Biostatistik dalam bidang kesehatan:

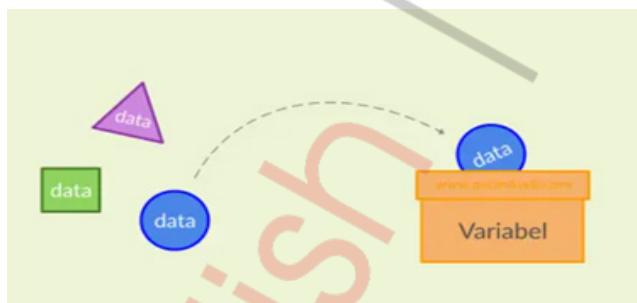
- Kepala Daerah X ingin menaksir cakupan pelayanan masyarakat lanjut usia (lansia) pada tahun 2022.
- Seorang ahli virologi ingin mengetahui probabilitas seseorang yang telah mendapatkan vaksin ulangan ke-2, akan terinfeksi penyakit "Y".
- Seorang ahli promosi kesehatan ingin membandingkan efektivitas penggunaan dua media terhadap peningkatan sikap dan perilaku sehat pada murid Sekolah Dasar.

C. Data dan Variabel

Variabel merupakan setiap karakteristik, jumlah, atau kuantitas yang dapat diukur, diklasifikasikan, dihitung, nilainya bervariasi yang merupakan objek yang menjadi titik perhatian suatu studi.



Data mengacu pada seperangkat nilai, yang biasanya diatur oleh variabel. Data didefinisikan sebagai fakta, informasi atau angka yang dihasilkan dari pengukuran maupun proses penghitungan. Misalnya, ketika seorang kader menimbang balita atau mengukur tinggi badan balita, yang merupakan fakta terdiri dari angka seperti berat badan anak 10 kilogram atau panjang badan bayi 64 centimeter.



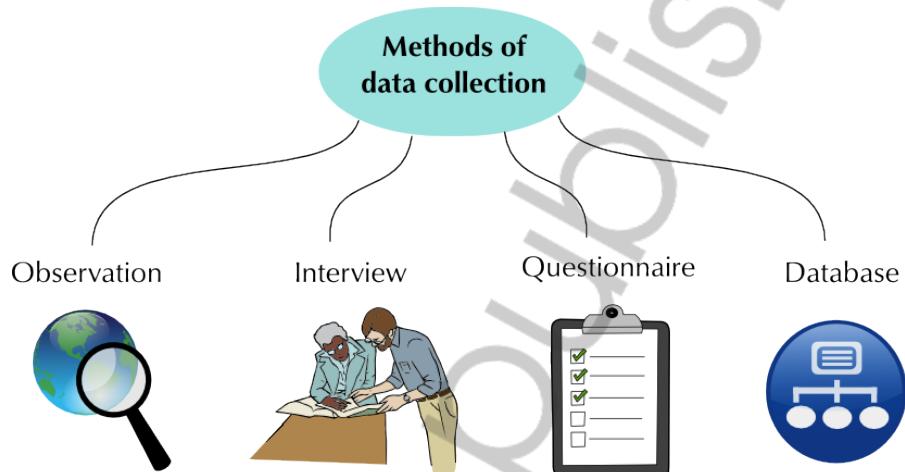
Bagaimana data dikumpulkan?

Pengumpulan data dilakukan untuk memperoleh informasi yang dibutuhkan dalam rangka mencapai tujuan.

Tujuan mengumpulkan data ialah memperoleh informasi yang berkualitas, dengan cara sistematis yang memungkinkan seseorang menjawab pertanyaan studi/ penelitian.



Data dikumpulkan dengan berbagai cara diantaranya yaitu: (1) observasi, (2) wawancara, (3) kuesioner dan (4) basis data.



1. Observasi

Pengamatan yang melibatkan sistem sensorik (mata dan telinga) untuk merekam informasi, perilaku maupun lingkungan. Observasi atau pengamatan harus dilakukan terstruktur di mana pengamatan dicatat terhadap daftar periksa/lembar observasi yang dipersiapkan.

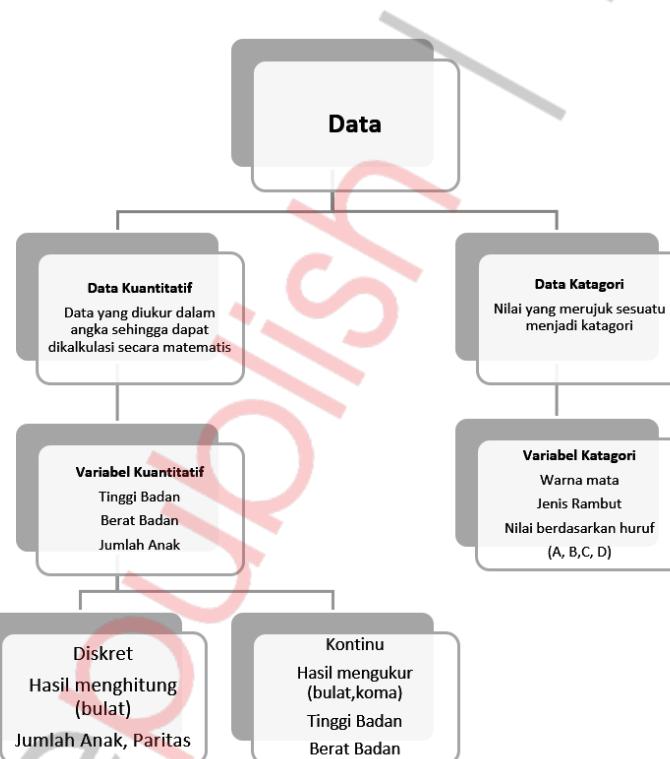
2. Interview

Interview/wawancara merupakan alat yang paling umum digunakan dalam mengumpulkan data. Dilakukan dengan satu orang pada satu waktu (wawancara individu) atau kelompok orang. Wawancara dapat dikelola secara formal atau informal. Selain itu wawancara dapat dilakukan secara tatap muka atau melalui media jarak jauh seperti telepon, Skype atau zoom.

Wawancara dapat dilaksanakan melalui pertanyaan tertulis melalui surat atau email. Wawancara juga dapat dilakukan terstruktur, semi-terstruktur atau terbuka. Wawancara terstruktur didasarkan pada serangkaian pertanyaan inti yang selalu ditanyakan dalam urutan yang sama. Wawancara semi-terstruktur juga berisi serangkaian pertanyaan inti, tetapi memungkinkan pewawancara

- untuk mengajukan pertanyaan tambahan, atau mengubah urutan pertanyaan yang diajukan.
3. *Questionnaire*/Kuesioner
Kuesioner dirancang untuk mengumpulkan dan mencatat informasi dari banyak orang, kelompok atau organisasi secara konsisten. Kuesioner adalah formulir yang berisi pertanyaan. Ini mungkin formulir cetak atau yang dirancang untuk diisi secara online.
 4. *Database*/Basis Data
Data yang dikumpulkan bersumber dari data rutin fasilitas kesehatan maupun berbasis hasil survei nasional. Aspek kunci dari basis data adalah data yang akurat karena data yang tidak akurat menyebabkan pengambilan keputusan yang kurang informatif.

Data berdasarkan jenisnya dibagi ke dalam dua bagian: kualitatif dan kuantitatif.



1. Kualitatif

Data dalam bentuk **kualitas** (baik/buruk) atau **pernyataan** (setuju/tidak). Data ini mendeskripsikan kualitas seseorang atau sesuatu yang dipelajari, atau merupakan kategori karena nilai yang diperoleh dapat dikelompokkan pada kategori tertentu.

2. Kuantitatif

Data dalam bentuk angka numerik

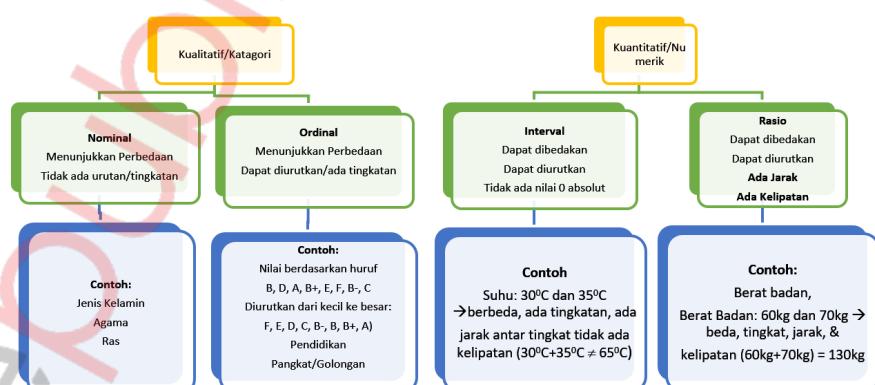
1) Kuantitatif Diskret: Hasil menghitung (bulat)

2) Kuantitatif Kontinu: Hasil mengukur (bulat,koma)

Skala pengukuran merupakan tingkat presisi dari di mana karakteristik diukur. Hasil dari suatu pengukuran akan berimplikasi terhadap bagaimana data disajikan atau disimpulkan maupun dianalisis.

Skala data dibagi dalam 4 (empat) bagian:

1. Nominal; kelompok data berdasarkan kesamaan karakteristik tertentu.
2. Ordinal; kelompok data dengan urutan, peringkat atau tingkatan.
3. Interval; data dalam bentuk numerik atau angka, dapat dibedakan, dapat diurutkan, memiliki jarak namun tidak memiliki nol (0) mutlak.
4. Rasio; data yang paling kompleks, dapat digunakan dalam operasi hitung angka dalam data rasio merupakan angka yang sesungguhnya, bukan hanya sebagai simbol sehingga ada jarak dan nada kelipatan.



Berdasarkan dari mana mendapatkannya terdapat tiga sumber data:

1. Primer

Data primer mengacu pada data tangan pertama yang dikumpulkan oleh peneliti sendiri. Contoh: hasil wawancara, hasil observasi, hasil *focus group discussion* (FGD).

2. Sekunder

Data sekunder yaitu data yang dikumpulkan oleh orang lain sebelumnya. Contoh data yang bersumber dari survei *public domain* (Riskesdas, SDKI, Susenas), Publikasi pemerintah, situs web, buku, data rutin (rekam medis dari pelayanan kesehatan).

3. Tersier

Sumber tersier adalah sumber dari publikasi yang merangkum dan mencerna informasi dalam sumber primer dan sekunder untuk memberikan latar belakang tentang suatu topik, ide, atau peristiwa. Ensiklopedia dan kamus biografi.

Sumber data kegiatan statistik juga berasal dari kebutuhan untuk menjawab pertanyaan atau mencari kebenaran. Misalnya, ahli epidemiologi menginginkan jawaban atas pertanyaan mengenai pola penyebaran suatu kasus penyakit baru yang ada di masyarakat. Promotor kesehatan menginginkan jawaban atas pertanyaan mengenai berbagai persepsi yang ada di masyarakat yang menghambat dalam pencarian pelayanan kesehatan. Pendekatan yang tepat untuk mencari jawaban atas berbagai pertanyaan akan memerlukan penggunaan statistik, harus tersedia data yang sesuai untuk dijadikan sebagai bahan mentah untuk penyelidikan. Data tersebut biasanya tersedia dari satu atau lebih sumber berikut:

1) Data rutin

Mayoritas instansi terutama penyedia pelayanan kesehatan menyimpan catatan transaksi sehari-hari dari aktivitasnya. Rekam medis rumah sakit, puskesmas serta klinik misalnya, menyimpan sejumlah besar informasi tentang pasien, sedangkan catatan akuntansi rumah sakit berisi banyak data tentang data seluruh

sumber daya.

2) Survei

Jika data yang diperlukan untuk menjawab pertanyaan tidak tersedia dari dokumen yang berasal dari data rutin, sumber logisnya bisa berupa hasil survei. Misalkan, ahli kebijakan kesehatan ingin mengetahui hubungan kinerja tenaga kesehatan dengan fasilitas, dapat menggunakan data hasil survei Riset Fasilitas Kesehatan). Survei lain yang dapat digunakan khususnya bidang kesehatan diantaranya: Riskesdas (Riset Kesehatan Dasar), SDKI (Survei Demografi dan Kesehatan Indonesia), *IFLS* (Indonesia Family Life Survey).

3) Eksperimen

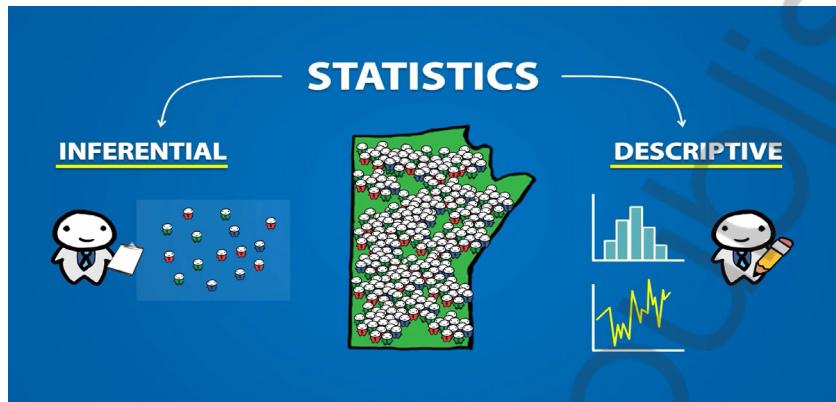
Seringkali data yang diperlukan untuk menjawab pertanyaan hanya tersedia sebagai hasil eksperimen. Seorang terapis ingin mengetahui strategi mana yang terbaik untuk perilaku berhenti merokok. Ia mungkin melakukan percobaan di mana strategi yang digunakan berbeda untuk memotivasi kepatuhan tidak merokok dicoba dengan orang yang berbeda. Evaluasi selanjutnya dari respons terhadap strategi yang berbeda memungkinkan terapis untuk memutuskan strategi/metode mana yang paling efektif.

4) Sumber eksternal

Data yang diperlukan untuk menjawab pertanyaan mungkin sudah ada dalam bentuk laporan yang dipublikasikan, bank data yang tersedia secara komersial, atau literatur penelitian. Dengan kata lain, dimungkinkan menemukan bahwa orang lain telah mengajukan pertanyaan yang sama, dan jawaban yang diperoleh mungkin dapat diterapkan pada situasi saat ini.

D. Statistik Deskriptif dan Inferensial

Dua bidang utama statistik dibagi dalam deskriptif dan inferensial.



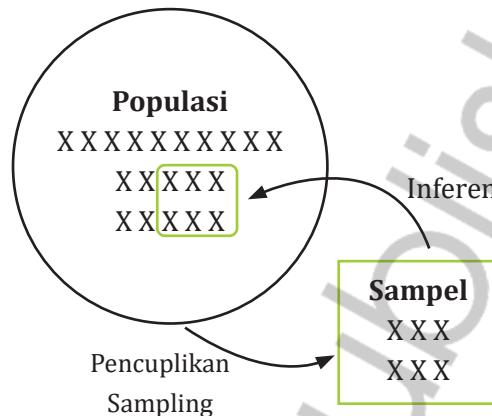
Inferensial

Deskriptif

Berhubungan dengan pengambilan sampel, analisis sampel tersebut untuk membuat justifikasi hasil sampel merupakan gambaran populasi. Meringkas dan mengomunikasikan informasi terkait sejumlah/sekelompok data dan menyajikannya dalam bentuk tabel, diagram grafik dan besaran lainnya.

Berhubungan dengan pengambilan data dan "berbicara" tentang hal tersebut. Membuat kesimpulan berdasarkan data yang dikumpulkan, membuat prediksi, generasi melampaui data yang tersedia.

Inferensi statistik merupakan proses menggunakan sampel untuk menyimpulkan sifat-sifat populasi. Prosedur statistik menggunakan data sampel untuk memperkirakan karakteristik seluruh populasi dari mana sampel itu diambil.



Gambar 1. 1. Pencuplikan dan Inferensi (penarikan kesimpulan)

Populasi biasanya terlalu besar untuk diukur seluruhnya atau sepenuhnya, sehingga peneliti harus menggunakan bagian yang dapat dikelola dari populasi itu untuk mempelajarinya.

Statistik Inferensial dikenal juga sebagai statistik induktif yaitu kumpulan cara atau metode yang dapat menggeneralisir nilai-nilai dari sampel yang sengaja dikumpulkan menjadi nilai populasi. Dengan metode statistik inferens kita dapat mengevaluasi informasi yang telah kita kumpulkan menjadi suatu pengetahuan baru.

Statistik inferensial melibatkan pengujian hipotesis (mengevaluasi beberapa ide tentang populasi menggunakan sampel) dan estimasi (memperkirakan nilai atau rentang potencial nilai dari beberapa karakteristik populasi berdasarkan sampel).

Tabel 1. 1. Perbedaan Statistik Deskriptif dan Statistik Inferensial

No	Perbedaan	Statistik Deskriptif	Statistik Inferensial
1	Tujuan	Menyajikan data	Menarik kesimpulan dari populasi berdasarkan data
2	Hasil akhir	Tabel, grafik, koefisien	Kemungkinan/ probabilitas

3	Uji	Langsung terlihat data tanpa dilakukan pengujian	Dilakukan uji hipotesis dan uji statistik
4	Fokus	Perumusan penyajian data di manfaatkan dalam terminologi statistika sehingga tidak terlalu menekankan kepada aspek stokastik/hubungan pasti yang berbasiskan teori probabilitas.	Data dikaitkan pada distribusi probabilitas untuk menggeneralisir nilai-nilai dari sampel yang sengaja dikumpulkan menjadi nilai populasi.

Rangkuman

Variabel dapat diklasifikasikan dalam banyak cara, misalnya sebagai variabel numerik dan kategori. Cara lain untuk mengklasifikasikan variabel adalah menurut skala pengukurannya.

E. Soal Latihan

Suatu studi mengumpulkan data dari ibu hamil sebagai berikut:

Tabel 1. 2. Data Ibu Hamil di Desa X Tahun 2022

No	Nama Ibu	Pendidikan	Umur	Berat Badan	Kadar Hb	Umur kehamilan (bulan)	Jumlah pemeriksaan kehamilan di pelayanan kesehatan
1	Asvi	SMP	22	54	12.3	4	4
2	Mira	SMA	28	47	11.6	5	2
3	Liana	Perguruan Tinggi	26	48	12.2	6	5
4	Desi	SMA	24	56	11.4	8	3
5	Wendi	SMA	31	44	12.6	7	5
6	Tiara	SMP	29	50	12.0	6	4

No	Nama Ibu	Pendidikan	Umur	Berat Badan	Kadar Hb	Umur kehamilan (bulan)	Jumlah pemeriksaan kehamilan di pelayanan kesehatan
7	Geni	Perguruan Tinggi	27	46	129	2	2
8	Vera	SMA	30	51	12.2	6	3
9	Peni	SMA	34	56	11.4	8	4
10	Yesi	Perguruan Tinggi	36	49	12.6	7	2

Tentukan jenis data dan skala pengukuran untuk setiap variabel tersebut!

BAB 2

PENYAJIAN DATA

A. Tujuan Pembelajaran

Setelah mempelajari Bab ini mahasiswa mampu membuat penyajian data dalam tekstual, tabel dan grafik sesuai dengan skala data.

B. Tekstual

Penyajian data harus dapat meringkas data, sehingga dapat menggambarkan informasi, sederhana, lugas dan komunikatif.

Metode penyajian harus ditentukan sesuai skala data, metode analisis yang akan digunakan, dan informasi yang harus ditekankan. Data yang disajikan tidak sesuai gagal menyampaikan informasi dengan jelas kepada pembaca.

Bahkan ketika informasi yang sama disampaikan, berbeda metode presentasi harus digunakan tergantung pada apa informasi spesifik akan ditekankan. Sebuah metode penyajian harus dipilih setelah dengan hati-hati menimbang keuntungan dan kerugian dari berbagai metode penyajian.

Data dapat disajikan dengan narasi atau tekstual atau tulisan. Penyajian data secara tekstual digunakan apabila datanya tidak besar dan dapat dengan mudah dipahami oleh pembaca hanya pada saat membaca paragraf.

Kumpulan data kecil dapat dengan mudah disajikan melalui presentasi tekstual. Contoh, data sederhana seperti; **ada 30 ibu yang memiliki anak Balita di Desa X, terdiri dari 20 anak perempuan sedangkan 10 anak lainnya laki-laki.** Hal ini lebih mudah dipahami melalui penyajian tekstual. Tidak diperlukan tabel atau grafik untuk menyajikan data ini karena dapat dipahami melalui teks. Jika informasi ini disajikan dalam grafik atau tabel, itu akan menempati ukuran yang tidak perlu ruang di halaman, tanpa meningkatkan pemahaman pembaca tentang data.

Tabel 2. 1. Kelebihan dan Kekurangan Penyajian Data Tekstual

Kelebihan	Kekurangan
1. Memungkinkan untuk menginterpretasikan data secara lebih rinci untuk menjelaskan temuan, menguraikan tren, dan memberikan informasi kontekstual	1. Menghasilkan data yang luas dalam bentuk kata-kata dan paragraph menimbulkan kesulitan bagi pembaca untuk menarik kesimpulan dalam sekejap.
2. Memungkinkan untuk menyajikan data kualitatif yang tidak dapat disajikan dalam bentuk grafik atau tabel.	2. Tidak cocok untuk kumpulan data besar yang memiliki terlalu banyak detail.
3. Membantu dalam menekankan beberapa poin penting dalam data terutama data secara kontekstual.	3. Dalam penyajian tekstual seseorang harus membaca seluruh teks untuk memahami dan memahami poin utamanya.

C. Distribusi Frekuensi

Data yang telah dikumpulkan, diolah, dianalisis harus juga dilakukan penyajian dan interpretasi. Data dapat disusun menurut beberapa cara. Jika data yang dimiliki terdiri dari banyak observasi, kita tidak dapat langsung mendapatkan informasi dari data tersebut. Untuk memudahkan maka data disusun dalam distribusi frekuensi. Distribusi frekuensi merupakan susunan angka menurut besarnya (kuantitas) atau menurut kategorinya (kualitas). Salah satu penyajian distribusi frekuensi adalah dengan menggunakan tabel.

Tabel menyampaikan informasi yang telah diubah menjadi kata atau angka dalam baris dan kolom. Tabel paling tepat untuk menyajikan informasi individu, dan dapat menyajikan informasi kuantitatif dan kualitatif. Tabel penting karena membuat pemahaman data lebih mudah. Jumlah data yang panjang menjadi ringkas dan tepat dengan tabel.

Tabel 2. 2. Kelebihan dan Kekurangan Penyajian Data Tabel

Kelebihan	Kekurangan
1. Dapat secara akurat menyajikan informasi yang tidak dapat disajikan dengan grafik. Angka seperti "132.145852" dapat dinyatakan secara akurat dalam sebuah tabel.	1. Interpretasi informasi membutuhkan waktu lebih lama dalam tabel daripada grafik, dan tabel tidak sesuai untuk mempelajari tren data.
2. Informasi dengan unit yang berbeda dapat disajikan bersama-sama. Misalnya, tekanan darah, detak jantung, jumlah obat yang diberikan, dan waktu anestesi dapat disajikan bersama dalam satu tabel.	2. Tidak mudah untuk mengidentifikasi dan selektif memilih informasi yang dibutuhkan.
3. Tabel berguna untuk meringkas dan membandingkan informasi kuantitatif dari variabel yang berbeda.	

Struktur tabel dalam penelitian terdiri dari: (1) Judul, (2) Tubuh, dan (3) *Headings (Stub dan Box Head)*.

Tabel ##. Judul

STUB	BOX HEAD			Total
BODY				
Total				Grand Total

Catatan kaki:

Sumber data:

Judul

Setiap tabel memiliki judul dan judul ini harus ringkas tetapi tepat. Identitas numerik pada tabel juga harus diberikan. Nomor bab harus mendahului nomor tabel agar mudah ditemukan di kemudian hari.

Judul perlu menjelaskan dengan jelas apa yang akan dilihat pembaca dalam tabel. Biasanya ditempatkan di atas tabel, harus jelas, singkat, dan lengkap. Judul yang baik akan menjawab: *what, when, dan where*.

Stub

Merupakan kolom paling kiri, termasuk kepala kolom tersebut. *Stub* memberi suatu keterangan/penjelasan secara terperinci tentang gambaran pada tiap baris dan badan tabel.

Box head

Box head Ini termasuk kepala kolom. *Box head* memberi keterangan/penjelasan secara terperinci tentang gambaran tiap kolom dari badan tabel

Body/Tubuh

Tubuh terdiri dari *sel-sel* (kolom dan baris) yang berisi data utama yang dianalisis.

Jenis-Jenis Tabel

1) Tabel Induk (*Master Table*)

Merupakan tabel yang berisikan semua hasil pengumpulan data masih dalam bentuk data asli atau mentah (*raw data*) tersedia secara rinci agar seluruh informasi dapat dibaca. Berikut contoh tabel induk:

Tabel 2. 3. Daftar Balita yang Berkunjung ke Puskesmas Tahun 2022

Nomor Rekam Medis	Usia (bln)	Jenis Kelamin	Berat Badan (kg)	Tinggi Badan (cm)	Lingkar kepala (cm)	ASI Ekslusif
112021	10	L	10,2	76	44,2	Ya
113045	12	P	11,1	75	46,5	Tidak

120130	20	L	13,00	80	47,0	Ya
123031	18	L	10,5	78	46,8	Ya
130001	11	P	8,25	73	45,5	Tidak
132003	14	P	12,1	77	46,5	Ya
140900	8	L	7,5	74	45,0	Ya
141010	13	P	9,0	75	43,4	Ya
142000	16	L	12,5	79	47,5	Tidak

2) Tabel Distribusi frekuensi

a) Tabel distribusi frekuensi

Distribusi frekuensi menyediakan kumpulan data yang informatif dan ringkas. Data yang diringkas memungkinkan penilaian pembaca yang lebih cepat dari sampel yang dijelaskan daripada membaca pernyataan kuantitatif individu. Distribusi frekuensi memberikan informasi kategoris tentang jumlah kejadian.

Metode penyusunan distribusi frekuensi data kuantitatif

1. Cari nilai maksimum dan minimum (Selisih nilai maksimum dan minimum disebut *Range* =R)
2. Tentukan jumlah kelas dan interval kelas.

Jumlah kelas: (rumus Sturgess)

$$M = 1 + 3,3 \log N$$

M = Jumlah kelas

N = Jumlah data (observasi)

Interval kelas: R/M

3. Hitung banyak observasi yang termasuk ke dalam setiap kelas, disebut frekuensi.

Data dibawah ini adalah usia dari 150 orang pengguna BPJS di Rumah Sakit Y di Samarinda Tahun 2022.

Tabel 2. 4. Usia Pengguna BPJS Rumah Sakit Y Samarinda tahun 2022

25	33	35	26	29	26	25	27	22
25	22	38	25	23	30	25	25	26
26	26	26	35	22	29	35	28	37
23	36	30	39	28	42	35	32	30
40	33	23	40	44	30	40	35	24
43	30	22	23	24	22	25	19	33
25	21	21	30	22	22	27	25	33
30	31	30	28	28	40	40	24	30
33	33	29	30	29	29	37	30	30
28	28	22	34	27	39	31	36	23
26	30	21	37	26	25	30	31	35
36	20	20	37	36	31	30	43	25
31	31	25	27	32	20	25	32	32
39	30	31	43	24	24	23	35	23
32	28	30	21	34	43	20	35	31
35	34	37	28	40	33	37	38	24
27	32	26	28	27	38			

Dari data diatas sulit diperoleh informasi mengenai usia rumah sakit tersebut. Akan lebih mudah untuk distribusi frekuensinya

Berdasarkan tabel 2.1 disusun distribusi frekuensi dengan langkah-langkah sebagai berikut:

1. Nilai maksimum dan minimum

Nilai minimum: 19

Nilai maksimum: 44

2. Range: $44-19 = 25$

Jumlah kelas dan interval kelas

Jumlah kelas (M): (rumus Sturgess)

$$M = 1 + 3,3 \log N = 1 + 3,3 \log 150 = 8,18$$

$$\text{Interval kelas} = R/M = 25/8,18 = 3,05$$

3. Masukkan data atau angka hasil setiap observasi umur yang telah dikategorikan dalam interval 3, hasilnya sebagai berikut

Tabel 2. 5. Distribusi Frekuensi Responden Berdasarkan Usia Pengguna BPJS

Rumah Sakit Samarinda Tahun 2022

Umur	Jumlah
19-21	9
22-24	21
25-27	28
28-30	31
31-33	21
34-36	16
37-39	12
40-42	7
43-45	5
Total	150

- b) Tabel distribusi frekuensi relatif (%)

Distribusi frekuensi dapat diubah menjadi distribusi frekuensi relatif, dengan cara menghitung bilangan frekuensi dalam bentuk persentase. Disebut sebagai “frekuensi relatif” karena frekuensi yang disajikan di sini bukanlah frekuensi yang sesungguhnya, melainkan frekuensi yang disajikan dalam bentuk angka persentase (%).

Dari tabel 2.2 dapat dibuat menjadi tabel distribusi frekuensi relatif sebagai berikut:

Tabel 2. 6. Distribusi Frekuensi Relatif Usia Pengguna BPJS Rumah Sakit Samarinda Tahun 2022

Umur	Jumlah	Frekuensi (%)
19-21	9	6,0
22-24	21	14,0
25-27	28	18,7
28-30	31	20,7
31-33	21	14,0
34-36	16	10,7
37-39	12	8,0
40-42	7	4,7
43-45	5	3,2
Total	150	100,0

- c) Tabel distribusi frekuensi kumulatif relatif

Frekuensi kumulatif ialah salah satu jenis tabel yang di dalamnya disajikan frekuensi yang terus meningkat atau selalu ditambah-tambahkan, baik dari bawah ke atas maupun dari atas ke bawah. Dari tabel 2.2 dapat dibuat menjadi tabel distribusi frekuensi kumulatif sebagai berikut:

Tabel 2. 6. Distribusi Frekuensi Relatif Usia Pengguna BPJS RS. Samarinda Tahun 2022

Umur	Jumlah	Frekuensi Kumulatif	Frekuensi Relatif Kumulatif
19-21	9	9	6%
22-24	21	30	20%
25-27	28	58	38,7%
28-30	31	89	59,4%
31-33	21	110	73,4%
34-36	16	126	84,1%
37-39	12	138	92,1%

Umur	Jumlah	Frekuensi Kumulatif	Frekuensi Relatif Kumulatif
40-42	7	145	96,8%
43-45	5	150	100%
Total	150		

Berdasarkan tabel 2.6 diketahui bahwa proporsi pengguna BPJS dengan umur dibawah 31 tahun adalah 59,4 persen.

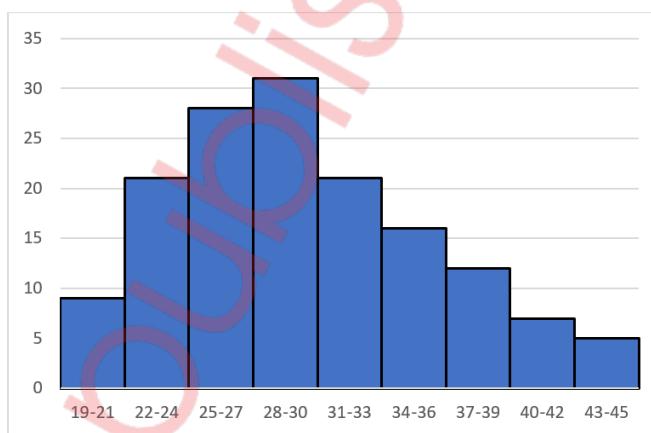
D. Grafik

Jenis-Jenis Grafik

1) Histogram

Histogram adalah tampilan grafis data menggunakan bar dengan ketinggian yang berbeda. Dalam histogram, setiap batang mengelompokkan angka ke dalam rentang. Batang yang lebih tinggi menunjukkan bahwa lebih banyak data berada dalam kisaran tersebut. Histogram menampilkan bentuk dan **penyebaran data dengan skala kontinu**. Akan lebih mudah apabila membuat histogram berdasarkan data yang telah dibuat dalam tabel distribusi frekuensi.

Berikut adalah contoh histogram yang dibuat berdasarkan data pada tabel 2.4 atau tabel 2.5



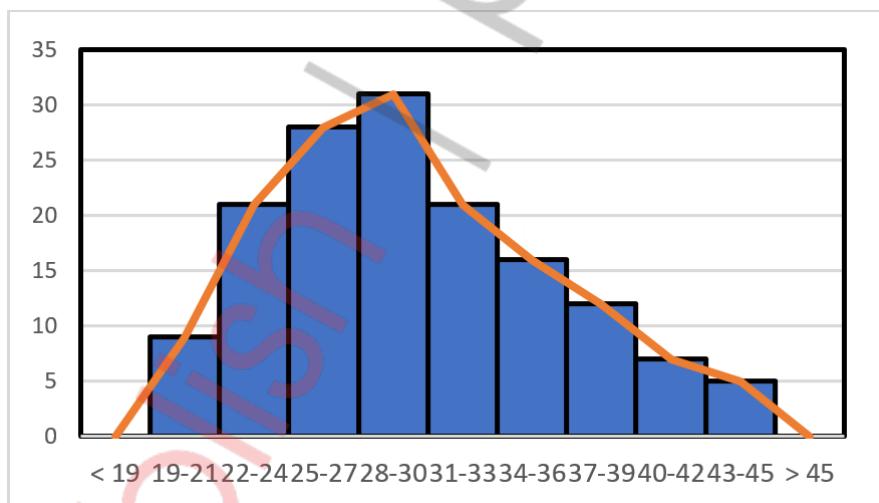
Gambar 2. 1. Distribusi Usia Pengguna BPJS RS.Samarinda 2022

2) Frekuensi Poligon

Dipergunakan untuk menyajikan suatu distribusi frekuensi dari data yang kontinu. Permukaan area frekuensi poligon sama luasnya dengan histogram

Dipergunakan untuk membandingkan sejumlah distribusi frekuensi pada sebuah gambar

Langkah-langkah membuat polygon: (1) membuat titik-titik tengah kelas interval yang berada pada bagian atas bar histogram. (2) menghubungkan titik dengan membentuk garis yang disebut frekuensi polygon, (3) frekuensi poligon harus tertutup di kedua ujungnya, sebab merupakan area diagram dan harus sama luasnya dengan bar histogram. Berikut adalah contoh histogram yang dibuat berdasarkan data pada tabel 2.4 atau tabel 2.5



Gambar 2. 2. Distribusi Usia Pengguna BPJS RS.Samarinda 2022

3) Ogive

Ogive merupakan grafik distribusi kumulatif, yang menjelaskan nilai data pada sumbu bidang horizontal dan frekuensi relatif kumulatif, frekuensi kumulatif atau frekuensi persen kumulatif pada sumbu vertikal.

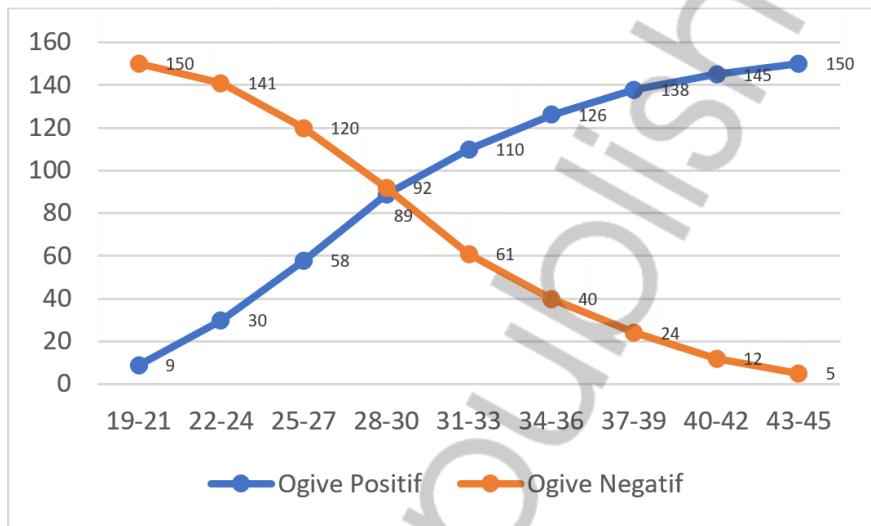
Frekuensi kumulatif didefinisikan sebagai jumlah dari semua frekuensi sebelumnya hingga titik saat ini. Untuk menemukan popularitas data yang diberikan atau kemungkinan data yang berada dalam rentang frekuensi tertentu, kurva Ogive membantu menemukan detail tersebut secara akurat.

Ogive dibuat dengan memplot titik yang sesuai dengan frekuensi kumulatif setiap interval kelas. Ogive sebagian digunakan untuk menggambarkan data dalam representasi bergambar. Ini membantu dalam memperkirakan jumlah pengamatan yang kurang dari atau sama dengan nilai tertentu.

Tabel 2. 8. Distribusi Frekuensi Kumulatif Usia Pengguna BPJS RS. Samarinda
Tahun 2022

Umur	Jumlah	Frek.	Frek.
		Kumulatif Kurang dari	Kumulatif Lebih dari
19-21	9	9	150
22-24	21	30	141
25-27	28	58	120
28-30	31	89	92
31-33	21	110	61
34-36	16	126	40
37-39	12	138	24
40-42	7	145	12
43-45	5	150	5
Total	150		

Dari tabel distribusi kumulatif kurang dan lebih dari yang telah dibuat dapat di plot menjadi ogive berikut:



Gambar 2. 3. Ogive Positif dan Negatif Usia Jumlah Pengguna BPJS RS Samarinda 2022

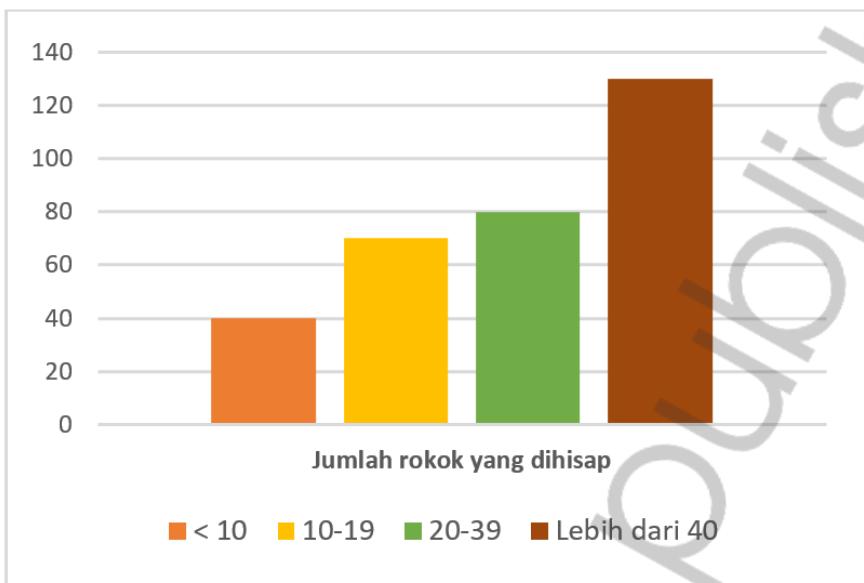
Dari gambar ogive kita dapat melihat bahwa 89 orang pengguna BPJS di RS Samarinda berumur kurang dari 30 tahun, lebih banyak pengunjung yang berusia diatas 30 tahun.

4) Diagram batang (*bar diagram*)

Dipergunakan untuk menyajikan suatu frekuensi data diskret atau data dengan skala nominal maupun ordinal. Berbeda dengan histogram, grafik batang terdiri balok terpisah. Beragam kategori diwakili sumbu horizontal (sumbu X), sedangkan sumbu Y atau sumbu vertical mewakili jumlah atau frekuensi.

a) Single bar

Digunakan jika klasifikasi didasarkan pada atribut-atribut dan jika atribut-atribut tersebut akan dibandingkan sehubungan dengan satu karakter. Contoh: Jumlah rokok yang dikonsumsi per hari, Jumlah Baduta stunting di berbagai wilayah yang berbeda, efek dari perawatan yang berbeda dan lain-lain. Berikut merupakan contoh dari single bar:

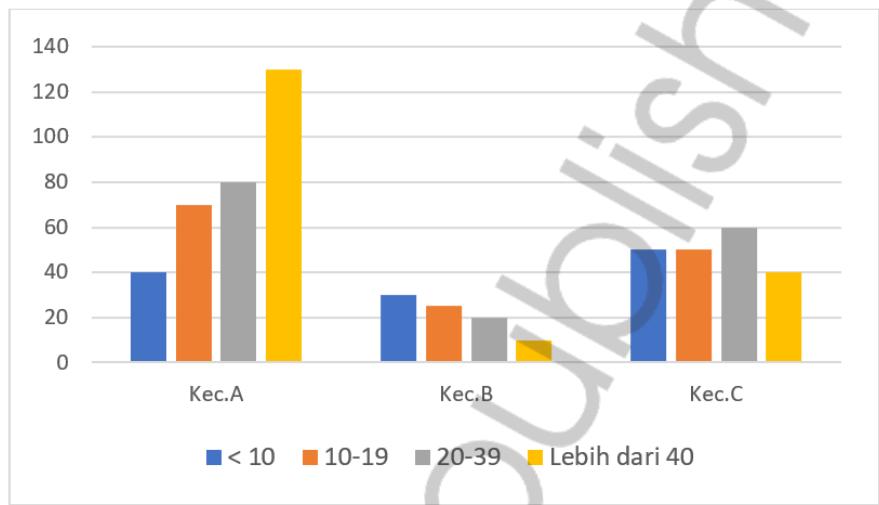


Gambar 2. 4. Gambaran Perilaku merokok pada remaja di Kecamatan A tahun 2022

Gambar 2.4 memperlihatkan 130 orang memiliki kebiasaan menghisap rokok berjumlah lebih dari 40 batang dalam 1 minggu.

b) Multiple bar

Diagram batang ganda/*multiple* merupakan variasi dari diagram batang. Untuk setiap atribut, dua atau lebih batang yang mewakili karakter atau kelompok terpisah harus ditempatkan berdampingan. Setiap batang dalam atribut akan ditandai atau diwarnai secara berbeda untuk membedakannya. Jenis penandaan atau pewarnaan yang sama harus dilakukan di bawah setiap atribut. Catatan kaki harus diberikan untuk menjelaskan tanda atau pewarnaan.

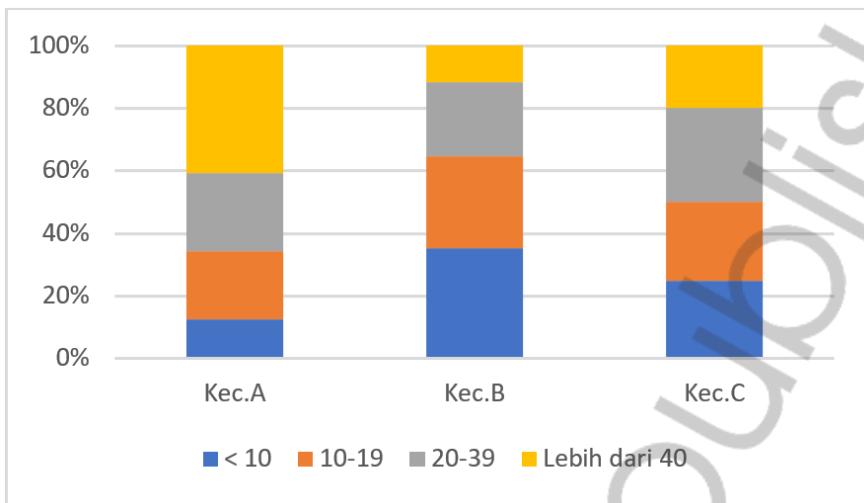


Gambar 2. 5. Gambaran Perilaku merokok pada remaja di Kecamatan A, B dan C Kabupaten X tahun 2022

Gambar 2.5 menunjukkan di ketiga kecamatan lebih banyak remaja memiliki kebiasaan menghisap rokok berjumlah lebih dari 40 batang dalam 1 minggu.

c) Component bar

Cara lain untuk menyajikan data tentang item yang berbeda adalah dengan menggunakan diagram batang komponen, seperti Gambar 6.1 di bawah ini. Alih-alih menggunakan dua atau tiga batang terpisah seperti dalam beberapa bagan batang, semua data untuk setiap tahun dimuat dalam satu batang. Setiap komponen, atau bagian, dari batang diarsir secara berbeda. Sebuah legenda sekali lagi digunakan untuk memberikan kunci makna.

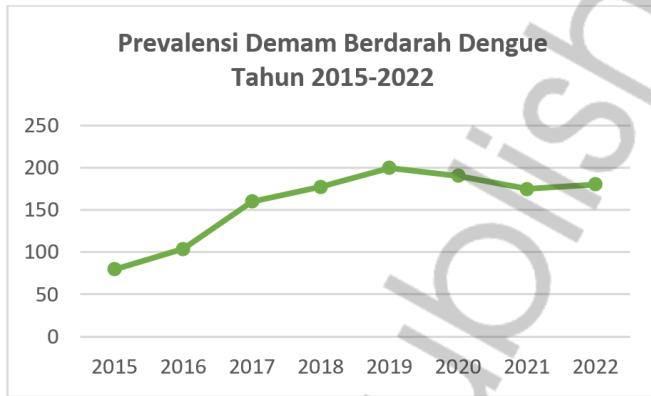


Gambar 2. 6. Gambaran Perilaku merokok pada remaja di Kecamatan A, B dan C Kabupaten X tahun 2022

Gambar 2.6 memperlihatkan bahwa proporsi remaja yang merokok lebih dari 40 batang per hari paling tinggi berada di Kecamatan A.

5) Diagram Garis (line diagram)

Diagram garis adalah perbandingan visual tentang bagaimana dua variabel—ditampilkan pada sumbu X dan Y—terkait atau berbeda satu sama lain. Sumbu horizontal menggambarkan perkembangan berkelanjutan, sering kali juga menggambarkan waktu. Sumbu vertikal atau sumbu Y merepresentasikan frekuensi dan sumbu X menggambarkan variabel yang mempengaruhi sumbu Y, atau dapat merepresentasikan waktu serta periode.

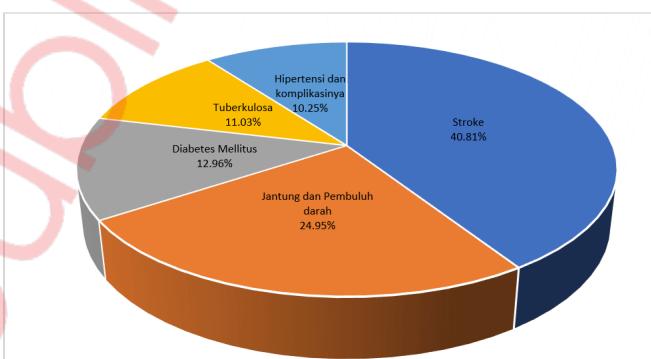


Gambar 2. 7. Prevalensi Demam Berdarah Dengue Tahun 2015-2022

Gambar 2.7 menunjukkan terjadi kecenderungan peningkatan prevalensi Demam Berdarah Dengue meningkat dari tahun 2015 sampai 2022 dengan tertinggi terdapat pada tahun 2019.

6) Diagram Pinca (pie diagram)

Umumnya data kategorik ditampilkan dalam bentuk diagram pie. Satu lingkaran dibagi dalam beberapa irisan yang sesuai dengan distribusi proporsi atau persentase masing-masing kategori. Diagram Pie bermanfaat dalam menyampaikan data yang terdiri dari kategori yang terbatas.

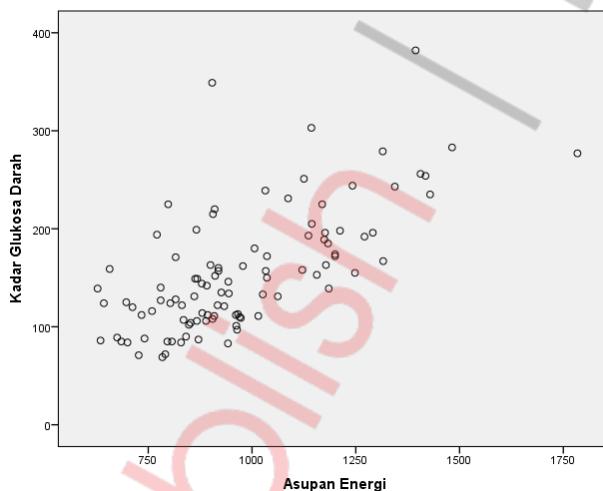


Gambar 2. 8. Penyebab Kematian Tertinggi di Kab. X Tahun 2022

Gambar 2.8 memperlihatkan bahwa stroke merupakan penyebab kematian tertinggi di Kabupaten X pada tahun 2022.

7) Diagram Tebar (scatter diagram)

Diagram tebar dapat mengidentifikasi kemungkinan hubungan antara perubahan yang diamati dalam dua variabel berskala numerik. Manfaat lain dari penggunaan diagram tebar adalah: menunjukkan apa yang terjadi pada variabel ketika yang lain diubah, menganalisis apakah dua variabel saling terkait, serta memberikan visual dan statistik untuk menguji kekuatan hubungan antara dua variabel tersebut.

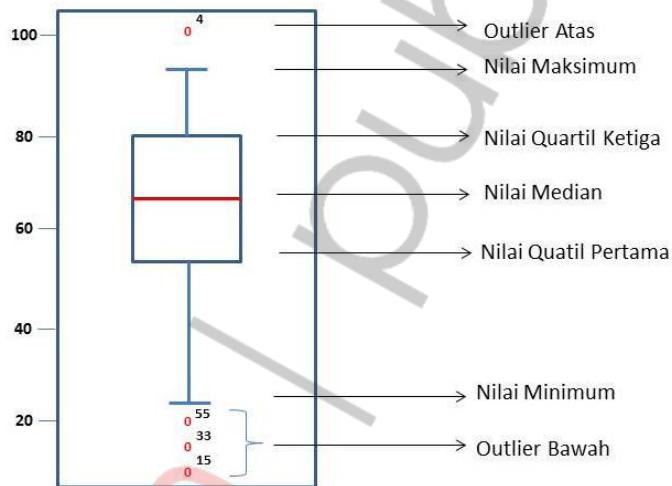


Gambar 2.9. Korelasi Asupan Energi dengan Kadar Gula dalam Darah

Dari gambar 2.9 diketahui bahwa terdapat korelasi asupan energi dengan kadar gula dalam darah.

8) Box Whisker Plot

Box plot atau a dikenal sebagai diagram box-and-whisker ialah suatu kotak persegi panjang dalam statistik deskriptif merupakan media untuk menyajikan secara grafik dari data numerik. Dalam penyajian *boxplot* dapat digambarkan secara horizontal maupun vertikal sebagaimana di bawah ini:

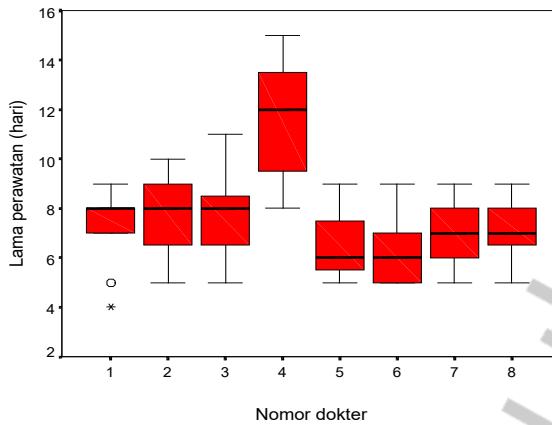


Gambar 2. 10. Bagian dari Box

Kotak (Box) terdiri dari:

- (1) Garis tengah adalah nilai kuartil 2 (Q2) atau median
- (2) Garis bawah adalah nilai kuartil 1 (Q1)
- (3) Garis bawah adalah nilai kuartil 3 (Q3)

Tali (whisker) batas bawah adalah nilai batas yang tidak lebih perbedaannya dari Q1 sebanyak $1\frac{1}{2}x(Q3-Q1)$ atau perbedaan *inter quartile*, sedangkan batas atas nilai yang paling jauh dan tidak lebih dari sebanyak $1\frac{1}{2}x(Q3-Q1)$. Tanda bintang adalah nilai yang menjadi pencikan (*outlier*), selanjutnya adanya lingkaran kecil adalah kandidat untuk *outlier*.



Gambar 2. 11. Gambaran Lama Perawatan Dokter di RS X.

Dari gambar 2.11 diketahui bahwa dokter nomor 4 memiliki median lama perawatan tertinggi dibandingkan dokter lainnya yaitu 12 hari.

9) Steam and Leaf Plot

Penyajian data dalam bentuk distribusi akan menghilangkan nilai aslinya dari data tersebut. Untuk menghilangkan kelemahan ini dibuat suatu penyajian yang disebut *stem and leaf*.

Berikut merupakan umur Ibu Balita di Posyandu Mawar yang telah disusun:

19	20	20	20	21	21	22
23	23	30	31	32	32	33
34	35	36	37	38	40	40
41	41	42	42			

BATANG	DAUN	Frekuensi
10	9 1	
20	0,0,0,1,1,2,3,3, 8	
30	0,1,2,2,3,4,5,6,7,8 10	
40	0,0,1,1,2,2, 6	

Gambar 2. 12. Umur 25 Ibu Balita di Posyandu Mawar

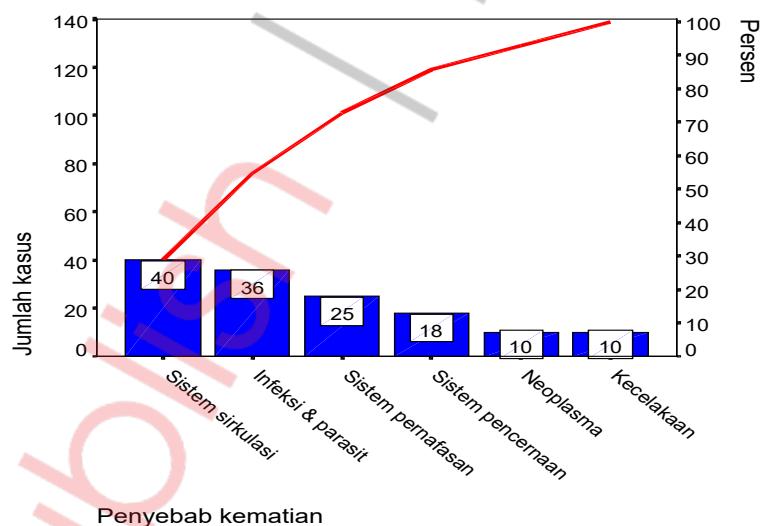
10) Pareto

Pareto merupakan dari diagram batang yang disusun dengan susunan tinggi rendahnya batang sehingga dengan mudah dapat diinterpretasi.

Langkah membuat Pareto:

- 1) Membuat daftar masalah sesuai dengan urutan frekuensi kejadian (dari tertinggi sampai terendah).
- 2) Menghitung Frekuensi kumulatif dan Persentase kumulatif
- 3) Gambarkan Frekuensi dalam bentuk grafik batang
- 4) Gambarkan kumulatif Persentase dalam bentuk grafik garis
- 5) Interpretasikan (terjemahkan) Pareto Chart tersebut.

Berikut merupakan diagram pareto distribusi penyebab kematian di Provinsi X.



Gambar 2. 12. Penyebab Kematian di Provinsi X

Proporsi penyebab kematian tertinggi di Provinsi X teridentifikasi adalah Sistem sirkulasi yaitu sebesar 40%.

E. Rangkuman

Penyajian data bertujuan untuk memudahkan menyampaikan informasi kepada sasaran atau khalayak. Data dapat disajikan dengan berbagai bentuk disesuaikan dengan kebutuhan serta jenis data. Baik tabel maupun grafik harus dibuat dengan lengkap, jelas serta mampu menjelaskan dirinya sendiri

F. Soal Latihan

Sajikan data pada Tabel 1.1 dengan penyajian yang paling sesuai dengan jenis datanya.

BAB 3

UKURAN SENTRAL DAN TENDENSI

A. Tujuan Pembelajaran

Setelah mempelajari Bab ini mahasiswa mampu menghitung ukuran sentral dan tendensi, mampu memilih teknik statistik deskriptif yang sesuai untuk meringkas kumpulan data serta mampu membuat ringkasan data dan mampu menghitung ukuran proporsi

B. Ukuran Tendensi Sentral

Dalam statistik, konsep dari nilai perwakilan disebut Tendensi Sentral. Tendensi Sentral adalah pengukuran statistik untuk menentukan nilai tunggal yang menggambarkan pusat distribusi dari suatu gugus data. Tujuan dari Tendensi Sentral adalah untuk mencari nilai tunggal yang paling mendekati atau paling mewakili seluruh kelompok.

1. Mean (Rerata)

Rata-rata Hitung atau *arithmatic mean* atau lebih dikenal dengan mean saja adalah nilai yang baik mewakili suatu data. Nilai ini sangat sering dipakai dan malah yang paling banyak dikenal dalam menyimpulkan sekelompok data.

Misalnya kalau kita mempunyai n pengamatan yang terdiri dari $x_1, x_2, x_3, \dots, x_n$, maka nilai rata-rata adalah:

$$\bar{x} = \sum \frac{x_i}{n} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

Simbol untuk mean

1. \bar{x} (x bar) = untuk sampel (statistik)
2. μ (miu) = untuk populasi (parameter)

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

Sifat dari Rerata diantaranya: tidak selalu satu nilai menjelaskan Tendensi Sentral dan tidak selalu satu nilai dapat mewakili nilai keseluruhan dalam seluruh situasi.

Contoh untuk gambar berikut, bagaimana kita memutuskan nilai mana yang **paling baik** untuk digunakan.

Solusinya dapat digunakan 3 metode untuk mengukur Tendensi Sentral: Mean, Median dan Modus.

Contoh:

Diketahui tinggi badan tujuh orang dewasa 148, 152, 156, 162, 168, 170, 172 centimeter
Rerata tinggi badan tujuh orang ini adalah:

$$\text{Rerata} = \frac{156 + 162 + 152 + 148 + 168 + 170 + 172}{7}$$

Rerata data berkelompok:

$$\chi = \chi_g + \frac{\sum(f \cdot d)}{n}$$

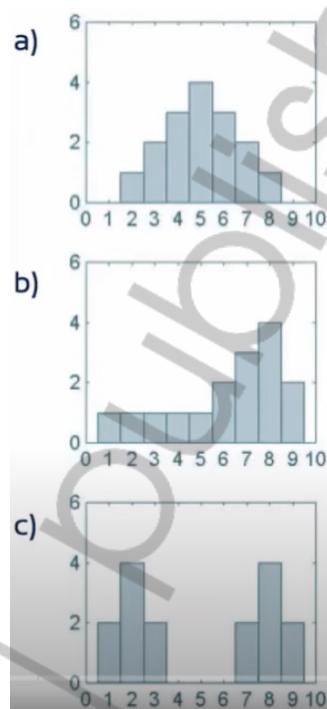
χ_g = guess mean (rata-rata sementara)

n = jumlah seluruh pengamatan

f = frekuensi

d = $(mp - \chi_g)$ selisih nilai tengah kelas (mp)

dengan rata-rata sementara (χ_g)



Berapakah rerata tinggi badan mahasiswa dari data di bawah ini:

Tabel 3. 1. Distribusi Tinggi Badan Siswa

Kelas	Frekuensi
150—154	5
155—159	10
160—164	15
165—169	35
170—174	20
175—179	10
180—185	5
Total	100

Langkah pertama adalah membuat *mid point* atau kelas tengah dari data, sebagai contoh:

- (1) Mid point dari kelas 150-155 diperoleh dari $(150 + 155)/2 = 152,5$
- (2) Mid point dari kelas 155-159 diperoleh dari $(155 + 160)/2 = 157,5$

Dan seterusnya, setelah itu tentukan rata-rata sementara adalah nilai tengah kelas keempat, yaitu 165-169. Nilai tengah kelas tersebut diberi nilai (Nol) selanjutnya hitung selisih (deviasi/d) masing-masing nilai *mid point* (mp) terhadap frekuensi nilai rata-rata sementara. Langkah terakhir kalikan frekuensi dengan deviasi (d).

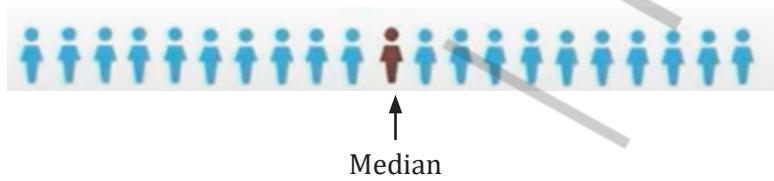
Kelas	f	mp	d (mp-167,5)	f.d
150—154	5	152,5	-15	-75
155—159	10	157,5	-10	-100
160—164	15	162,5	-5	-75
165—169	35	167,5	0	0
170—174	20	172,5	5	100

Kelas	f	mp	d (mp-167,5)	f.d
175—179	10	177,5	10	100
180—185	5	182,5	15	75
Total	100			$\Sigma=25$

$$X = 167,5 + \frac{25}{100} = 167,5$$

2. Median (Nilai Tengah)

Median adalah nilai yang terletak pada observasi yang di tengah, kalau data tersebut telah disusun (*array*). Nilai median disebut juga nilai letak di mana merupakan nilai yang membagi 2 (50%) distribusi data yang telah diurutkan



Posisi median adalah:

$$= \frac{N+1}{2}$$

Nilai median adalah nilai pada posisi tersebut. Contoh: Diketahui tinggi badan tujuh orang dewasa

148, 152, 156, 162, 168, 170, 172 centimeter

$$= \frac{N+1}{2} + \frac{7+1}{2} = 4$$

Nilai observasi ketiga adalah 162, maka dikatakan median adalah 162 centimeter. Jika datanya berjumlah genap maka posisi median terletak antara dua nilai, maka nilai median adalah rata-rata dari kedua nilai tersebut.

Median Data Berkelompok

Langkah menentukan median data per kelompok: (1) Tentukan posisi median = $(n/2)$ dan (2) Hitung nilai median dg rumus. Rumus Median adalah sebagai berikut:

$$Md = L_{md} + \frac{i * \left[\frac{n}{2} - f_k \right]}{f_{md}}$$

Lmd = Lower limit kelas median

i = interval kelas

n = jumlah pengamatan

fk = frekuensi kumulatif sebelum kelas median

fmd = frekuensi kelas median

Dari tabel 3.1 dapat dihitung median sebagai berikut:

Kelas	f	fk
150—154	5	5
155—159	10	15
160—164	15	30
165—169	35	65
170—174	20	85
175—179	10	95
180—185	5	100
Total	100	

- 1) Nilai Md berada pada kelas 165—169
- 2) Kelas tersebut disebut **kelas median**
- 3) $L_{md} = 165$ $f_{md} = 35$
 $fk = 30$ Interval = 5
 $Md = 165 + \{5(50-30)/35\} = 167,86$

3. Mode (Modus)

Merupakan nilai yang paling sering muncul atau nilai frekuensi pengamatan yang paling besar. Nilai observasi yang frekuensinya paling banyak dinyatakan dengan puncak dari kurva distribusinya. Kurva satu puncak disebut unimodal), dua puncak disebut bimodal, banyak puncak disebut multimodal.

Contoh: Dari pengamatan tinggi badan 10 orang dewasa muda didapatkan data sebagai berikut:

152, 153, 155, 155, 155, 156, 157, 160, 162, 162 centimeter.

Dari pengamatan tersebut ditemui nilai 155 cm sebanyak tiga kali.

Dengan demikian maka nilai median adalah 55 centimeter.

Modus pada data berkelompok

$$Mo = L_{mo} + \left(\frac{i * \Delta_1}{\Delta_1 + \Delta_2} \right)$$

L_{mo} = Lower limit kelas mode

i = interval kelas

Δ_1 = Selisih frekuensi kelas mode dengan frekuensi kelas sebelum kelas mode

Δ_2 = Selisih frekuensi kelas mode dengan frekuensi kelas setelah kelas mode

Dari tabel 3.1 berikut nilai modus:

Kelas	Frekuensi
150—154	5
155—159	10
160—164	15
165—169	35
170—174	20
175—179	10
180—185	5
Total	100

Frekuensi terbanyak pada kelas 165-169

Berarti kelas tsb disebut **kelas mode**

$$L_{Mo} = 165$$

$$\Delta_1 = 35 - 15 = 20$$

$$\Delta_2 = 35 - 20 = 15$$

$$i = 5$$

$$\text{Nilai Mo} = 165 + \{(5 \times 20) / (20 + 15)\} = 167,86$$

C. Ukuran Variabilitas

Dengan melihat pada nilai rata-rata saja informasi yang ada dapat menuju pada kesalahan interpretasi. Sebagai contoh dari dua kelompok data diketahui rata-ratanya sama, jika hanya dari informasi ini dinyatakan bahwa dua kelompok ini sama namun pada kenyataannya variasinya data di dalam kelompok masing-masing bisa berbeda.

Nilai variasi atau deviasi adalah nilai yang menunjukkan bagaimana bervariasi data di dalam kelompok data itu terhadap nilai rata-ratanya., sehingga makin besar nilai variasi maka makin bervariasi pula data tersebut.

1. Range (Nilai Kisaran)

Range adalah nilai yang menunjukkan perbedaan nilai pengamatan yang paling besar dengan nilai yang paling kecil.

Contoh:

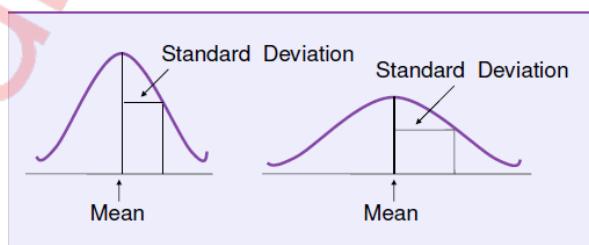
Diketahui tinggi badan tujuh orang dewasa

148, 152, 156, 162, 168, 170, 172 centimeter

Range-nya adalah: $172 - 148 = 24$

2. Varians dan Standar Deviasi (Simpang Baku)

Varians adalah rata-rata perbedaan antara mean dengan nilai masing-masing observasi.



Rumus:

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

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

Contoh: dari data tinggi badan tujuh orang dewasa 148, 152, 156, 162, 168, 170, 172 centimeter dapat dihitung Varian dan standar deviasi:

Varians

$$= \frac{(148 - 161.14)^2 + (152 - 161.14)^2 + (156 - 161.14)^2 + (162 - 161.14)^2 + (168 - 161.14)^2 + (170 - 161.14)^2 + (172 - 161.14)^2}{6}$$

$$= 87.81$$

$$\text{Standar Deviasi} = \sqrt{87.81} = 9,37$$

3. Quartile

Suatu ukuran yang membagi distribusi data menjadi 4 (empat) bagian yang sama

- 1) Kuartil 1 (K1) merupakan nilai yang membagi jumlah pengamatan menjadi dua bagian, yaitu sebesar 25% dibawah K1 dan 75% diatas K1.
- 2) K2 atau sama dengan Median merupakan nilai yang membagi jumlah pengamatan menjadi dua bagian yaitu sebesar 50% dibawah K2 dan 50% diatas K2.
- 3) K3 atau sama dengan Median merupakan nilai yang membagi jumlah pengamatan menjadi dua bagian yaitu sebesar 75% dibawah K2 dan 25% diatas K2
- 4) **Nilai kuartil** adalah nilai observasi pada posisi kuartil tersebut

$$\text{Posisi - Kuartile} = \frac{k_i}{4}(n)$$

4. Interquartile Range (Jarak Antar Kuartil)

Hampir sama dengan range, hanya ditentukan oleh selisih (K3) kuartile-3 dengan (K1) kuartile-1. Dihitung dengan rumus:

$$IQR = K3 - K1$$
$$DQ = \frac{K3 - K1}{2}$$

5. Desil

Suatu ukuran yang membagi kumpulan data menjadi 10 bagian yang sama. Prinsipnya sama dengan kuartil, hanya dengan desil kita dapat memperoleh informasi yang lebih detail. Perhitungan desil

$$Posisi - Desil = \frac{d_i}{10}(n)$$

Nilai desil adalah nilai observasi pada posisi desil tersebut.

6. Persentil

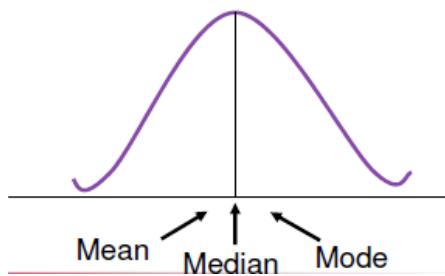
Suatu ukuran yang membagi kumpulan data menjadi 100 bagian yang sama. Prinsipnya sama dengan kuartil/desil, hanya dengan persentil kita dapat memperoleh informasi yang lebih detail lagi. **Nilai persentil** adalah nilai observasi pada posisi persentil tersebut.

$$Posisi - Persentil = \frac{p_i}{100}(n)$$

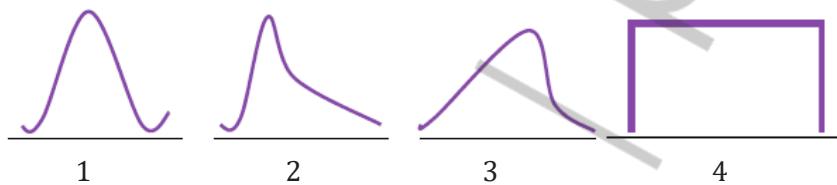
D. Normalitas Data

Kelompok data yang dikumpulkan selanjutnya dilakukan penyebaran, maka berpeluang untuk terdistribusi secara normal. Data yang terdistribusi normal tersebut memiliki sebaran pola yang **normal/terarah**.

Nilai mean, median dan modus berkaitan dengan distribusi data atau normalitas data.



1. Jika $\chi = Md = Mo$ (berhimpitan), maka kurvanya disebut simetris.
2. Jika $\chi > Md$ atau $Md < \chi$, kurvanya menceng ke kanan
3. Jika $\chi < Md$, kurvanya menceng/miring ke kiri
4. Jika semua observasi (nilai pengamatan) mempunyai frekuensi yang sama



E. Ukuran Proporsi atau Persentase

Proporsi merupakan suatu pecahan (fraksi) di mana numerator (pembilang) adalah bagian dari denominator (penyebut) atau dengan perkataan lain, proporsi merupakan perbandingan sebagian terhadap keseluruhan. Proporsi merupakan salah satu bentuk dari rasio. Proporsi digunakan untuk melihat komposisi suatu variabel dalam populasinya. Proporsi dapat dalam bentuk desimal, pecahan atau persentase (%).

Ciri proporsi: (1) Tidak mempunyai satuan (dimensi), karena satuan dari pembilang dan penyebutnya sama, sehingga saling meniadakan dan (2) Nilainya antara 0 sampai 1 atau 0 – 100%.

$$\text{Proporsi} = \frac{\text{Jumlah orang atau peristiwa dengan karakteristik khusus}}{\text{Total keseluruhan peristiwa yang ada pada pembilang}} \times k$$

F. Rangkuman

Dalam menggambarkan data dengan menggunakan ukuran ringkas, penting untuk memilih ukuran tendensi sentral yang paling akurat mewakili data.

G. Soal Latihan

1. Menggunakan data: 3, 4, 6, 1, 10, 6

a. Berapakah mean, median dan range?

b. Berapakah varians dan standar deviasinya?

Hitung mean, median, modus, standar deviasi dan varians.

Untuk data-data berikut pada nomor 2 dan 3:

2. Dua puluh siswa secara acak diperiksa kadar kolesterolnya dengan hasil sebagai berikut:

3. 260	4. 210	5. 244	6. 233	7. 269
8. 158	9. 221	10. 198	11. 214	12. 246
13. 164	14. 225	15. 254	16. 184	17. 206
18. 209	19. 213	20. 179	21. 257	22. 221

3. Selanjutnya berikut hasil konsumsi vitamin C dari dua puluh lima siswa dengan hasil sebagai berikut:

23. 760	24. 790	25. 715	26. 750	27. 785
28. 735	29. 780	30. 760	31. 785	32. 730
33. 740	34. 715	35. 735	36. 770	37. 740
38. 715	39. 770	40. 700	41. 725	42. 785
43. 700	44. 790	45. 735	46. 705	47. 740

BAB 4

TABULASI SILANG

A. Tujuan Pembelajaran

Setelah mempelajari bab ini, mahasiswa mampu menghitung persentase baris, kolom dan persentase total serta interpretasinya.

B. Pengertian Tabulasi

Tabulasi dalam Kamus Besar Bahasa Indonesia (KBBI) adalah penyusunan menurut lajur yang telah tersedia; penyajian data dalam bentuk tabel atau daftar untuk memudahkan pengamatan dan evaluasi. Dalam proses pengolahan data tabulasi secara umum diartikan sebagai proses penempatan data dalam bentuk tabel dengan cara membuat tabel yang berisikan data sesuai dengan kebutuhan analisis.

Tabulasi silang (*crosstabs*) adalah metode untuk mentabulasikan beberapa variabel yang berbeda ke dalam suatu matriks. Hasil tabulasi silang disajikan ke dalam suatu tabel dengan variabel-variabel yang tersusun sebagai kolom dan baris. Tabulasi silang merupakan metode analisis yang paling sederhana tetapi memiliki daya menerangkan cukup kuat untuk menjelaskan hubungan antarvariabel.

C. Persentase Baris

Menghitung persentase baris dapat dilakukan dengan cara yang ada pada tabel di bawah ini:

Variabel	Frekuensi	%
X	n_1	$n_1/N \times 100$
Y	n_2	$n_2/N \times 100$
Jumlah	N	100%

Contoh cara menghitung persentase baris:

Jenis Kelamin	N	%
Laki-laki	20	$20/50 \times 100 = 40\%$
Perempuan	30	$30/50 \times 100 = 60\%$
Jumlah	50	100%

Berdasarkan tabel di atas dapat dilihat sebuah variabel jenis kelamin dengan jumlah laki-laki ($n_1=20$) dan perempuan ($n_2=30$) didapatkan persentase laki-laki sebanyak 40% dan persentase perempuan sebanyak 60%.

D. Persentase Kolom

Menghitung persentase kolom dapat dilakukan dengan cara yang ada pada tabel di bawah ini:

Variabel 1	X	Y	Jumlah
Frekuensi	n_1	n_2	N
%	$n_1/N \times 100$	$n_2/N \times 100$	100%

Contoh cara menghitung persentase kolom:

Jumlah Mahasiswa	Biostatistika		Dasar Gizi		Jumlah
	n	%	n	%	
Lulus	15	$15/40 \times 100 = 37,5\%$	25	$25/40 \times 100 = 62,5\%$	40 (100%)
Tidak Lulus	5	$5/10 \times 100 = 50\%$	5	$5/10 \times 100 = 50\%$	10 (100%)

Berdasarkan tabel di atas, dapat disimpulkan mahasiswa peminatan biostatistika yang lulus sebanyak 15 orang (37,5%) peminatan gizi sebanyak 25 orang (62,5%). Sedangkan mahasiswa yang tidak lulus pada peminatan biostatistika dan gizi berjumlah sama yaitu masing-masing sebanyak 5 orang (50%).

E. Persentase Total

Menghitung persentase kolom dapat dilakukan dengan cara yang ada pada tabel di bawah ini:

Variabel 1	Variabel 2		Jumlah	%
	A	B		
X	n_{AX}	n_{BX}	Nx	$Nx/N_{total} \times 100$
Y	n_{AY}	n_{BY}	Ny	$Ny/N_{total} \times 100$
Jumlah	Na	Nb	N_{total}	
%	$Na/N_{total} \times 100$	$Nb/N_{total} \times 100$		100 (100%)

Contoh cara menghitung persentase total:

Berat Bayi saat Lahir	Usia Ibu Melahirkan		Jumlah
	Usia Risiko	Usia Tidak Risiko	
BBLR	45	10	$55/100 \times 100 = 55\%$
Tidak BBLR	15	30	$45/100 \times 100 = 45\%$
Jumlah	$60/100 \times 100 = 60\%$	$40/100 \times 100 = 40\%$	100 (100%)

Berdasarkan tabel di atas, dapat ditarik kesimpulan bahwa persentase total ibu dengan usia risiko sebesar 60% dan ibu dengan usia tidak risiko 40%. Sedangkan jumlah bayi yang mengalami BBLR sebanyak 55% dan tidak BBLR 45%.

F. Rangkuman

Tabulasi silang merupakan metode untuk mentabulasikan beberapa variabel yang berbeda ke dalam suatu matriks. Hasil tabulasi silang disajikan ke dalam suatu tabel dengan variabel-variabel yang tersusun sebagai kolom dan baris. Persentase yang dihitung pada baris, kolom dan total data dapat memberikan gambaran mengenai kondisi sebuah masalah, terutama yang berkaitan dengan masalah kesehatan masyarakat.

G. Soal Latihan

Hitunglah nilai persentase baris, persentase kolom dan persentase total dari tabel di bawah ini!

Tabel. Site of infection for serogroup B meningococcal disease in region X.

Site of infection	Pre-epidemic	Early epidemic	Recent epidemic	Total
Sepsis	66	45	40	151
Meningitis	39	32	39	110
Both	39	32	34	105
Unknown	6	8	26	42
Total	150	117	141	408

BAB 5

TEORI PROBABILITAS

A. Tujuan Pembelajaran

Mahasiswa mampu memecahkan masalah menggunakan konsep probabilitas dalam pengkajian dan analisis permasalahan bidang kesehatan masyarakat.

B. Konsep Probabilitas

Segala sesuatu dalam kehidupan kita ini bersifat tidak pasti. Artinya, kita tidak bisa mengetahui bagaimana hasil akhir dari sebuah kejadian. Terlebih lagi bila kejadian tersebut berkaitan dengan masa yang akan datang.

Contoh:

1. Jika seorang ibu memiliki dominan tangan kiri maka belum tentu memiliki anak yang kidal juga.
2. Apakah bila seseorang tidak menjadi perokok aktif maka akan terhindar dari kanker paru-paru.

Ketika menghadapi ketidakpastian tersebut, maka biasanya orang akan mengandalkan kemungkinan atau peluang yang mungkin terjadi. Dari kemungkinan dan peluang tersebut maka melahirkan sebuah teori yang dikenal sebagai teori *probabilitas*.

C. Permutasi dan Kombinasi

1. Permutasi

Permutasi adalah suatu penyusunan atau pengaturan beberapa objek ke dalam suatu urutan tertentu.

Contoh: Ada 3 objek, yaitu XYZ. Pengaturan objek-objek tersebut ialah XYZ, XZY, YZX, ZXY, ZYX, YXZ yang disebut permutasi. Jadi permutasi 3 objek menghasilkan enam pengaturan dengan cara yang berbeda,

Contoh: Pengaturan 4 huruf dari 6 huruf pertama dalam abjad menghasilkan 360 cara yang berbeda.

Posisi 1	Posisi 2	Posisi 3	Posisi 4
Tersedia pilihan	Tersedia pilihan	Tersedia pilihan	Tersedia pilihan
6	5	4	3

Dari contoh tersebut, dapat diketahui bahwa banyaknya permutasi yang mungkin bergantung pada:

1. Ukuran kelompok asalnya
2. Banyaknya objek yang dipilih dari kelompok itu.

Rumus-Rumus Permutasi

1. Permutasi dari n objek tanpa pengembalian

- a. Permutasi dari n objek seluruhnya

$$nPn = n!$$

- b. Permutasi sebanyak r dari n objek

$$nPn = \frac{n!}{(n-r)!} \quad (n \geq r)$$

- c. Permutasi Melingkar

Sejumlah objek yang berbeda dapat disusun secara teratur dalam suatu lingkaran dalam $(n-1)!$ cara.

2. Permutasi dari n objek dengan pengembalian

$$nP_r = n.r$$

3. Permutasi dari n objek yang sama

$$nP_{n_1, n_2, n_3, \dots} = \frac{n!}{n_1! n_2! n_3! \dots}$$

2. Kombinasi

Kombinasi adalah suatu penyusunan beberapa objek tanpa memperhatikan urutan objek tersebut.

Persamaan kombinasi

1. Kombinasi r dari n objek yang berbeda

$$C_r^n = \frac{n!}{r! (n-r)!} \quad (n \geq r)$$

2. Hubungan permutasi dengan kombinasi

$$P_r^n = r! \ C_r^n \text{ atau } C_r^n = \frac{P_r^n}{r!}$$

D. Distribusi Binomial dan Poisson

1. Distribusi Binomial

Distribusi binomial adalah Suatu distribusi teoretis yang menggunakan variabel random diskret yang terdiri dari dua kejadian yang berkomplementer. Misalnya, kejadian sukses-gagal, ya-tidak, baik-buruk, angka-gambar dan sebagainya.

Ciri-ciri distribusi binomial adalah:

- Setiap percobaan hanya memiliki dua peristiwa seperti ya-tidak, sukses-gagal.
- Probabilitas satu peristiwa adalah tetap, tidak berubah untuk setiap percobaan.
- Percobaannya bersifat independen artinya peristiwa dari suatu percobaan tidak mempengaruhi atau dipengaruhi peristiwa dalam percobaan lainnya.
- Jumlah atau banyaknya percobaan yang merupakan komponen percobaan binomial harus tertentu (bilangan random diskret).

Rumus binomial suatu peristiwa

Probabilitas suatu peristiwa dapat dihitung dengan mengalikan kombinasi susunan dengan probabilitas salah satu susunan.

$$P(X = x) = b(x; n; p) = C_x^n \cdot p^x \cdot q^{n-x}$$

Keterangan:

x = banyaknya peristiwa sukses

n = banyaknya percobaan

p = probabilitas peristiwa sukses

q = 1-p = probabilitas peristiwa gagal

Probabilitas dari peristiwa binomial lebih dari satu sukses disebut sebagai probabilitas binomial kumulatif. Probabilitas binomial kumulatif dapat dihitung dengan persamaan berikut

$$\begin{aligned}
 P_{KB} &= \sum_{x=0}^n C_x^n \cdot p^x \cdot q^{n-x} \\
 &= \sum_{x=0}^n P(X=x) \\
 &= P(x=0) + P(x=1) + P(x=2) + \dots + P(x=n)
 \end{aligned}$$

Rata-rata, varians, dan simpangan baku pada probabilitas binomial dapat dihitung dengan persamaan berikut:

$$\begin{aligned}
 \text{rata-rata } (\mu) &= n \cdot p \\
 \text{varians } (\sigma^2) &= n \cdot p \cdot q \\
 \text{simpangan baku } (\sigma) &= \sqrt{n \cdot p \cdot q}
 \end{aligned}$$

2. Distribusi Poisson

Distribusi poisson adalah distribusi nilai-nilai bagi suatu variabel random X yaitu banyaknya hasil percobaan yang terjadi dalam suatu interval waktu tertentu atau di suatu daerah tertentu.

Ciri-ciri dan manfaat dari distribusi poisson dapat dilihat pada tabel berikut:

Ciri-Ciri	Manfaat
<ol style="list-style-type: none"> 1. Banyaknya hasil percobaan yang terjadi dalam suatu interval waktu atau suatu daerah tertentu tidak tergantung pada banyaknya hasil percobaan yang terjadi pada interval waktu atau daerah lain yang terpisah 2. Probabilitas terjadinya hasil percobaan selama suatu interval waktu yang singkat atau dalam suatu daerah kecil, sebanding dengan panjang 	<p>Menghitung probabilitas terjadinya peristiwa menurut satuan waktu, ruang, luas, panjang tertentu seperti menghitung probabilitas dari:</p> <ol style="list-style-type: none"> 1. Banyaknya telepon per menit atau banyaknya mobil yang lewat selama 5 menit di suatu ruas jalan 2. Banyaknya bakteri dalam 1 tetes per 1 L air

<p>interval waktu atau besarnya daerah tersebut dan tidak bergantung pada banyaknya hasil percobaan yang terjadi di luar interval waktu atau daerah tersebut</p> <p>3. Probabilitas lebih dari satu hasil percobaan yang terjadi dalam interval waktu yang singkat atau dalam daerah yang kecil dapat diabaikan</p>	<p>3. Banyaknya kesalahan ketik per halaman</p> <p>4. Banyaknya kecelakaan mobil di jalan tol selama seminggu</p> <p>5. Menghitung distribusi probabilitas binomial apabila nilai n besar ($n \geq 30$) dan p kecil ($p < 0,1$)</p>
---	--

Rumus Distribusi Poisson

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

Keterangan:

λ = rata-rata terjadinya suatu peristiwa

E = bilangan alam = 2,71828

E. Distribusi Normal dan T

1. Distribusi Normal

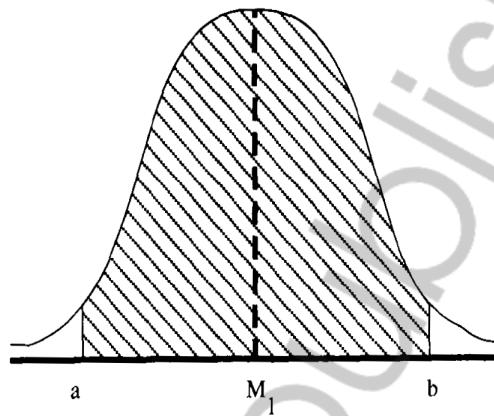
Distribusi Normal adalah model distribusi kontinyu yang paling penting dalam teori probabilitas. Distribusi Normal diterapkan dalam berbagai permasalahan. Distribusi normal memiliki kurva berbentuk lonceng yang simetris. Dua parameter yang menentukan distribusi normal adalah rataan/ekspektasi (μ) dan standar deviasi (σ).

Fungsi kerapatan probabilitas dari distribusi normal diberikan dalam rumus berikut:

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

Di mana μ adalah rata-rata, σ adalah standar deviasi dan $\pi = 3,14159\dots$

Contoh grafik fungsi kerapatan probabilitas dari distribusi normal.



Grafik fungsi distribusi normal tersebut di atas membentang dari minus tak hingga hingga tak hingga. Hanya saja, semakin jauh dengan rata-rata (M_1), nilai probabilitas akan semakin mendekati nol.

2. Distribusi T

Distribusi T atau distribusi student-t banyak diterapkan pada statistik inferensial ketika nilai varians dari suatu populasi tidak diketahui. Distribusi-t merupakan distribusi sampling yang digunakan untuk pengujian statistik dengan kondisi ukuran sampel yang relatif kecil. Dengan menggunakan distribusi-t, kita bisa melakukan pengujian statistik untuk memperkirakan parameter yang tidak diketahui (seperti standar populasi) dengan cara menggunakan distribusi sampel untuk mengubah standar deviasi populasi.

Karena sampel terdiri atas kelompok nilai yang berbeda, nilai standar deviasi juga cenderung selalu berubah dan semakin bervariasi antar satu kelompok sampel dengan sampel lainnya. Jika standar deviasi populasi diketahui, peneliti akan cenderung menggunakan distribusi normal sebagai pendekatan pengujian statistik yang lebih valid.

Distribusi-t digunakan dalam estimasi statistik sampel kecil

(small statistic estimation). Dalam hal ini, para ahli sepakat menyimpulkan bahwa ukuran sampel kecil yaitu di bawah 30 unit. Ini juga sejalan dengan Theorema Limit Pusat (Central Limit Theorem).

Formula penggunaan distribusi T terdiri dari dua jenis, yaitu:

1. Jika standar deviasi populasi diketahui:

$$t_0 = \frac{-\mu_0}{S}$$

$$t_0 = \frac{-\mu_0}{\frac{\sigma}{\sqrt{n}}}$$

2. Jika standar deviasi populasi tidak diketahui, dengan asumsi bahwa sampel yang digunakan memiliki distribusi standar normal di mana nilai rata-rata = 0 dan nilai varians = 1, maka:

$$t_0 = \frac{-\mu_0}{S}$$

$$t_0 = \frac{-\mu_0}{\frac{S}{\sqrt{n}}}$$

Keterangan:

σ = standar deviasi dari populasi

S = standar deviasi dari sampel

F. Rangkuman

Penggunaan konsep probabilitas dalam pengkajian dan analisis permasalahan bidang kesehatan masyarakat dirasa sangat penting. Beberapa distribusi probabilitas yang biasa digunakan adalah Distribusi Binomial, Distribusi Poisson, Distribusi Normal serta Distribusi T.

G. Soal Latihan

1. Sebuah survei kebersihan gigi memperlihatkan bahwa 2 dari 5 orang sudah pergi ke dokter gigi dalam beberapa bulan terakhir. Apabila ada 12 orang terpilih secara acak, hitunglah probabilitas 4 diantaranya pergi ke dokter dua bulan lalu?

2. Diketahui probabilitas untuk terjadi shock pada saat imunisasi dengan vaksinasi meningitis adalah 0,0005. Kalau di suatu kota jumlah orang yang dilakukan vaksinasi sebanyak 4000. Hitunglah peluang tepat tiga orang akan terjadi shock!
3. Dalam suatu ujian biostatistik terdapat 300 siswa yang mengikuti ujian tersebut. Rata-rata dari hasil ujian yaitu 70 dengan simpangan baku hasil ujian tersebut adalah 10. Jika data nilai hasil ujian siswa tersebut berdistribusi normal, maka berapa persen mahasiswa yang mendapat nilai A jika syarat untuk mendapatkan nilai A adalah nilai lebih dari 85?

BAB 6

CENTRAL LIMIT THEOREM

A. Tujuan Pembelajaran

Mahasiswa mampu memahami Central limit theorem (teori limit pusat); nilai standard error (galat baku), distribusi sampling nilai mean dan proporsi, sifat distribusi probabilitas normal; dan mengaplikasikan distribusi normal standar

B. Central Limit Theorem

Teorema Limit Pusat (Central Limit Theorem) menyatakan bahwa semakin besar ukuran sampel maka bentuk distribusi sampling dari rata-rata sampel selalu cenderung mendekati distribusi normal. Sehingga menyebabkan sampel tersebut memiliki karakteristik yang juga terdapat pada distribusi normal. Di mana besar kemungkinan rata-ratanya akan sangat dekat atau bahkan hampir sama dengan nilai parameternya, begitu juga dengan nilai variansnya.

C. Distribusi Sampling

Distribusi sampling adalah distribusi dari besaran-besaran statistik, seperti rata-rata, simpangan baku, proporsi atau persentase yang mungkin muncul dari sampel-sampel. Distribusi dari rata-rata sampel disebut distribusi sampling rata-rata, distribusi dari proporsi sampel disebut distribusi sampling proporsi dan sebagainya.

Contoh:

Jika besar populasi adalah 4 ($N=4$), misalkan A, B, C, D, kemudian diambil sampel berukuran 2 ($n=2$) maka diperoleh 6 sampel, yaitu AB, AC, AD, BC, BD, CD (sampelnya tanpa pengembalian). Dari keenam sampel tersebut dihitung rata-ratanya, maka didapatkan 6 rata-rata sampel. Enam rata-rata sampel tersebut membentuk suatu distribusi, disebut distribusi sampling rata-rata atau rata-rata distribusi sampel.

Demikian pula dengan perhitungan simpangan baku, distribusi varians dan distribusi proporsinya.

Jenis-Jenis Distribusi Sampling

Berdasarkan besaran statistik yang digunakan, dikenal beberapa jenis distribusi sampling, yaitu distribusi sampling rata-rata, proporsi, beda dua rata-rata dan beda dua proporsi.

1. Distribusi Sampling Rata-Rata

Distribusi sampling rata-rata adalah distribusi dari besaran rata-rata yang muncul dari sampel-sampel. Misalkan sebuah populasi berukuran 4, yang anggotanya adalah 1,2,3,4 dan sampelnya berukuran 2. Maka distribusi sampling rata-ratanya dapat dilihat pada tabel di bawah ini:

Sampel 1: 1;2 dengan rata-rata = 1,5

Sampel 2: 1;3 dengan rata-rata = 2

Sampel 3: 1;4 dengan rata-rata = 2,5

Sampel 4: 2;3 dengan rata-rata = 2,5

Sampel 5: 2;4 dengan rata-rata = 3

Sampel 6: 3;4 dengan rata-rata = 3,5

	f	Probabilitas
1,5	1	0,16
2	1	0,16
2,5	2	0,34
3	1	0,16
3,5	1	0,16
Jumlah	6	1,00

Pada distribusi sampling rata-rata berlaku hal-hal berikut:

- Pemilihan sampel dari populasi terbatas

Untuk pengambilan sampel tanpa pengembalian atau $(\frac{n}{N}) > 5\%$:

$$\mu_x = \mu$$

$$\sigma_x = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

Untuk pengambilan sampel dengan pengembalian atau $\leq 5\%$:

$$\mu_x = \mu$$

$$\sigma_x = \frac{\sigma}{\sqrt{n}}$$

- b. Pemilihan sampel dari populasi yang tidak terbatas

Apabila populasi memiliki ukuran yang tidak terbatas dan berdistribusi secara normal dengan rata-rata μ dan simpangan baku σ , maka rata-rata sampel dengan sampel berukuran n dan dipilih dengan pengembalian atau tanpa pengembalian dari populasi tersebut akan memiliki distribusi normal dengan rata-rata dan simpangan baku:

$$\mu_x = \mu$$

$$\sigma_x = \frac{\sigma}{\sqrt{n}}$$

- c. Daftar distribusi normal untuk distribusi sampling rata-rata

Daftar distribusi normal untuk distribusi sampling rata-rata, dapat digunakan rumus:

$$Z = \frac{-\mu}{\sigma_x}$$

Untuk populasi terbatas atau $(\frac{n}{N}) > 5\%$:

$$Z = \frac{-\mu}{\sigma_x} \text{ atau } Z = \frac{-\mu}{\frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}}$$

Untuk populasi tidak terbatas atau $(\frac{n}{N}) \leq 5\%$:

$$Z = \frac{-\mu}{\sigma_x} \text{ atau } Z = \frac{-\mu}{\frac{\sigma}{\sqrt{n}}}$$

2. Distribusi Sampling Proporsi

Distribusi sampling proporsi adalah distribusi dari proporsi yang diperoleh dari semua sampel sama besar yang mungkin dari

satu populasi. Distribusi sampling proporsi biasanya digunakan untuk mengetahui persentase atau perbandingan antara dua hal yang saling berkomplementer, seperti peristiwa perokok dan bukan perokok, kejadian BBLR dan tidak BBLR, Perilaku berisiko dan tidak berisiko dan sebagainya.

Pada distribusi sampling proporsi, berlaku hal-hal berikut:

Untuk pengambilan sampel dengan pengembalian atau jika $(\frac{n}{N}) \leq 5\%$, memiliki rata-rata dan simpangan baku:

$$\mu_p = P$$

$$\sigma_p = \sqrt{\frac{P(1-P)}{n}} = \sqrt{\frac{PQ}{n}}$$

Keterangan:

P = proporsi kejadian sukses

Q = proporsi kejadian gagal (1-P)

Untuk pengambilan sampel tanpa pengembalian atau jika $(\frac{n}{N}) > 5\%$ memiliki rata-rata dan simpangan baku:

$$\mu_p = P$$

$$\begin{aligned}\sigma_p &= \sqrt{\frac{P(1-P)}{n}} = \sqrt{\frac{N-n}{N-1}} \\ &= \sqrt{\frac{PQ}{n}} \sqrt{\frac{N-n}{N-1}}\end{aligned}$$

Daftar distribusi normal untuk distribusi sampling proporsi dapat ditentukan sebagai berikut:

Jika n besar maka nilai Z adalah,

$$Z = \frac{p - P}{\sigma_p}$$

Jika n sangat kecil maka nilai Z adalah,

$$Z = \frac{p \pm \frac{1}{2n} - P}{\sigma_p}$$

3. Distribusi sampling yang lain

Distribusi sampling yang lain yaitu distribusi sampling beda dua rata-rata dan distribusi sampling beda dua proporsi.

D. Distribusi Probabilitas Normal

Distribusi probabilitas normal disebut pula distribusi gauss yang merupakan probabilitas yang paling banyak digunakan dalam berbagai analisis statistika.

Beberapa karakteristik dari distribusi probabilitas dan kurva normal adalah:

1. Kurva berbentuk genta atau lonceng dan memiliki satu puncak yang terletak di tengah. Nilai rata-rata hitung (μ) = median (Md) = modus (Mo). Nilai $\mu = Md = Mo$ yang berada di tengah membelah kurva menjadi dua bagian yaitu setengah di bawah nilai $\mu = Md = Mo$ dan setengah di atas nilai $\mu = Md = Mo$.
2. Distribusi probabilitas dan kurva normal berbentuk kurva simetris dengan rata-rata hitungnya (μ).
3. Distribusi probabilitas dan kurva normal bersifat asimptotis.
4. Kurva mencapai puncak pada saat $X = \mu$.
5. Luas daerah di bawah kurva normal adalah 1; $\frac{1}{2}$ di sisi kanan nilai tengah dan $\frac{1}{2}$ di sisi kiri.

E. Distribusi Normal Baku (Standar)

Distribusi normal baku adalah distribusi probabilitas acak normal dengan nilai tengah nol dan simpangan baku 1. Seringkali disebut dengan distribusi Z . Hal yang perlu dilakukan pada distribusi probabilitas normal standar adalah mengubah atau membakukan distribusi aktual dalam bentuk distribusi normal baku, dikenal dengan nilai Z atau skor Z .

Nilai Z adalah jarak yang berbeda antara sebuah nilai X yang dipilih dari rata-rata μ , dibagi dengan standar deviasinya, σ .

Rumus nilai Z adalah:

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

Z = Skor Z atau nilai normal baku

X = Nilai dari suatu pengamatan atau pengukuran

μ = Nilai rata-rata hitung suatu distribusi

σ = Standar deviasi

F. Rangkuman

Teorema Limit Pusat merupakan sebuah teori yang menyatakan bahwa jika ukuran sampel semakin besar, maka sifat dari rata-rata distribusi peluang sampelnya (*sample mean distribution*) akan semakin mendekati distribusi normal. Dalam beberapa teori statistika, dikatakan bahwa jumlah sampel minimal yang digunakan untuk teorema ini adalah 30 observasi.

G. Soal Latihan

1. Berat badan mahasiswa suatu perguruan tinggi mempunyai distribusi normal dengan rata-rata (μ) 60 dan deviasi standar (σ) 10. Tentukan nilai variabel normal standar bagi mahasiswa yang memiliki besar badan 70 dan 50?

BAB 7

TEORI ESTIMASI

A. Tujuan Pembelajaran

Mahasiswa mampu menganalisis permasalahan kesehatan masyarakat dengan estimasi (perkiraan) interval kepercayaan dari nilai rerata populasi dan estimasi (perkiraan) interval kepercayaan dari nilai proporsi populasi.

B. Estimasi

Estimasi adalah proses menggunakan sampel statistik untuk menduga atau menaksir hubungan parameter populasi yang tidak diketahui. Estimasi merupakan suatu pernyataan mengenai parameter populasi yang diketahui berdasarkan informasi dari sampel (sampel random) yang diambil dari populasi. Jadi dengan estimasi tersebut keadaan parameter populasi dapat diketahui.

C. Estimasi Interval Kepercayaan dari Nilai Rerata Populasi

Estimasi interval untuk rata-rata ditentukan sebagai berikut.

1. Untuk Sampel Besar ($n > 30$)

- Untuk populasi tidak terbatas atau populasi terbatas yang pengambilan sampelnya dengan pengembalian dan σ diketahui

Untuk populasi yang tidak terbatas atau dari populasi terbatas yang pengambilan sampelnya dengan pengembalian serta nilai σ diketahui, pendugaan interval untuk rata-rata dirumuskan:

$$-z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < +z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

- Untuk populasi terbatas, pengambilan sampel tanpa pengembalian, dan σ diketahui

Untuk populasi yang terbatas, di mana pengambilan sampelnya tanpa pengembalian dan nilai σ diketahui atau $(\frac{n}{N}) > 5\%$ pendugaan interval untuk rata-rata dirumuskan:

$$-z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} < \mu < +z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

2. Untuk Sampel kecil ($n \leq 30$)

Untuk sampel kecil di mana pengambilan sampelnya dengan pengembalian dan σ tidak diketahui, pendugaan interval untuk rata-rata dirumuskan:

$$-t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} < \mu < +t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

di mana

$$s = \sqrt{\frac{\sum X^2}{n-1} - \frac{(\sum X)^2}{n(n-1)}}$$

D. Estimasi Interval Kepercayaan dari Nilai Proporsi Populasi

Pendugaan interval untuk proporsi ditentukan sebagai berikut:

1. Untuk Sampel Besar ($n > 30$)

a. Untuk populasi tidak terbatas

Untuk populasi yang tidak terbatas, pendugaan interval untuk proporsi dirumuskan sebagai berikut:

$$-z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}} < P < +z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}}$$

b. Untuk populasi terbatas dan pengambilan sampel tanpa pengembalian

Untuk populasi yang terbatas serta pengambilan sampelnya tanpa pengembalian atau $(\frac{n}{N}) > 5\%$, maka untuk menduga interval proporsinya dirumuskan:

$$-z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}} \sqrt{\frac{N-n}{N-1}} < P < +z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}} \sqrt{\frac{N-n}{N-1}}$$

2. Untuk Sampel kecil ($n \leq 30$)

Untuk sampel kecil estimasi interval untuk proporsi dirumuskan:

$$- t_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}} < P < + t_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}}$$

E. Rangkuman

Menganalisis permasalahan kesehatan masyarakat dengan estimasi interval kepercayaan dari nilai rerata populasi dan estimasi interval kepercayaan dari nilai proporsi populasi akan dapat menunjukkan keadaan parameter populasi yang sebenarnya.

F. Soal Latihan

Penelitian terhadap sampel 25 orang penderita kelainan darah tertentu sebuah rumah sakit, 3 diantaranya mengidap kanker darah. Dengan interval keyakinan 95%, tentukan proporsi orang yang mengidap kanker darah!

BAB 8

UJI HIPOTESIS

A. Tujuan Pembelajaran

Setelah mempelajari Bab ini, mahasiswa mampu menelaah (memilih, menginterpretasikan) uji hipotesis, menghitung serta menginterpretasikan uji beda proporsi.

B. Hipotesis

1. Pengertian Hipotesis

Hipotesis adalah pernyataan keyakinan yang digunakan dalam evaluasi nilai populasi. Hypo artinya sementara/lemah kebenarannya dan thesis artinya pernyataan/teori. Dengan demikian hipotesis berarti pernyataan sementara yang perlu diuji kebenarannya. Untuk menguji kebenaran sebuah hipotesis digunakan pengujian yang disebut pengujian hipotesis. Di dalam pengujian hipotesis dijumpai dua jenis hipotesis, yaitu hipotesis nol (H_0) dan hipotesis alternatif (H_a).

1) Hipotesis Null (H_0)

Hipotesis yang menyatakan tidak ada perbedaan suatu kejadian antara kedua kelompok (kejadian pada kedua kelompok adalah sama) atau hipotesis yang menyatakan tidak ada hubungan antara variabel satu dengan variabel yang lain.

Contoh:

Tidak ada perbedaan berat badan bayi antara mereka yang dilahirkan dari ibu yang mendapatkan antenatal terpadu dan komprehensif dengan *antenatal care* tidak terpadu dan komprehensif (kurang dari 6 kali).

Tidak ada hubungan *antenatal care* dengan berat badan bayi yang dilahirkan.

2) Hipotesis Alternatif (H_a)

Hipotesis yang menyatakan ada perbedaan kejadian antara kedua kelompok atau hipotesis yang menyatakan ada hubungan variabel satu dengan variabel yang lain.

Contoh:

Ada perbedaan berat badan bayi antara mereka yang dilahirkan dari ibu yang mendapatkan antenatal terpadu dan komprehensif dengan berat badan bayi yang dilahirkan dari ibu dengan *antenatal care* tidak terpadu dan komprehensif.

Ada hubungan *antenatal care* dengan berat badan bayi yang dilahirkan.

2. Arah Uji Hipotesis

Bentuk pernyataan hipotesis alternatif akan menentukan arah uji statistik, apakah satu arah (*one tail*) atau dua arah (*two tail*).

- 1) One tail (satu arah/sisi): bila hipotesis alternatifnya menyatakan adanya arah, hal yang satu *lebih tinggi atau lebih rendah* dari hal yang lain. ($A > B$ atau $B < A$)

Contoh:

Rata-rata berat badan bayi dari ibu hamil yang mendapatkan antenatal terpadu dan komprehensif **lebih tinggi** dibandingkan rata-rata berat badan bayi dari ibu dengan *antenatal care* tidak terpadu dan komprehensif.

- 2) Two Tail (dua sisi) merupakan hipotesis alternatif yang hanya menyatakan adanya perbedaan, **tanpa** melihat apakah hal yang satu lebih tinggi atau lebih rendah dari hal yang lain. ($C \neq D$)

Contoh:

Berat badan bayi dari ibu hamil dengan *antenatal care* terpadu dan komprehensif **berbeda** dibandingkan berat badan bayi dari ibu dengan antenatal yang tidak terpadu dan komprehensif.

Dengan kata lain: *Ada perbedaan* berat badan bayi antara mereka yang dilahirkan dari ibu dengan *antenatal care* terpadu dan komprehensif dari ibu dengan *antenatal care* tidak terpadu dan komprehensif.

Hipotesis ini biasanya dipakai untuk masalah-masalah yang belum banyak diteliti, belum jelas pengaruhnya.

Contoh Penulisan Hipotesis:

Suatu penelitian ingin mengetahui hubungan antara jenis kelamin dengan tekanan darah, maka hipotesisnya adalah sbb:

$$H_0: \mu_A = \mu_B$$

Tidak ada perbedaan mean tekanan darah antara laki-laki dan perempuan, atau Tidak ada hubungan antara jenis kelamin dengan tekanan darah

$$H_a: \mu_A \neq \mu_B$$

Ada perbedaan mean tekanan darah antara laki-laki dan perempuan, atau Ada hubungan antara jenis kelamin dengan tekanan darah.

C. Area Penolakan

Wilayah penolakan (juga disebut wilayah kritis) adalah area grafik di mana kita akan menolak hipotesis nol jika hasil penghitungan termasuk dalam area tersebut. Dengan kata lain, jika hasil penghitungan termasuk dalam area tersebut, maka hasilnya signifikan secara statistik.

Misal seorang peneliti, memilih tingkat alfa yang diterima, misalnya, jika ingin 95% yakin bahwa hasil kita signifikan, kita akan memilih tingkat alfa 5% (100% - 95%). Level 5% itu adalah wilayah penolakan. Untuk tes satu ekor, 5% akan berada dalam satu ekor. Untuk uji dua sisi, daerah penolakan akan berada dalam dua sisi, sehingga masing-masing sisi 2.5%.

Satu sisi (ekor kiri)	Dua sisi	Satu sisi (ekor kanan)
$H_0: \mu_X = \mu_0$ $H_1: \mu_X < \mu_0$	$H_0: \mu_X = \mu_0$ $H_1: \mu_X \neq \mu_0$	$H_0: \mu_X = \mu_0$ $H_1: \mu_X > \mu_0$
		

D. Kesalahan Pengambilan Keputusan

Dalam pengujian hipotesis kita selalu dihadapkan suatu kesalahan pengambilan keputusan.

Ada dua jenis kesalahan pengambilan keputusan dalam uji statistik, yaitu:

1. Kesalahan Tipe I (α)

Merupakan kesalahan menolak H_0 padahal sesungguhnya H_0 benar. Artinya: menyimpulkan adanya perbedaan padahal sesungguhnya tidak ada perbedaan. Peluang kesalahan tipe satu (I) adalah α atau sering disebut Tingkat signifikansi (*significance level*). Sebaliknya peluang untuk tidak membuat kesalahan tipe I adalah sebesar $1-\alpha$, yang disebut dengan Tingkat Kepercayaan (*confidence level*).

2. Kesalahan Tipe II (β)

Merupakan kesalahan tidak menolak H_0 padahal sesungguhnya H_0 salah. Artinya: menyimpulkan tidak ada perbedaan padahal sesungguhnya ada perbedaan. Peluang untuk membuat kesalahan tipe kedua (II) ini adalah sebesar β . Peluang untuk tidak membuat kesalahan tipe kedua (II) adalah sebesar $1-\beta$, dan dikenal sebagai Tingkat Kekuatan Uji (*power of the test*).

Tabel 8. 1. Kesalahan Yang Mungkin Terjadi Pada Uji Hipotesis

Sampel	Populasi	
	H_0 Salah	H_0 Benar
H_0 ditolak (H_0 Salah)	$1-\beta$	α (type-I error)
H_0 gagal ditolak (H_0 Benar)	β (type-II error)	$1-\alpha$
Jumlah	1.0	1.0

E. Langkah Uji

1. Nyatakan H_0 dan H_a

H_0 : Tidak ada perbedaan A dengan B *atau* $A = B$

H_a : Ada perbedaan A dengan B *atau* $A \neq B$ (2-tailed)

H_a : A lebih besar dari B *atau* $A > B$ (1-tailed)

H_a : P lebih kecil dari Q *atau* $P < Q$ (1-tailed)

H_a merupakan *hypothesis penelitian yang diyakini oleh peneliti*

2. Tentukan derajat kemaknaan (α): 0.01, 0.05, 0.1

0.01 → Berhubungan dengan obat + nyawa manusia

0.05 → Berhubungan dengan kesehatan/kesakitan

0.10 → Berhubungan dengan sosial budaya

3. Tentukan jenis uji statistik yang cocok:

a. Uji beda proporsi, beda rata-rata, uji-korelasi?

b. Tentukan area penolakan H_0 dg melihat nilai-_{tabel} pada α tertentu

4. Hitung nilai-uji-statistik

5. Hitung nilai p/p-value atau bandingkan nilai-statistik dengan nilai tabel.

Nilai-p adalah probabilitas bahwa hipotesis nol itu benar. P adalah singkatan dari probabilitas dan mengukur seberapa besar kemungkinan bahwa setiap perbedaan yang diamati antara kelompok disebabkan oleh kebetulan. Nilai yang mendekati 0 menunjukkan bahwa perbedaan yang diamati tidak mungkin disebabkan oleh kebetulan, sedangkan nilai P yang mendekati 1 menunjukkan tidak ada perbedaan antara kelompok selain karena kebetulan.

6. Keputusan:

Jika $|$ nilai uji-_{statistik} $| \geq$ nilai-_{tabel} atau $p\text{-value} \leq \alpha$, maka H_0 ditolak
Jika $|$ nilai uji-_{statistik} $| <$ nilai-_{tabel} atau $p\text{-value} > \alpha$, maka H_0 gagal ditolak

Uji 2-sisi (2-tailed) menggunakan tanda nilai mutlak $||$

Uji 1-sisi (1-tailed) tanpa menggunakan tanda nilai mutlak

7. Kesimpulan:
 - jika H_0 ditolak \rightarrow simpulkan H_a
 - jika H_0 gagal ditolak \rightarrow simpulkan H_0

F. Penentuan Uji Statistik

Penentuan jenis uji statistik yang cocok tergantung pada:

1. Skala pengukuran: Nominal Ordinal Interval Rasio.

2. Tujuan uji statistik:

- a. Uji 1-sampel
- b. Uji 2-sampel
- c. Uji k-sampel
- d. Uji korelasi atau hubungan

3. Asumsi uji statistik:

- Parametrik:
 - Sampel random
 - Distribusi normal
 - Varians sama
- Non-parametrik (tidak perlu asumsi)

G. Menguji Beda Mean 1 Sampel

Tujuan pengujian adalah untuk mengetahui perbedaan mean populasi dengan mean data sampel penelitian. Oleh karena tujuan pengujian ini adalah membandingkan data satu sampel dengan satu populasinya.

1. Uji Beda Mean Satu Sampel

Berdasarkan ada ~~tidaknya~~ standar deviasi di populasi (σ) maka jenis uji beda mean satu sampel dibagi 2 jenis:

Berdasarkan ada ~~tidaknya~~ standar deviasi di populasi (σ) maka jenis uji beda mean satu sampel dibagi 2 jenis:

Jika nilai σ diketahui, digunakan uji z, rumusnya:

$$Z = \frac{x - \mu}{\sigma / \sqrt{n}}$$

Bila nilai σ tidak diketahui, digunakan uji t, rumusnya:

$$t = \frac{x - \mu}{Sd / \sqrt{n}}$$

Ket: $df = n - 1$

X = rata-rata data sampel

μ = rata-rata data populasi

σ = standar deviasi pada data di populasi

Sd = standar deviasi pada data sampel

n = jumlah sampel yang diteliti

Contoh:

Dalam penelitian tentang penyakit jantung di Provinsi X diperoleh rata-rata umur 53 dengan standar deviasi 5.5, kemudian pada pengujian terhadap 100 sampel diperoleh rata-rata umur 54.85. Peneliti ingin menguji apakah ada perbedaan rata-rata umur pada sampel dan populasi.

Penyelesaian:

1. Hipotesis

$H_0: \mu = 53$

Tidak ada perbedaan rata-rata umur rata-rata penderita jantung pada studi di Provinsi X dengan rata-rata penderita jantung pada populasi

$H_a: \mu \neq 53$

Ada perbedaan rata-rata umur rata-rata penderita jantung pada studi di Provinsi X dengan rata-rata penderita jantung pada populasi.

Jika dilihat dari hipotesis alternatif yang hanya ingin mengetahui perbedaan, maka jenis statistik yang digunakan adalah *two tail* (dua arah).

2. Level of Significance

Digunakan alpha $\alpha = 0,05$ (5%)

3. Pemilihan uji statistik

Jenis uji yang digunakan adalah uji Z, karena ingin membandingkan nilai populasi & data sampel dengan σ diketahui.

4. Penghitungan

$$Z = \frac{x - \mu}{\sigma / \sqrt{n}} = \frac{54.85 - 53}{5.5 / \sqrt{100}} = \frac{1.85}{0.55} = 3.36$$

5. Keputusan

Selesai menghitung selanjutnya kita ingin mengetahui apakah keputusan uji apakah H_0 ditolak atau gagal ditolak.

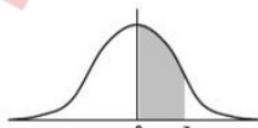
Dari nilai $Z=3,36$ diperoleh peluang $0,4996$ (tabel Z), berarti nilai p -nya $= 0,5 - 0,4996 = 0,004$.

Karena arah uji ini adalah *two tail* maka nilai $p = 2 \times 0,004 = 0,008$. Jadi nilai **$p=0,008$** .

Dengan melihat hasil p dan membandingkannya dengan $\alpha = 0,05$, maka **terlihat nilai p lebih kecil dari nilai α , sehingga kita memutuskan hipotesis nol (H_0) ditolak.**

Dengan demikian dapat disimpulkan pada $\alpha 5\%$ bahwa secara statistik **ada perbedaan rata-rata** umur rata-rata penderita jantung pada studi di Provinsi X dengan rata-rata penderita jantung pada populasi.

Kumulatif sebaran frekuensi normal
(Area di bawah kurva normal baku dari 0 sampai z)



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319

1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997

Jika peneliti tidak mengetahui besarnya standar deviasi populasi serta hanya mengambil populasi sebanyak 25 sampel penderita jantung. Standar deviasi sampel didapatkan sebesar 2,3 tahun. Peneliti ini ingin menguji apakah rata-rata umur rata-rata penderita jantung pada studi di Provinsi X berbeda dengan rata-rata penderita jantung pada populasi.

Penyelesaian

1. Hipotesis

$$H_0: \mu = 53$$

$$H_a: \mu \neq 53$$

Jika dilihat dari hipotesis alternatif yang hanya ingin mengetahui perbedaan, maka jenis statistik yang digunakan adalah *two tail* (dua arah).

2. Level of Significance

Digunakan alpha $\alpha = 0,05$ (5%)

3. Pemilihan uji statistik

Jenis uji yang digunakan adalah uji t, karena ingin membandingkan nilai populasi & data sampel dengan σ tidak diketahui, yang diketahui adalah Sd.

4. Penghitungan

$$t = \frac{x - \mu}{Sd / \sqrt{n}} = \frac{54.85 - 53}{2.3 / \sqrt{25}} = \frac{1.85}{0.46} = 4.02$$

$$Df = 25 - 1 = 24$$

5. Keputusan

Selesai menghitung selanjutnya kita ingin mengetahui apakah keputusan uji apakah H_0 ditolak atau gagal ditolak.

dk	α untuk Uji Satu Pihak (<i>one tail test</i>)					
	0,25	0,10	0,05	0,025	0,01	0,005
	α untuk Uji Dua Pihak (<i>two tail test</i>)					
0,50	0,20	0,10	0,05	0,02	0,01	
1	1,000	3,078	6,314	12,706	31,821	63,657
2	0,816	1,886	2,920	4,303	6,965	9,925
3	0,765	1,638	2,353	3,182	4,541	5,841
4	0,741	1,533	2,132	2,776	3,747	4,604
5	0,727	1,476	2,015	2,571	3,365	4,032
6	0,718	1,440	1,943	2,447	3,143	3,707
7	0,711	1,415	1,895	2,365	2,998	3,499
8	0,706	1,397	1,860	2,306	2,896	3,355
9	0,703	1,383	1,833	2,262	2,821	3,250
10	0,700	1,372	1,812	2,228	2,764	3,169
11	0,697	1,363	1,796	2,201	2,718	3,106
12	0,695	1,356	1,782	2,179	2,681	3,055
13	0,692	1,350	1,771	2,160	2,650	3,012
14	0,691	1,345	1,761	2,145	2,624	2,977
15	0,690	1,341	1,753	2,131	2,602	2,947
16	0,689	1,337	1,746	2,120	2,583	2,921
17	0,688	1,333	1,740	2,110	2,567	2,898
18	0,688	1,330	1,734	2,101	2,552	2,878
19	0,687	1,328	1,729	2,093	2,539	2,861
20	0,687	1,325	1,725	2,086	2,528	2,845
21	0,686	1,323	1,721	2,080	2,518	2,831
22	0,686	1,321	1,717	2,074	2,508	2,819
23	0,685	1,319	1,714	2,069	2,500	2,807
24	0,685	1,318	1,711	2,064	2,492	2,797

25	0,684	1,316	1,708	2,060	2,485	2,787
26	0,684	1,315	1,706	2,056	2,479	2,779
27	0,684	1,314	1,703	2,052	2,473	2,771
28	0,683	1,313	1,701	2,048	2,467	2,763
29	0,683	1,311	1,699	2,045	2,462	2,756
30	0,683	1,310	1,697	2,042	2,457	2,750
40	0,681	1,303	1,684	2,021	2,423	2,704
60	0,679	1,296	1,671	2,000	2,390	2,660
120	0,677	1,289	1,658	1,980	2,358	2,617
∞	0,674	1,282	1,645	1,960	2,326	2,576

Dari nilai df 24 diperoleh nilai tabel sebesar 2.064 (two tail 0.05). Dengan melihat hasil hitung dan membandingkannya dengan nilai tabel, maka **terlihat nilai hitung (4.02) lebih besar dari nilai tabel (2,064)**, sehingga kita memutuskan hipotesis nol (H₀) ditolak.

Dengan demikian dapat disimpulkan pada α 5% bahwa secara statistik **ada perbedaan rata-rata** umur rata-rata penderita jantung pada studi di Provinsi X dengan rata-rata penderita jantung pada populasi.

Apabila ingin mencari nilai p/p value menggunakan tabel distribusi t, maka cara mencarinya adalah sebagai berikut:

	.10	.05	.025	.01	.005
1			
2			
..					
24	1,318	1,711	2,064	2,492	2,797

↑
t=4,02

Nilai p yang diperoleh adalah lebih kecil dari 0.005.

Dengan melihat nilai p dan membandingkannya dengan $\alpha = 0,05$, maka **terlihat nilai p lebih kecil dari nilai α , sehingga kita memutuskan hipotesis nol (H₀) ditolak.**

Menguji Beda Proporsi

Tujuannya adalah untuk mengetahui/menguji perbedaan proporsi populasi dengan proporsi data sampel penelitian.

Hipotesis:

$$H_0: p = P_0$$

$$H_0: p = P_0$$

$$H_a: p \neq P_0$$

$$H_a: p > P_0 \text{ atau } p < P_0$$

two tail

one tail

$$Z = \frac{p - P_0}{\sqrt{(P_0.Q)/n}}$$

p = proporsi data sampel penelitian

P_0 = proporsi data populasi

$$Q = 1 - P$$

Contoh:

Data Dinas Kependudukan Kab.X tahun 2021 menyebutkan bahwa 40% perempuan tidak menggunakan kontrasepsi. Kepala dinas ingin membuktikan apakah sekarang penggunaan kontrasepsi masih tetap seperti tahun 2021 atau sudah berubah.

Untuk pengujian ini diambil sampel acak sebanyak 25 perempuan dan dilakukan wawancara pada ibu baru setahun terakhir melakukan persalinan, dan ternyata terdapat 45% yang mengaku tidak menggunakan kontrasepsi. Ujilah apakah ada perbedaan proporsi tidak digunakannya kontrasepsi antara Laporan Dinas Kependudukan dengan sampel penelitian, dengan alpha 5%.

1. Hipotesis

$$H_0: p = 0.4$$

Tidak ada perbedaan proporsi tidak digunakannya kontrasepsi antara data dinas dengan data sampel

$$H_a: \mu \neq 0.4$$

Ada perbedaan proporsi tidak digunakannya kontrasepsi antara data dinas dengan data sampel.

Jika dilihat dari hipotesis alternatif yang hanya ingin mengetahui perbedaan, maka jenis statistik yang digunakan adalah *two tail*

(dua arah).

2. *Level of Significance*

Digunakan alpha $\alpha = 0,05$ (5%)

3. Pemilihan uji statistik

Jenis uji yang digunakan adalah uji Z uji beda proporsi, karena ingin membandingkan nilai proporsi populasi & data.

4. Penghitungan

$$Z = \frac{p - P}{\sqrt{(PQ)/n}} = \frac{0.45 - 0.40}{\sqrt{(0.40 \times 0.60)/25}} = \frac{0.05}{\sqrt{(0.40 \times 0.60)/25}} = 2,55$$

5. Keputusan

Selesai menghitung selanjutnya kita ingin mengetahui apakah keputusan uji apakah H_0 ditolak atau gagal ditolak.

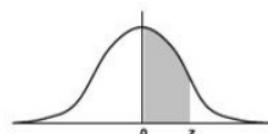
Dari nilai $Z=2,55$ diperoleh peluang 0,4946 (tabel Z), berarti nilai p -nya $= 0,5 - 0,4946 = 0,0054$.

Karena arah uji ini adalah *two tail* maka nilai $p = 2 \times 0,0054 = 0,0108$. Jadi nilai $p=0,0108$.

Dengan melihat hasil p dan membandingkannya dengan $\alpha = 0,05$, maka **terlihat nilai p lebih kecil dari nilai α , sehingga kita memutuskan hipotesis nol (H_0) ditolak.**

Dengan demikian dapat disimpulkan pada $\alpha 5\%$ bahwa secara statistik **ada perbedaan proporsi tidak digunakannya kontrasepsi antara data dinas dengan data sampel.**

Kumulatif sebaran frekuensi normal
(Area di bawah kurva normal baku dari 0 sampai z)



Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389

1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997

H. Soal Latihan

1. Untuk setiap pernyataan ini, buatlah hipotesis Null dan Hipotesis Alternatif:
 - a. Apakah rata-rata umur terjadinya penyakit akut pada anak sekolah berbeda dari 11.5 tahun?
 - b. Apakah IQ anak sekolah di Sekolah X lebih tinggi dari IQ anak sekolah pada umumnya yaitu diatas 100?
 - c. Apakah rata-rata kadar hemoglobin dari kelompok pekerja berbeda dari 16 g/cc?
2. Standar kadar serum kolesterol untuk laki-laki adalah 200 mg/100 ml dengan standar deviasi 16.67 mg. Pada sampel 49 laki-laki kelebihan berat badan, diperoleh rata-rata 211 mg. Apakah kadar serum kolesterol pada laki-laki kelebihan berat badan lebih tinggi dibandingkan laki-laki pada umumnya, dengan $\alpha = 0.05$.

BAB 9

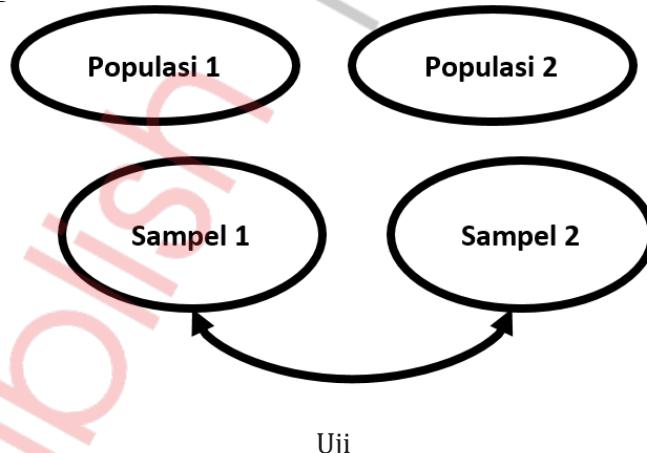
UJI BEDA DUA RERATA

A. Tujuan Pembelajaran

Mahasiswa mampu menganalisis permasalahan kesehatan masyarakat berdasarkan uji beda dua rerata.

B. Uji Beda Dua Rerata Independen

Pengambil kebijakan maupun penelitian dalam bidang kesehatan seringkali ingin melihat apakah dua populasi memiliki karakteristik yang sama atau berbeda satu sama lain dan jika berbeda, perbedaannya bukanlah suatu kebetulan (*by chance*). Untuk menguji hal tersebut sampel diambil dari dua populasi yang berbeda dengan asumsi mereka mewakili nilai yang ada di masing-masing populasi untuk selanjutnya nilai kedua sampel tersebut dibandingkan. Ilustrasi gambarnya adalah sebagai berikut:



Contoh:

- Ingin diketahui apakah ada perbedaan rata-rata tinggi badan baduta antara ibu dengan pengetahuan gizi baik dan pengetahuan gizi kurang.

- Ingin diketahui apakah ada perbedaan rata-rata hb ibu sebelum dan sesudah pemberian tablet Fe.

Sebelum dilakukan pengujian statistik dua kelompok data, perlu diperhatikan apakah dua kelompok data tersebut berasal dari kelompok yang sama/independen atau berasal dari dua kelompok yang sama/dependen/berpasangan. Kelompok data independen adalah bila data kelompok yang satu tidak tergantung pada data kelompok kedua, misalnya membandingkan rata-rata tinggi badan baduta pada kelompok ibu pengetahuan gizi baik dan rata-rata berat badan bayi pada kelompok ibu pengetahuan gizi kurang, sedangkan kelompok dikatakan dependen apabila dua kelompok data yang dibandingkan datanya saling mempunyai ketergantungan, misalnya data hb ibu sebelum dan sesudah pemberian tablet Fe. Kita harus mencocokkan dan memastikan setiap data ibu yang diambil sebelum pemberian tablet Fe sama dengan data ibu sesudah pemberian tablet Fe (subjek yang sama), jika tidak ada data yang bisa dibandingkan.

Alasan pentingnya untuk mengenali karakteristik data adalah karena terdapat dua uji beda 2 rata-rata/mean yang berbeda. Uji beda dua mean dibagi dalam dua kelompok, yaitu: uji beda dua mean independen/ *independen sample t-test* (untuk 2 sampel yang independen) dan uji beda mean dependen/ *paired t-test* (untuk 2 sampel dependen/data sampel sebelum dan sesudah).

Asumsi berikut harus dipenuhi untuk uji beda 2 mean:

1. Data pada kedua sampel berdistribusi normal
2. Kedua kelompok data independen
3. Variabel yang dihubungkan adalah kategorik dan numerik.

Prosedur untuk menguji perbedaan 2 sampel:

1. Buat Hipotesis nol dan Hipotesis alternatif
2. Pilih level signifikansi, (biasanya $\alpha=0.05$ atau $\alpha=0.01$)
3. Hitung melalui rumus statistik
4. Tentukan area kritis (*critical region*)
5. Ho akan ditolak jika nilai t berdasarkan uji statistik berada pada area kritis (*tail*) dan Ho akan ditolak jika nilai t berdasarkan uji

statistik berada pada area kritis (*tail*) dan H_0 gagal ditolak jika nilai t berdasarkan uji statistik berada diluar pada area kritis (*tail*).

6. Hitung nilai p (p value)
7. Bandingkan nilai p dengan nilai α
8. Hitung *confidence interval* (95% atau 99%)
9. Buat kesimpulan

Prinsip pengujinya adalah melihat perbedaan variasi kedua kelompok data. Oleh karena itu dalam pengujian ini diperlukan informasi apakah varian kedua kelompok yang diuji sama atau tidak. Bentuk varian ini akan berpengaruh pada standar error yang akhirnya akan membedakan rumus pengujinya. Untuk itu perlu dilakukan pengujian varians dengan uji F. Tujuan uji ini adalah untuk mengetahui varian antara kelompok data satu, apakah sama dengan kelompok data yang kedua.

Adapun rumus untuk uji Varians (Uji F):

$$F = \frac{S_1^2}{S_2^2}$$

Dengan:

Varian yang lebih besar sebagai pembilang (numerator)

Varian yang lebih kecil sebagai penyebut (denominator)

Jika p value < a, H_0 gagal ditolak : Varian sama

Jika p value > a, H_0 ditolak : Varian berbeda

$df_1 = n_1 - 1$ dan $df_2 = n_2 - 1$

Uji untuk Varian Sama

Menggunakan uji Z atau uji t. Uji Z digunakan apabila standar populasi deviasi populasi (σ) diketahui dan jumlah sampel besar (>30). Apabila dua syarat tersebut tidak dipenuhi maka dilakukan uji t. Pada umumnya σ sulit diketahui sehingga digunakan uji t.

$$t = \frac{x_1 - x_2}{Sp \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}}}$$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}$$

dengan:

x_1 atau x_2 = rata-rata sampel klp 1 atau 2

n_1 atau n_2 = jumlah sampel klp 1 atau 2

S_1 atau S_2 = standar deviasi sampel klp 1 atau 2

df = degree of freedom (derajat kebebasan)

$df = n_1 + n_2 - 2$

Uji untuk Varian Berbeda

$$t = \frac{x_1 - x_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Dengan degree of freedom:

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

C. Uji Beda Dua Rerata Dependen

Bertujuan untuk menguji perbedaan mean antara dua kelompok data yang dependen. Contoh kasus:

- Apakah ada perbedaan rata-rata berat badan bayi sebelum dan sesudah pemberian ASI ekslusif.
- Apakah ada perbedaan berat badan balita sebelum dan sesudah pemberian makanan tambahan.

Syarat:

1. Distribusi data normal
2. Kedua kelompok data dependen/pair

3. Jenis variabel yang dihubungkan numerik dan kategori

Formula uji T dependen:

$$t = \frac{d}{Sd_d/\sqrt{n}}$$

$$df = n - 1$$

d = rata-rata deviasi/selisih sampel 1 dengan sampel 2

Sd_d = Standar deviasi dari deviasi/selisih sampel 1 dan sampel 2

Contoh Data:

50 ibu/responden diambil untuk menjadi sampel, diperoleh data sebagai berikut:

Tabel 9. 1. Distribusi data Ibu di Kabupaten X Tahun 2022

No_resp	Umur	BB_Bayi	ANC	Hb 1	Hb 2	Pendidikan	Sosek
1	19	3600	0	9.8	14.6	3	3
2	18	3300	1	9.6	12.4	3	3
3	17	2800	1	9.1	12.5	3	2
4	19	2500	0	9.8	14.7	1	2
5	19	2400	0	9.3	13.4	1	2
6	19	2100	1	9.4	13.5	1	1
7	20	2300	1	9.4	13.5	1	1
8	20	2600	0	9.6	14.1	2	3
9	20	2700	0	9.5	13.5	2	3
10	20	2300	0	9.5	13.6	2	2
11	21	2000	0	9.5	13.7	1	1
12	21	3000	0	9.9	14.8	1	2
13	21	2200	1	9.6	14.2	1	1
14	22	2800	0	9.7	14.3	2	1
15	22	2900	0	9.7	14.3	2	2
16	22	3000	0	9.8	14.6	2	2
17	22	2900	1	9.7	15.5	3	3

No_resp	Umur	BB_Bayi	ANC	Hb 1	Hb 2	Pendidikan	Sosek
18	22	2900	0	10.1	15.5	3	3
19	23	2600	1	9.4	13.4	3	1
20	23	2400	1	9.2	12.7	3	2
21	23	3400	1	9.3	13.3	3	2
22	23	3100	1	9.5	13.7	1	2
23	23	2600	1	9.5	13.8	1	2
24	24	3500	0	9.6	14.2	4	3
25	24	3000	0	9.7	14.5	4	2
26	24	3100	0	9.8	10.2	4	2
27	24	3200	0	9.8	10.2	4	2
28	25	2700	1	9.2	9.1	3	2
29	25	2800	0	9.2	9.1	3	3
30	26	2200	1	11.2	10	1	1
31	26	2900	0	11.2	10	1	3
32	26	2500	1	11.2	10	1	3
33	26	2600	1	13.2	12.3	3	3
34	26	3400	0	13.2	12.3	3	3
35	27	2700	1	11.2	10	4	3
36	27	2500	1	11.2	10	4	1
37	27	2900	0	11.2	10	4	2
38	28	3300	0	11.2	10	3	3
39	28	3700	0	11.2	10	3	2
40	29	3500	0	11.2	10	3	2
41	29	3000	0	13.2	12.3	4	2
42	30	2800	0	13.2	12.3	4	3
43	31	2800	0	13.2	12.3	4	3
44	32	2400	1	13.2	12.3	4	2
45	32	3300	0	13.2	12.3	4	3
46	32	3200	0	13.2	12.3	4	3
47	34	3100	0	11.1	11.5	2	2
48	34	2700	1	11.1	11.5	2	2

No_resp	Umur	BB_Bayi	ANC	Hb 1	Hb 2	Pendidikan	Sosek
49	35	3100	0	10.2	9.8	3	2
50	35	3200	0	10.2	9.8	3	3

Contoh 1:

Dari data diatas ingin diketahui apakah ada perbedaan rata-rata berat badan bayi lahir antara ibu dengan pemeriksaan kehamilan (anc) yang baik dan tidak.

Penyelesaian:

Diketahui anc baik

$$X_1 = 3135.71$$

$$Sd_1 = 257.069$$

$$n_1 = 28$$

anc kurang

$$X_2 = 2486.364$$

$$Sd_2 = 233.596$$

$$n_2 = 22$$

Prosedur untuk menguji perbedaan 2 sampel Independen:

1. Hipotesis nol dan Hipotesis alternatif

Ho: Tidak ada perbedaan rata-rata berat badan bayi lahir antara ibu dengan pemeriksaan kehamilan (anc) yang baik dan tidak.

Ha: Ada perbedaan rata-rata berat badan bayi lahir antara ibu dengan pemeriksaan kehamilan (anc) yang baik dan tidak.

2. *Two Tail* (dua arah), Level signifikansi $\alpha=0.05$

3. Uji Homogenitas Varians

Ho: Varian berat badan bayi dengan ibu dengan anc baik sama dengan varian berat badan ibu dengan anc kurang

Ha: Varian berat badan bayi dengan ibu dengan anc baik berbeda dengan varian berat badan ibu dengan anc kurang

$$F = \frac{S_1^2}{S_2^2} = \frac{257,0694^2}{233,596^2} = 1,21$$

$$df_1 = 28-1 = 27 \text{ (Numerator)}$$

$$df_2 = 22-1 = 21 \text{ (Denominator)}$$

Tabel F:

Denominator	.. 24	∞	P value
20	0,100	1,61	$F=1,21$
	0,050	1,84	
	0,025	2,09	
	0,010	2,42	
	0,005	2,69	
	0,001	3,35	

- Hasilnya dari tabel terlihat bahwa nilai $F=1,21$ berada di atas nilai 1,61 karena tabel memiliki pola semakin ke atas nilainya semakin kecil sehingga diperoleh p value lebih besar dari 0,10 (karena untuk pola p value semakin ke atas semakin besar).
- Perbandingan p value dan α : p value $>\alpha$ (lebih besar dari 0,10 $> 0,05$) sehingga keputusannya H_0 gagal ditolak atau **varians sama**.
- Uji beda dua mean dengan varian sama:

$$\begin{aligned}
 7. \quad t &= \frac{x_1 - x_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{3135,71 - 2486,36}{\sqrt{\frac{247,07}{28+22}}} = \frac{649,35}{\sqrt{70,39}} = 9,225 \\
 s_p^2 &= \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)} \\
 &= \frac{(28 - 1)257,069^2 + (22 - 1)233,596^2}{(28 - 1) + (22 - 1)} = 61,045,617
 \end{aligned}$$

$$Sp = 247,07$$

- Diperoleh nilai t hitung sebesar 9,225
- $df = n_1 + n_2 - 2 = 28 + 22 - 2 = 48$
- Hitung nilai p (p value) dengan tabel t

df \ α	0.10	0.05	0.025	0.01	0.005	P value??
40	1,303	1,684	2,021	2,423	2,704	
60						t=9,225

Nilai $t = 9,225$ dalam tabel berada di sebelah kanan dari 2,704 karena terlihat pola tabel nilai t semakin ke kanan nilainya semakin besar, sehingga dari tabel diatas terlihat bahwa p value berada pada posisi lebih kecil dari 0,005 (memiliki pola semakin ke kanan semakin kecil nilainya).

10. Bandingkan nilai p dengan nilai α

$0,05 < 0,005$, H_0 ditolak sehingga kesimpulannya ada perbedaan rata-rata berat badan bayi lahir antara ibu dengan pemeriksaan kehamilan (anc) yang baik dan tidak.

Contoh 2:

Dari data yang tersedia ingin diketahui apakah ada perbedaan rata-rata hb pada ibu sebelum dan sesudah pemberian tablet Fe

Jawaban

Prosedur untuk menguji perbedaan 2 sampel Dependent:

1. Hipotesis nol dan Hipotesis alternatif

H_0 : Tidak ada perbedaan rata-rata kadar hb ibu sebelum dan sesudah pemberian tablet Fe

H_a : Ada perbedaan rata-rata kadar hb ibu sebelum dan sesudah pemberian tablet Fe.

2. Level signifikansi $\alpha=0.05$

3. Hitung dengan rumus:

$$t = \frac{d}{Sd_d / \sqrt{n}}$$

No_resp	Hb 1	Hb 2	deviasi	No_resp	Hb 1	Hb 2	deviasi
1	9.8	14.6	-4.8	26	9.1	12.6	-3.5
2	9.6	12.4	-2.8	27	9.2	12.8	-3.6
3	9.1	12.5	-3.4	28	9.4	12.2	-2.8
4	9.8	14.7	-4.9	29	9.4	12.3	-2.9
5	9.3	13.4	-4.1	30	9.4	12.9	-3.5
6	9.4	13.5	-4.1	31	9.6	14.2	-4.6
7	9.4	13.5	-4.1	32	9.8	14.7	-4.9
8	9.6	14.1	-4.5	33	9.9	15.2	-5.3
9	9.5	13.5	-4	34	9.6	15.6	-6
10	9.5	13.6	-4.1	35	9.5	13.8	-4.3
11	9.5	13.7	-4.2	36	9.5	13.9	-4.4
12	9.9	14.8	-4.9	37	9.5	13.7	-4.2
13	9.6	14.2	-4.6	38	9.2	13.1	-3.9
14	9.7	14.3	-4.6	39	9.6	14.2	-4.6
15	9.7	14.3	-4.6	40	9.7	14.2	-4.5
16	9.8	14.6	-4.8	41	9.5	15.5	-6
17	9.7	15.5	-5.8	42	10	15.4	-5.4
18	10.1	15.5	-5.4	43	10	15.3	-5.3
19	9.4	13.4	-4	44	9	12.4	-3.4
20	9.2	12.7	-3.5	45	9.7	15.2	-5.5
21	9.3	13.3	-4	46	9.9	15.1	-5.2
22	9.5	13.7	-4.2	47	9.6	14.1	-4.5
23	9.5	13.8	-4.3	48	9.6	14.1	-4.5
24	9.6	14.2	-4.6	49	9.3	13.2	-3.9
25	9.7	14.5	-4.8	50	9.3	13.2	-3.9

$$df = n - 1 = 50 - 1 = 49$$

d = rata-rata deviasi/selisih sampel 1 dengan sampel 2 = -4,39

Sd_d = standar deviasi dari deviasi/selisih sampel 1 dan sampel 2 = 0,759

$$t = \frac{d}{Sd_d/\sqrt{n}} = \frac{-4,39}{0,759/\sqrt{50}} = -40.9$$

4. Hitung nilai p (p value) dengan tabel t

df \ α	0.10	0.05	0.025	0.01	0.005	P value??
40	1,303	1,684	2,021	2,423	2,704	
60						t=9,225

Dari tabel diatas terlihat bahwa nilai 40.9 berada di sebelah kanan dari 2,704 karena terlihat pola tabel nilai t semakin ke kanan nilainya semakin besar, sehingga dari tabel diatas terlihat bahwa p value berada pada posisi lebih kecil dari 0,005 (memiliki pola semakin ke kanan semakin kecil nilainya).

5. Bandingkan nilai p dengan nilai α

$0,05 < 0,005$, H_0 ditolak sehingga kesimpulannya ada perbedaan rata-rata kadar hb ibu sebelum dan sesudah pemberian tablet Fe.

D. Rangkuman

Dalam menyelesaikan masalah hipotesis dua sampel secara komprehensif, pertama-tama perlu ditentukan apakah sampel tersebut merupakan uji independen atau dependen.

E. Soal Latihan

1. Perhatikan data tekanan darah diastolik berikut:

Kelompok	Rerata	Standar Deviasi	n
Vegetarian	72.9	11.7	40
Non-Vegetarian	73.5	11.4	43
Laki-Laki	74.9	12.0	38
Perempuan	71.8	11.0	45

- Apakah terdapat perbedaan yang signifikan antara rerata tekanan darah diastolik antara:
- Vegetarian dan non-vegetarian
 - Laki-laki dan perempuan
2. Sepuluh percobaan terhadap hewan percobaan dilakukan untuk melihat denyut nadi per menit. Sebelum dan sesudah intervensi. Berikut hasilnya:

Hewan	Sebelum	Sesudah
1	70	115
2	84	128
3	88	146
4	110	171
5	105	158
6	120	115
7	110	110
8	67	140
9	79	131
10	86	157

Apakah data tersebut dapat membuktikan bahwa kondisi eksperimen mampu meningkatkan denyut jantung intervensi per menit (gunakan $\alpha = 0.05$).

BAB 10

UJI BEDA LEBIH DARI DUA RERATA

A. Tujuan Pembelajaran

Mahasiswa mampu menelaah (menghitung, menyajikan, dan menginterpretasikan, serta membandingkan) hasil uji Analisis Varians; mampu menguraikan prosedur Analisis Varians; mampu menjelaskan istilah-istilah umum yang terkait dengan prosedur Analisis Varians; serta mampu melakukan uji hipotesis Analisis Varians.

B. Analisis Varians

Uji Anova atau Analisis Varians (ANOVA) dikembangkan untuk memungkinkan peneliti untuk menguji hipotesis perbandingan lebih dari dua kelompok. Prinsip uji Anova adalah kita membandingkan variansi tiga kelompok sampel atau lebih. Lebih dari sekadar membandingkan nilai mean (rata-rata), uji anova juga mempertimbangkan keragaman data yang ditunjukkan dalam nilai varians.

Prinsip uji ANOVA adalah melakukan telaah variabilitas data menjadi dua sumber yaitu variasi dalam kelompok (*within*) dan variasi antar kelompok (*between*). Bila variasi *within* dan *between* sama (nilai perbandingan kedua varian sama dengan 1) maka *mean-mean* yang dibandingkan tidak ada perbedaan. Sebaliknya bila hasil perbandingan kedua varian tersebut menghasilkan lebih dari 1, maka mean yang dibandingkan menunjukkan ada perbedaan.

ANOVA terdiri dari dua jenis yaitu analisis varian satu faktor (*one way*) ~~dan~~ analisis varian dua faktor (*two way*). Apa saja asumsi yang harus dipenuhi dalam uji Anova sebagai bentuk dari model linier, berikut diantaranya:

1. Varian homogen
2. Sampel/kelompok independen
3. Data berdistribusi normal
4. Jenis data yang dihubungkan adalah numerik dengan Kategorik.

C. Prosedur Analisis Varians

Berikut adalah prosedur dalam melakukan analisis varians:

1. Mengumpulkan sampel dan mengelompokkan sampel berdasarkan karakteristik tertentu

Untuk memudahkan pengelompokan dan perhitungan, buatlah tabel data sesuai dengan kategori berisi sampel dan kuadrat dari sampel tersebut. Hitung juga total dari sampel dan kuadrat sampel dari masing-masing kelompok. Selain itu, tentukan pula hipotesis nol (H_0) dan hipotesis alternatif (H_1).

2. Menentukan tipe anova

Apakah termasuk tipe satu arah, tipe dua arah tanpa interaksi atau tipe dua arah dengan interaksi.

3. Menghitung variabilitas dari seluruh sampel

Pengukuran total variabilitas atas data dapat dikelompokkan menjadi tiga bagian:

- a. Total of sum squares (SSt) – jumlah kuadrat total (jkt).

Merupakan jumlah kuadrat selisih antara skor individual dengan rata-rata totalnya.

- b. Sum Square Between (SSb) – jumlah kuadrat kolom (jkk).

Variansi rata-rata kelompok sampel terhadap rata-rata keseluruhannya. Variansi di sini lebih terpengaruh karena adanya perbedaan perlakuan antar kelompok.

- c. Sum Square within (SSw) – jumlah kuadrat galat (jkg).

Variansi yang ada dalam masing-masing kelompok. Banyaknya variansi akan tergantung pada banyaknya kelompok, dan variansi di sini tidak terpengaruh/tergantung oleh perbedaan perlakuan antar kelompok.

4. Menghitung derajat kebebasan (degree of freedom)

Derasat kebebasan atau degree of freedom (dilambangkan dengan v , dof , atau df) dalam ANOVA akan sebanyak variabilitas. Oleh karena itu, ada tiga macam derajat kebebasan yang akan kita hitung:

- a. Derajat kebebasan untuk JKT

Merupakan derajat kebebasan dari Jumlah kuadrat total (JKT) ini akan kita lambangkan dengan dof JKT

- b. Derajat kebebasan untuk JKK

Merupakan derajat kebebasan dari Jumlah kuadrat kolom (JKK) ini akan kita lambangkan dengan dof JKK

- c. Derajat kebebasan untuk JKG

Merupakan derajat kebebasan dari Jumlah kuadrat galat (JKG) ini akan kita lambangkan dengan dof JKG

Derajat kebebasan juga memiliki sifat hubungan yang sama dengan sifat hubungan variabel, yakni:

$$\text{dof JKT} = \text{dof JKK} + \text{dof JKG}$$

5. Menghitung varians antar kelompok dan varians dalam kelompok

Variance dalam ANOVA, baik untuk antar kelompok maupun dalam kelompok sering disebut dengan kuadrat tengah atau deviasi rata-rata kuadrat (mean squared deviation) dan dilambangkan dengan MS atau KT. Dengan demikian, maka mean squared deviation masing-masing dapat dicari dengan rumus sebagai berikut:

$$KTK = JKK/\text{dof jkk}$$

$$KTG = JKG/\text{dof jkg}$$

6. Menghitung nilai F hitung

Menghitung nilai distribusi F (Fhitung) berdasarkan perbandingan variance antar kelompok dan *variance* dalam kelompok. Fhitung didapatkan dengan rumus di bawah ini:

$$F\text{hitung} = KTK/KTG$$

7. Menentukan nilai F tabel

Menentukan nilai F berdasarkan tabel (Ftabel) juga dihitung berdasarkan nilai derajat kebebasan (langkah ke-4) menggunakan tabel distribusi-F.

8. Membandingkan nilai F hitung dengan F tabel

Jika Fhitung > Ftabel: tolak H0

Jika $F_{hitung} \leq F_{tabel}$: terima H_0

9. Menarik kesimpulan

Menarik kesimpulan berdasarkan hasil perbandingan nilai F_{hitung} dan F_{tabel} , apakah terdapat perbedaan rata-rata antar kelompok atau tidak terdapat perbedaan rata-rata antar kelompok.

Sumber Keragaman (SK)	Jumlah Kuadrat (JK)	Derajat Bebas (db)	Kuadrat Tengah (KT)	F_{hitung}
Kolom (K)	JKK	db JKK	$KTK = JKK/db JKK$	$F_{hitung} = KTK/KTG$
Galat (G)	JKG	db JKG	$KTG = JKG/db JKG$	
Total (T)	JKT	db JKT		

D. Uji Hipotesis Analisis Varians

Dalam analisis varians hanya satu hipotesis yang digunakan yaitu hipotesis dua arah (*two tail*), artinya hipotesis ini hanya melihat apakah ada perbedaan rata-rata antara kelompok. Apabila ingin diketahui berapa perbedaan masing-masing kelompok dapat dilakukan uji lanjutan setelahnya. Berikut hipotesis dalam Anova:

$H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_n$, Tidak ada perbedaan yang nyata antara rata-rata hitung dari n kelompok

$H_1: \mu_1 \neq \mu_2 \neq \mu_3 \neq \dots \neq \mu_n$, Ada perbedaan yang nyata antara rata-rata hitung dari n kelompok

E. Rangkuman

Anova merupakan singkatan dari “analysis of variance”. *Analysis of Variance* adalah salah satu uji komparatif yang digunakan untuk menguji perbedaan mean (rata-rata) data lebih dari dua kelompok. Untuk melakukan uji Anova, harus dipenuhi beberapa asumsi, yaitu: Sampel berasal dari kelompok yang independen, varian antar kelompok harus homogen dan data masing-masing kelompok berdistribusi normal.

Prinsip Uji Anova adalah melakukan analisis variabilitas data menjadi dua sumber variasi yaitu variasi di dalam kelompok (*within*) dan variasi antar kelompok (*between*).

Bila variasi *within* dan *between* sama (nilai perbandingan kedua varian mendekati angka satu), maka berarti tidak ada perbedaan efek dari intervensi yang dilakukan, dengan kata lain nilai mean yang dibandingkan tidak ada perbedaan.

Sebaliknya bila variasi antar kelompok lebih besar dari variasi di dalam kelompok, artinya intervensi tersebut memberikan efek yang berbeda, dengan kata lain nilai mean yang dibandingkan menunjukkan adanya perbedaan.

F. Soal Latihan

Pada suatu laboratorium gizi, empat jenis susu formula yaitu A, B, C dan D yang sudah siap diminum bayi dicoba disimpan di tempat terbuka pada kondisi yang sama. Ingin diketahui apakah kualitas keempat jenis susu formula tersebut sama atau tidak. Lamanya (dalam jam) sampai susu formula tersebut basi, dapat dilihat pada tabel di bawah ini.

Jenis A	Jenis B	Jenis C	Jenis D
3	8	9	4
5	8	7	5
4	7	6	3
3	9	6	3
4	5	8	4
4	10	9	2
3	7	6	3

Ujilah dengan tingkat kepercayaan 95% ada tidaknya perbedaan kekuatan antara susu formula yang satu dengan susu formula yang lain!

BAGIAN 11

UJI KORELASI DAN REGRESI

1. Tujuan pembelajaran

Setelah mengikuti pembelajaran ini mahasiswa diharapkan untuk mampu mahasiswa mampu menelaah (menghitung, menyajikan, dan menginterpretasikan, serta membandingkan uji korelasi dan regresi linier, mampu menjelaskan perbedaan *Pearson's Product Moment Correlation Coefficient* dengan *Spearman's Rank Order Correlation Coefficient*.

2. Pendahuluan

Penelitian sosial, khususnya seperti ilmu Kesehatan masyarakat, Pendidikan, farmasi dan lainnya, banyak mencari korelasi antar variable. Uji korelasi atau hubungan merupakan bagian dari penelitian analitik yang minimal melibatkan dua variable yang berbeda. Korelasi banyak digunakan sebagai alat uji hipotesis dalam penelitian non-eksperimen. Kekuatan korelasi dieksperesikan melalui sebuah angka yang disebut "nilai koefisien" dan dinotasikan dengan huruf "r". Korelasi pearson merupakan salah satu jenis korelasi yang dikenalkan pertama kali oleh Galton pada tahun 1886 dan kemudian diresmikan oleh Karl Pearson tahun 1896. Ide utama dibalik koefisien korelasi adalah menghitung indeks yang merefleksikan seberapa banyak dua rangkaian pengukuran yang terkait satu sama lain. (Abdi et al., 2009)

Analisis korelasi digunakan untuk mengukur hubungan antara dua variable kontinu (misalnya hubungan variable bebas dengan variable terikat). Dalam analisis korelasi, koefisien korelasi sampel dilambangkan dengan r yang memiliki kisaran nilai -1 dan +1 yang mengkuantifikasi arah dan kekuatan hubungan linier antara dua variabel. Korelasi antara dua variable bisa positif ataupun negatif.

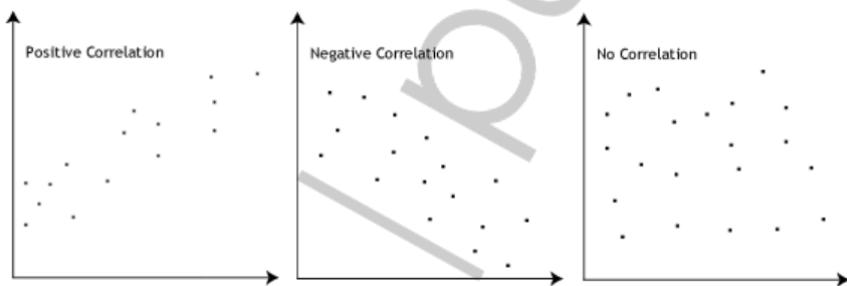
Interpretasi koefisien korelasi adalah sebagai berikut

- Jika nilai korelasi lebih besar dari 0, maka variable tersebut dikatakan berkorelasi (+) positif. Dua variable yakni x dan

y berkorelasi positif jika x meningkat, maka y cenderung mengikuti yakni mengalami peningkatan juga, sedangkan jika x menurun, y cenderung menurun

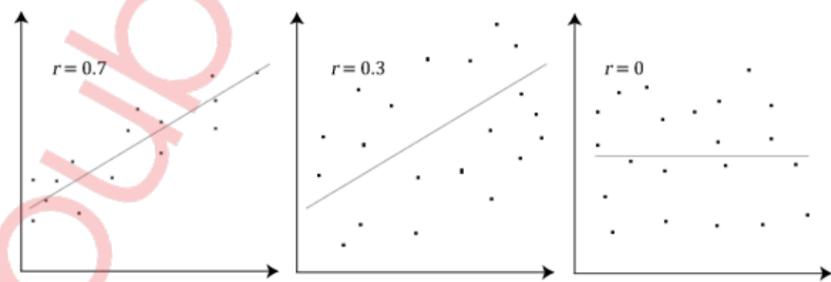
- b. Jika korelasi kurang dari 0, maka variable dikatakan berkorelasi negatif. Dua variable yakni x dan y memiliki korelasi negatif jika x meningkat, y justru cenderung menurun, sedangkan Ketika x menurun, y cenderung meningkat
- c. Jika korelasi tepat memiliki nilai 0, maka variable dikatakan tidak berkorelasi. Tidak ada hubungan linier antara x dan y

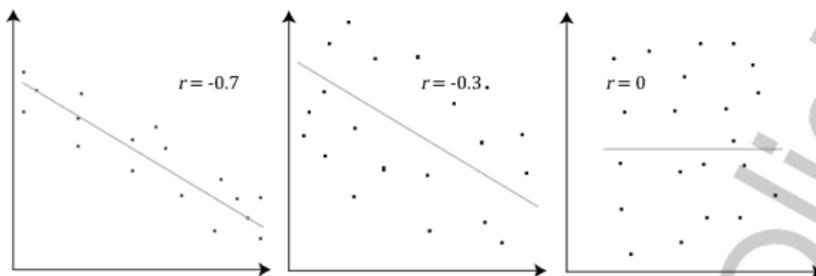
Perhatikan visualisasi berikut ini



Gambar Diagram Arah Hubungan Korelasi.

Hubungan positif jika nilai yang ditunjukkan lebih besar dari nol (0). Sedangkan nilai kurang dari nol menunjukkan hubungan negatif. Dan nilai nol menunjukkan bahwa tidak ada hubungan antara kedua variabel.





Gambar ... Diagram yang menunjukkan hubungan yang berbeda antar koefisien korelasi.

Semakin kuat hubungan kedua variable, semakin dekat koefisien korelasi, yakni menjadi +1 dan -1 tergantung apakah hubungan tersebut memiliki arah yang positif atau negatif.

3. Pearson product moment correlation

Korelasi antara variable x dan y dilambangkan dengan r, adalah

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

Di mana

r = koefisien korelasi pearson

n = jumlah data atau kasus

X = skor variabel independen

Y = skor variabel dependen

XY = skor hasil dari kedua variabel

Estimasi titik r dari korelasi pertama kali diusulkan oleh seorang ilmuwan Bernama Karl Pearson pada tahun 1896, empat tahun sebelum beliau mengembangkan chi-kuadrat uji independensi untuk tabel kontingensi. Sehingga seringkali disebut korelasi Pearson.

Sifat atau karakteristik dari korelasi adalah sebagai berikut

- Dalam mengukur kekuatan asosiasi linier, korelasi akan valid jika model garis lurus untuk hubungan x dan y. karena nilai r sebanding dengan kemiringan prediksi persamaan linier

- b. Nilai korelasi berada antara -1 dan +1
- c. Semakin besar nilai absolut r, semakin kuat hubungan liniernya. Variable dengan korelasi 0.80 lebih kuat secara linier daripada variable dengan korelasi 0.40.
- d. Nilai r tidak bergantung pada unit variable

Berikut ini ditampilkan contoh kasus untuk korelasi Pearson.

- a. Contoh kasus

Studi sederhana dilakukan dengan melibatkan 17 bayi untuk menyelidiki hubungan antara usia kehamilan saat lahir, diukur dalam satuan minggu dan berat lahir diukur dalam gram. Data ditunjukkan pada tabel berikut

Tabelusia kehamilan dan berat badan lahir

No identitas bayi	Usia kehamilan (minggu)	Berat badan lahir (gram)
1	34.7	1895
2	36	2030
3	29.3	1440
4	40.1	2835
5	35.7	3090
6	42.4	3827
7	40.3	3260
8	37.3	2690
9	40.9	3285
10	38.3	2920
11	38.5	3430
12	41.4	3657
13	39.7	3685
14	39.7	3345
15	41.1	3260

No identitas bayi	Usia kehamilan (minggu)	Berat badan lahir (gram)
16	38	2680
17	38.7	2005

Studi hendak memperkirakan hubungan antara usia kehamilan dan berat lahir. Dalam contoh kasus ini, berat lahir merupakan variable dependen (y) dan usia kehamilan adalah variable independen (x). Teori mengatakan bahwa ada hubungan positif antara usia kehamilan dan berat lahir. Bayi dengan masa usia kehamilan yang lebih pendek berisiko dilahirkan dengan berat badan yang lebih rendah, dan bayi dengan usia kehamilan yang lebih lama lebih mungkin untuk dilahirkan dengan bobot yang lebih tinggi.

- b. Formula korelasi

Rumus untuk mencari koefisien korelasi adalah

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

- c. Hipotesis null dan hipotesis alternatif

$H_0: \rho = 0$: koefisien korelasi populasi sama dengan nol

$H_1: \rho \neq 0$: koefisien korelasi populasi tidak sama dengan nol

- d. Penyelesaian

Dalam tahapan ini, proses penyelesaian perhitungan menggunakan bantuan Microsoft Excel. Data tersebut disajikan sebagai berikut:

no	x	y	xy	x^2	y^2
1	34.7	1895	65756.5	1204.09	3591025
2	36	2030	73080	1296	4120900
3	29.3	1440	42192	858.49	2073600

3	29.3	1440	42192	858.49	2073600
4	40.1	2835	113683.5	1608.01	8037225
5	35.7	3090	110313	1274.49	9548100
6	42.4	3827	162264.8	1797.76	14645929
7	40.3	3260	131378	1624.09	10627600
8	37.3	2690	100337	1391.29	7236100
9	40.9	3285	134356.5	1672.81	10791225
10	38.3	2920	111836	1466.89	8526400
11	38.5	3430	132055	1482.25	11764900
12	41.4	3657	151399.8	1713.96	13373649
13	39.7	3685	146294.5	1576.09	13579225
14	39.7	3345	132796.5	1576.09	11189025
15	41.1	3260	133986	1689.21	10627600
16	38	2680	101840	1444	7182400
17	38.7	2005	77593.5	1497.69	4020025
total (Σ)	652,1	49334	1921163	25173,21	150934928

Gambar ... Proses Perhitungan menggunakan bantuan MS Excel

Dari hasil perhitungan pada gambar di atas, maka dapat disajikan dalam rumus r berikut

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

$$r = \frac{1921163 - \frac{652,1 \times 49334}{17}}{\sqrt{(25173,21) - \frac{652,1^2}{17}} \times \sqrt{(150934928) - \frac{49334^2}{17}}} = \frac{28768,4}{35189,906}$$

$$= 0,817519$$

Nilai r hitung mencapai 0,817 sedangkan dengan menggunakan tingkat kesalahan 5% dan besar sampel ($n = 17$) adalah 0,482

Tabel nilai kritis untuk r Pearson Product Moment								
dk=n-2	Probabilitas 1 ekor							
	0,10	0,05	0,025	0,01	0,005	0,0025	0,001	0,0005
	0,20	0,10	0,05	0,02	0,01	0,01	0,002	0,001
1	0,951	0,988	0,997	1,000	1,000	1,000	1,000	1,000
2	0,800	0,900	0,950	0,980	0,990	0,995	0,998	0,999
3	0,687	0,805	0,878	0,934	0,959	0,974	0,986	0,991
4	0,608	0,729	0,811	0,882	0,917	0,942	0,963	0,974
5	0,551	0,669	0,754	0,833	0,875	0,906	0,935	0,951
6	0,507	0,621	0,707	0,789	0,834	0,870	0,905	0,925
7	0,472	0,582	0,666	0,750	0,798	0,836	0,875	0,898
8	0,443	0,549	0,632	0,715	0,765	0,805	0,847	0,872
9	0,419	0,521	0,602	0,685	0,735	0,776	0,820	0,847
10	0,398	0,497	0,576	0,658	0,708	0,750	0,795	0,823
11	0,380	0,476	0,553	0,634	0,684	0,726	0,772	0,801
12	0,365	0,458	0,532	0,612	0,661	0,703	0,750	0,780
13	0,351	0,441	0,514	0,592	0,641	0,683	0,730	0,760
14	0,338	0,426	0,497	0,574	0,623	0,664	0,711	0,742
15	0,327	0,412	0,482	0,558	0,606	0,647	0,694	0,725
16	0,317	0,400	0,468	0,543	0,590	0,631	0,678	0,708
17	0,308	0,389	0,456	0,529	0,575	0,616	0,662	0,693
18	0,299	0,378	0,444	0,516	0,561	0,602	0,648	0,679
19	0,291	0,369	0,433	0,503	0,549	0,589	0,635	0,665
20	0,284	0,360	0,423	0,492	0,537	0,576	0,622	0,652

Gambar ... Tabel Kritis untuk r Pearson Product Moment

Koefisien korelasi sampel bernilai positif dan kuat yakni mendekati angka 1. Sekali lagi diperjelas bahwa kisaran korelasi sampel adalah diantara -1 dan +1

4. Spearman's rank order correlation coefficient

Korelasi *rank* dari Spearman terkadang juga dikenal sebagai Spearman's rho yang digunakan untuk menjelaskan hubungan antara dua variable ordinal (atau satu variable ordinal dan satu numerik). Perhitungan korelasi *rank* spearman dilambangkan dengan melibatkan pengurutan peringkat nilai-nilai pada masing-masing karakteristik dari terendah ke tertinggi.

Formula untuk menghitung nilai r_s adalah

$$r_s = 1 - \frac{6 \sum d^2}{N(N^2 - 1)}$$

Berikut contoh kasus yang akan diselesaikan dengan uji korelasi *rank spearman*. Tabel berikut menyajikan data yang dikutip dari Stephens (2009)

Jumlah tornado (x)	Jumlah kematian (y)	Ranking variable x	Ranking variable y
35	0	1	2
85	0	4	2
90	2	6	5
94	1	7	4
50	0	2	2
76	3	3	6
98	4	8	7.5
104	4	9	7.5
88	5	5	9
128	7	10	10

Pada contoh kasus di atas, ingin diketahui apakah ada hubungan jumlah tornado dengan jumlah kematian yang terjadi di sebuah lokasi. Berikut akan disajikan bagaimana penyelesaian menggunakan bantuan Ms Excel.

Jumlah tornado (x)	Jumlah kematia n (y)	Ranking variable x	Ranking variable y	rank difference	d^2
35	0	1	2	-1	1
85	0	4	2	2	4
90	2	6	5	1	1
94	1	7	4	3	9
50	0	2	2	0	0

76	3	3	6	-3	9
98	4	8	7.5	0.5	0.25
104	4	9	7.5	1.5	2.25
88	5	5	9	-4	16
128	7	10	10	0	0
			Σ		42.5

Dengan menggunakan rumus berikut

$$r_s = 1 - \frac{6 \sum d^2}{N(N^2 - 1)} = 1 - \frac{6 \times 42,5}{10(10^2 - 1)} = 0,742424$$

H_0 akan ditolak jika nilai r hitung lebih besar dari nilai r tabel. Berikut merupakan tabel kritis untuk koefisien korelasi rank spearman

Table I.1 Critical Values for Spearman's Rho (ρ)

	One-tailed level of significance (p)			
	.05	.025	.01	.005
	Two-tailed level of significance (p)			
n ↓	.10	.05	.02	.01
4	1.000	—	—	—
5	.900	1.000	1.000	—
6	.829	.886	.943	1.000
7	.714	.786	.893	.929
8	.643	.738	.833	.881
9	.600	.700	.783	.833
10	.564	.648	.745	.794
11	.536	.618	.709	.755

Dengan $n = 10$ dan tingkat kesalahan 5% untuk uji dua arah maka nilai r tabel adalah 0,648. Karena .

Maka ditolak. Yang artinya ada korelasi jumlah tornado yang terjadi dengan angka kematian pada penduduk dilokasi x.

5. Analisis Regresi Linier Sederhana

Analisis regresi merupakan Teknik untuk menilai hubungan antara satu variable hasil (variable respons atau variable dependen) atau lebih dengan variable pembaur (disebut predictor, penjelas atau variable independen). Istilah variable independen dan dependen dianggap kurang sesuai pada interpretasi karena istilah ini tidak menyiratkan dengan kuat hubungan sebab akibat. Dalam analisis regresi, variable respons dilambangkan dengan y dan variable penjelas dinotasikan dengan x.

Analisis regresi sangat umum dan diterapkan secara luas. Misalkan ingin dinilai hubungan antara kolesterol total dan indeks massa tubuh (BMI). Kolesterol Total adalah variable dependen dan BMI adalah variable independen. Analisis ini dapat diperluas dengan mempertimbangkan predictor potensial seperti usia, jenis kelamin dan status merokok. Ketika terdapat satu variable dependen kontinu dan satu variable bebas, analisis yang diperlukan disebut analisis regresi linier sederhana. Analisis ini mengasumsikan bahwa ada hubungan linier antara kedua variable. Data BMI dan kolesterol total dalam sampel $n = 20$ peserta. Rerata BMI adalah $27,4 \text{ kg/m}^2$ standar deviasi adalah 3,7. Rata-rata kadar kolesterol total adalah $205,9 \text{ mg/dL}$ dengan standar deviasi 30,8.

Berdasarkan perhitungan dengan korelasi, diketahui bahwa terdapat hubungan positif antara BMI dengan kolesterol total. Peserta dengan BMI lebih rendah lebih mungkin untuk memiliki kadar kolesterol total yang lebih rendah, dan peserta dengan BMI yang lebih tinggi cenderung memiliki total kadar kolesterol yang lebih tinggi juga. Korelasi antara BMI dan kolesterol total adalah 0,78. Dalam analisis regresi linier sederhana, persamaan garis yang menggambarkan hubungan antara variable bebas dan terikat adalah

$$\hat{y} = b_0 + b_1 x$$

Di mana \hat{y} adalah nilai yang diprediksi atau diharapkan, x adalah variable independent atau predictor, b_0 adalah estimasi intercept dan b_1 adalah perkiraan kemiringan. Perpotongan y dan

kemiringan diperkirakan dari data sampel dan meminimalkan jumlah selisih kuadrat antara (y) dan nilai prediksi yang diamati \hat{y} . Selisih ini disebut sebagai residual. Estimasi kuadrat terkecil dari perpotongan dan kemiringan y dihitung sebagai berikut:

$$b_1 = r \frac{s_y}{s_x} \text{ dan } b_0 = \bar{Y} - b_1 \bar{X}$$

Di mana r adalah koefisien korelasi sampel, \bar{X} dan \bar{Y} adalah nilai rata-rata x dan y serta s_x dan s_y adalah simpangan baku dari variable bebas x dan variable terikat y . estimasi kuadrat terkecil dari koefisien regresi b_0 dan b_1 menggambarkan hubungan antara BMI dan kolesterol total. Nilai $b_0 = 28,07$ dan $b_1 = 6,49$.

$$b_1 = r \frac{s_y}{s_x} = 0,78 \frac{30,8}{3,7} = 6,49$$

$$b_0 = \bar{Y} - b_1 \bar{X} = 205,9 - 6,49(27,4) = 28,07$$

Sehingga model persamaan yang dihasilkan adalah

$$\hat{y} = 28,07 + 6,49x$$

6. Evaluasi

a. Hitunglah koefisien korelasi *product moment* pada data di bawah ini

1) Data pertama

11. X	12. 1	13. 3	14. 6	15. 10	16. 12
17. Y	18. 5	19. 13	20. 25	21. 41	22. 49

Data kedua

23. X	24. 1	25. 3	26. 5	27. 7	28. 9
29. Y	30. 44	31. 34	32. 24	33. 14	34. 4

2) Data ketiga

35. X	36. 1	37. 1	38. 3	39. 5	40. 5
41. Y	42. 5	43. 1	44. 3	45. 1	46. 5

3) Data keempat

47. X	48. 1	49. 3	50. 6	51. 9	52. 11
53. Y	54. 12	55. 28	56. 37	57. 28	58. 12

- b. Pada sebuah percobaan laboratorium, hendak diketahui apakah ada korelasi angka bakteri dan usia bakteri dalam hitungan hari. Data sebagai berikut:

Age (x)	1	2	3	4	5	6	7	8
Jumlah bakteri (y)	34	106	135	181	192	231	268	300

Referensi Chapter 12

- Abdi, H., Edelman, B., Valentin, D., & Dowling, W. J. (2009). Experimental design and analysis for psychology. In *Oxford University Press*.
- Stephens, L. (2009). *Schaum's Outline of Beginning Statistik, Second Edition*. <https://books.google.co.uk/books?id=M7bV0GkTeY0C&pg=PA5&dq=interval+level+data&hl=en&sa=X&ved=0ahUKEwj-Khf-whd3iAhXsURUIHUJqBckQ6AEIKjAA#v=onepage&q=interval level data&f=false%0Ahttp://books.google.at/books?id=M7bV0G-kTeY0C>

BAGIAN 12

UJI CHI-SQUARE

1. Tujuan Pembelajaran

Setelah mengikuti pembelajaran ini mahasiswa mampu menelaah (menghitung, menyajikan, dan menginterpretasikan, serta membandingkan) hasil uji chi-square; mampu menjelaskan penggunaan tabel kontingensi; menjelaskan penggunaan analisis chi-square dan kemaknaan yang diperoleh dari uji chi-square, serta mampu melakukan perhitungan uji chi-square; menyajikan dan menginterpretasikan hasil uji chi-square.

2. Tabel contingency

Dalam penyajian deskriptif untuk menggambarkan sebuah kondisi nilai observasi ataupun harapan, salah satu bentuk yang paling sering digunakan adalah tabel kontingensi. Tabel ini dapat memberikan informasi observasi per pengamatan dan observasi keseluruhan dari data yang diperoleh. Tabel kontingensi atau dikenal juga dengan tabulasi silang adalah tabel yang berisi nilai frekuensi atau banyaknya amatan berdasarkan dari dua peubah kategorik. Tabel kontingensi dibuat berdasarkan banyaknya jumlah baris dan kolom. Tabel ini disajikan untuk memberikan gambaran hasil penelitian. Tabel ini digunakan dalam perhitungan inferensial untuk pengujian hipotesis misalnya perhitungan uji "t" dan atau . Tabel ini dapat dinamakan sesuai dengan banyaknya baris dan banyaknya kolom sehingga dikenal dengan nama tabel 2x2 atau 2x3 dan sebagainya.

Perhatikan tabel di bawah ini

Kejadian	A_1	A_2	A_k
Frekuensi teramati	x_1	x_2	x_k
Frekuensi ekspektasi	np_1	np_2	np_k

dalam menentukan apakah proporsi "keberhasilan" P dalam

suatu sampel dengan ukuran n yang diambil dari suatu populasi binomial berbeda secara signifikan dari proporsi populasi keberhasilan p . (Spiegel, Murray;Schiller, John;Srinivasan, 2004)

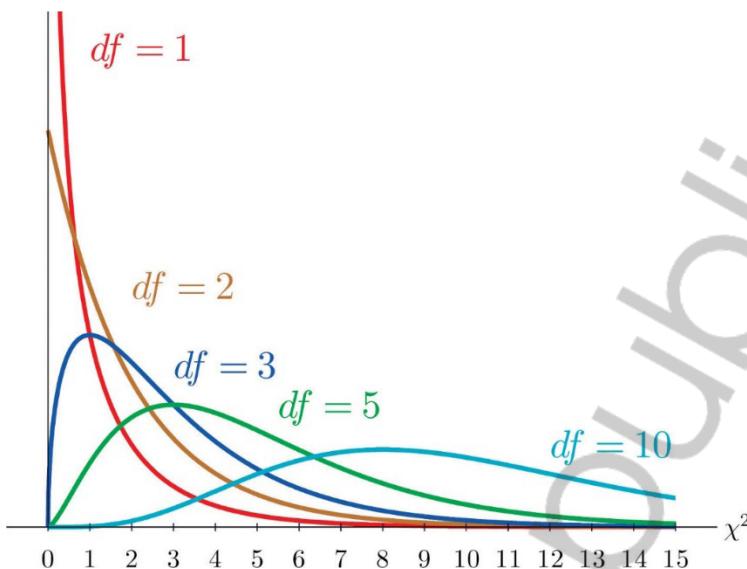
3. Analisis chi-square dan kemaknaan

Salah satu bentuk uji hipotesis dengan tujuan komparatif atau mencari perbedaan, terdapat salah satu uji yang sering digunakan yakni Chi-Square. Uji ini dapat digunakan untuk membandingkan sampel dengan populasi, membandingkan antar dua variable dan lebih dari dua variable. Skala data yang digunakan dalam uji ini adalah minimal nominal hingga ordinal (kategorik). Dalam bagian analisis inferensial, chi-square masuk dalam jenis non-parametrik.

Uji chi-square (diucapkan kai kuadrat) merupakan tes nonparametric yang paling banyak digunakan dalam penelitian ilmu sosial. Dikenal juga sebagai uji chi-kuadrat pearson. Uji ini digunakan untuk membandingkan distribusi satu variable dengan distribusi yang telah diketahui atau distribusi populasi.

Banyak penelitian melibatkan membuat kesimpulan terkait hubungan antara variable respons dan variable penjelas pada populasi tertentu. Dalam populasi itu, sebuah asosiasi ada jika distribusi probabilitas dari variable respons berubah dalam beberapa cara sebagai nilai dari perubahan variable penjelas. Uji signifikansi yang kemudian disebut uji chi-kuadrat yang digunakan untuk menentukan apakah dua variable kategori secara *statistik independent* atau terkait (EJD, Agresti and Finlay, 1998).

Sebuah uji statistik yang mengikuti distribusi chi-square yang memiliki bentuk positif dan miring (*skewness*). Uji statistik ini digunakan pula untuk *uji goodness of fit* dan *uji independensi*. Uji ini berdasarkan pada perbandingan antara nilai observasi dan frekuensi. Distribusi chi-kuadrat muncul dalam pengujian hipotesis mengenai independensi dua variable acak dan mengenai apakah variable acak diskrit mengikuti distribusi tertentu. Variable acak chi-kuadrat adalah variable acak yang mengasumsikan hanya nilai positif dan mengikuti distribusi chi-kuadrat (shafer, douglas; zhang, 2012).



Gambar 13.1 Menunjukkan beberapa distribusi chi-kuadrat untuk derajat kebebasan yang berbeda

Nilai peubah acak chi-kuadrat dengan $df = k-1$ ($df = degrees of freedom$). Berikutnya sebuah contoh ilustrasi apakah dua variable acak mengambil nilai secara independen atau tidak atau apakah nilai yang satu memiliki hubungan dengan nilai yang lain. Dengan demikian hipotesis akan diungkapkan dalam kalimat.

Terdapat teori bahwa jenis kelamin bayi dalam kandungan berhubungan dengan detak jantung bayi. Bayi perempuan cenderung memiliki detak jantung yang lebih tinggi. Teori ini hendak diuji dengan kemudian mencatat detak jantung dari 40 bayi yang diambil selama pemeriksaan prenatal terakhir ibu sebelum melahirkan. Untuk masing-masing 40 sampel ini dilakukan pencatatan pada jenis kelamin dan detak jantung. Kedua ukuran ini disebut faktor. Berikut rumusan hipotesisnya

H_0 = Jenis kelamin dan detak jantung bayi tidak memiliki hubungan (independent)

H_1 = Jenis kelamin dan detak jantung bayi memiliki hubungan (not independent)

Faktor jenis kelamin memiliki dua kategori yakni laki-laki dan perempuan. Faktor kedua, yakni detak jantung menjadi dua tingkat yaitu rendah dan tinggi. Batas diantara keduanya adalah 145 detak per menit. Jika denyut jantung di bawah 145 denyut per menit maka akan dikategorikan rendah dan jika di atas 145 dianggap tinggi. Uraian tersebut menghasilkan tabel kontingensi 2x2 sebagai berikut

Tabel 13.1 Jenis Kelamin Bayi dan Detak Jantung

		Detak jantung		Row total
		Rendah	Tinggi	
Jenis kelamin	Perempuan	11	7	18
	Laki-laki	17	5	22
Total kolom		28	12	Total = 40

Tabel di atas menyajikan nilai observasi, sehingga dibutuhkan nilai harapan (*expected value*) yang dapat dihitung dengan prosedur sebagai berikut

$$\text{expected value} = \frac{R \times C}{\text{total } (n)}$$

Sehingga dapat dihitung seperti uraian di bawah ini

$$\text{expected value baris 1 kolom 1} = \frac{18 \times 28}{40} = 12,6$$

$$\text{expected value baris 1 kolom 2} = \frac{18 \times 12}{40} = 5,4$$

$$\text{expected value baris 2 kolom 1} = \frac{22 \times 28}{40} = 15,4$$

$$\text{expected value baris 2 kolom 2} = \frac{22 \times 12}{40} = 6,6$$

Selanjutnya tabel 13.1 akan menjadi

Tabel 13.2 Nilai Observasi dan Harapan Jenis Kelamin dan Detak Jantung Bayi

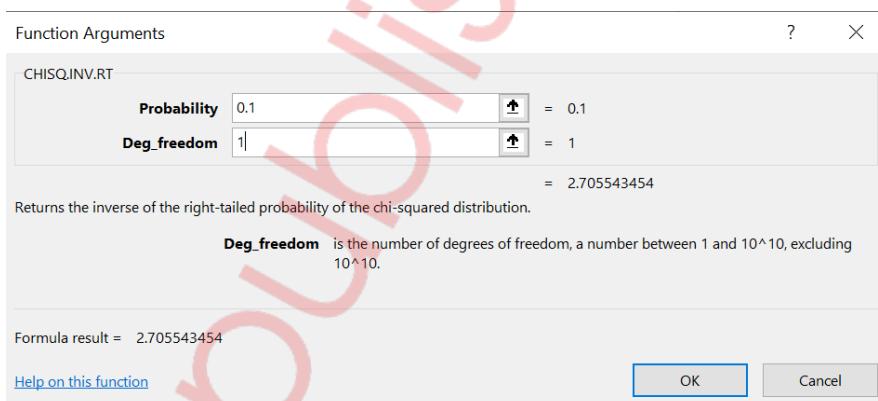
Rendah

Jenis kelamin		Detak jantung		Row total
		Tinggi		
Perempuan		Observasi:11 Harapan: 12.6	Observasi:7 Harapan: 5,4	R = 18
	Laki-laki	Observasi:17 Harapan: 15,4	Observasi: 5 Harapan: 6,6	R = 22
Total kolom		C = 28	C = 12	n = 40

$$chi - squared \chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

$$\begin{aligned} \chi^2 &= \sum \frac{(O_i - E_i)^2}{E_i} \\ &= \frac{(11 - 12.6)^2}{12.6} + \frac{(7 - 5.4)^2}{5.4} + \frac{(17 - 15.4)^2}{15.4} + \frac{(5 - 6.6)^2}{6.6} = 1.231361 \end{aligned}$$

Nilai hitung chi-kuadrat adalah 1,23. Selanjutnya dari $df = k-1 = 2 - 1 = 1$ dengan tingkat signifikansi sebesar 10 % maka berikut ilustrasi mencari nilai tabel chi-kuadrat menggunakan bantuan Ms Excel



Gambar 13.2 Pencarian nilai tabel chi-kuadrat dengan MS Excel

Jika menggunakan tabel chi-kuadrat probabilitas maka tampilannya seperti ini

df	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	---	---	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278

Gambar 13.3 Tabel Probabilitas Chi-Kuadrat

Nilai yang diperoleh adalah 2.706. Selanjutnya akan dibandingkan dengan nilai hitung 1.231.

$1.231 < 2.706$, keputusannya adalah gagal menolak . Pada tingkat signifikansi 10%, data tidak memberikan cukup bukti untuk menyimpulkan bahwa detak jantung dan jenis kelamin memiliki relasi atau hubungan.

4. Asumsi Uji Chi – Square

Sebagaimana yang telah dijelaskan pada poin sebelumnya, bahwa uji ini masuk dalam golongan nonparametric sehingga tidak mengikuti asumsi seperti tes-tes pada parametrik. Dapat dikatakan, bahwa data tidak perlu mengikuti bentuk kurva normal. Beberapa asumsi yang diperlukan untuk melaksanakan uji ini adalah sebagai berikut

- Variable yang dianalisis memiliki data kategori dan diukur pada tingkat nominal dengan dua atau lebih yang saling eksklusif dan lengkap
- Seluruh pengamatan harus independen satu sama lain. Yang bermakna setiap peserta memberikan data hanya untuk satu sel gabungan. Ini disebut sebagai asumsi independensi
- Data yang dianalisis harus berupa frekuensi atau jumlah bukan skor
- Ukuran sampel harus cukup besar. Maka perhatikan aturan ini
 - Tidak ada sel yang memiliki frekuensi harapan (*expected value*) kurang dari 1.

- 2) Tidak lebih dari 20% sel harus memiliki frekuensi yang diharapkan kurang dari 5. Misalnya dalam tabel 3x4 (total 15 sel) tidak lebih dari tiga sel ($15 \times 20\%$) harus memiliki frekuensi yang diharapkan kurang dari lima kasus.
- 3) Dalam tabel 2x2, semua sel harus memiliki frekuensi yang diharapkan dari lima kasus atau lebih (per sel).

Sebagai sebuah alternatif, jika ukuran sampel kecil atau asumsi aturan 20% tidak terpenuhi, dapat dipertimbangkan penggunaan uji eksak Fisher. Tes ini digunakan untuk membandingkan antara dua variable dikotomis, yakni tabel kontingensi khusus 2×2 . Oleh karena itu jika dimiliki tabel lebih dari 2×2 (seperti 3×4 , 4×3 dan lainnya) maka rekomendasinya adalah pemampatan atau pencutan menjadi tabel 2×2 . Lalu uji ini digunakan jika ada asumsi frekuensi yang diharapkan dilanggar. Uji ini akan dibahas lebih lanjut pada kesempatan berikutnya.

5. Langkah perhitungan uji chi-square

Fungsi chi-square dapat digunakan untuk uji kesesuaian (*goodness of fit*), uji perbandingan dua sampel dengan koreksi kontinuitas dari Yates dan Fisher eksak, serta uji perbandingan chi-square untuk lebih dari dua variable.

Berikut prosedur untuk melakukan perhitungan chi-square

a. Contoh kasus

NCHS (*National Center for Health Statistik*s) menyediakan data tentang distribusi berat badan (dalam kategori) penduduk di sebuah negara A pada tahun 2002. Distribusi didasarkan pada nilai spesifik indeks massa tubuh (BMI) dihitung sebagai berat badan dalam kilogram dan tinggi badan dalam meter kuadrat.

Berikut kategorinya

- 1) *Underweight* $< 18,5$
- 2) Normal $18,5 - 24,9$
- 3) Berat badan berlebih $25 - 29,9$
- 4) Obesitas ≥ 30

Penduduk Amerika dikategorikan sebagai berikut

- 1) Kurus 2% (0,02)
- 2) Berat badan normal 39% (0,39)
- 3) Berat badan berlebih 36% (0,36)
- 4) Obesitas 23% (0,23)

Besar sampel yang diperoleh adalah $n = 3536$ partisipan yang hadir pada pertemuan ke-7 di sebuah institusi penelitian. Berikut tampilan datanya:

Tabel 13.3 Distribusi IMT di Pusat Penelitian Negara A

	Kurus < 18,5	Berat normal 18,5 – 24,9	Berat badan berlebih 25,0 – 29,9	Obesitas 30+	Total
Frekuensi observasi (O)	20	932	1374	1000	3326
Frekuensi harapan (E)	66,5	1297,1	1197,4	765,0	3326

- b. Membentuk hipotesis dan menentukan tingkat kepentingan (*level of significance*)

$$H_0: p_1 = 0,02, p_2 = 0,39, p_3 = 0,36, p_4 = 0,23 \text{ atau setara (equivalently)}$$

H_0 : distribusi dari respond adalah 0,02, 0,39, 0,36 dan 0,23

H_1 : H_0 tidak benar (*false*)

$$\alpha = 0,05$$

- c. Memilih uji statistik yang sesuai

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Yang perlu dilakukan adalah pengecekan minimal besar sampel apakah adekuat atau tidak. Secara spesifik, perlu dilakukan pengecekan minimal $n \geq 5$. Besar sampel pada studi ini mencapai $n = 3326$ dengan proporsi pada hipotesis null adalah 0.02, 0.39, 0.36 dan 0.23. dengan demikian

A	B	C
3326	0.02	66.52
	0.39	1297.14
	0.36	1197.36
	0.23	764.98

Gambar 13.4 Perhitungan nilai harapan berdasarkan nilai observasi per kategori IMT

Besar sampel lebih dari adekuat, sehingga uji statistik dengan chi-square dapat dilanjutkan.

- d. Menyiapkan aturan keputusan

Derajat kebebasan atau dan tingkat kepentingan 5%.

Chi-square Distribution Table

d.f.	.995	.99	.975	.95	.9	.1	.05	.025	.01
1	0.00	0.00	0.00	0.00	0.02	2.71	3.84	5.02	6.63
2	0.01	0.02	0.05	0.10	0.21	4.61	5.99	7.38	9.21
3	0.07	0.11	0.22	0.35	0.58	6.25	7.81	9.35	11.34
4	0.21	0.30	0.48	0.71	1.06	7.78	9.49	11.14	13.28
5	0.41	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09
6	0.68	0.87	1.24	1.64	2.20	10.64	12.59	14.45	16.81
7	0.99	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.72
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14
15	4.60	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58

Gambar 13.5 Tabel Chi-Square

Berdasarkan derajat kebebasan dan tingkat kepentingan, maka nilai tabel adalah 7,81. Maka aturan keputusan adalah

Tolak H_0 if $X^2 \geq 7,81$

- e. Menghitung uji statistik

Dengan menggunakan nilai observasi dan harapan dapat dihitung nilai statistik sebagai berikut

$$\chi^2 = \frac{(20 - 66,5)^2}{66,5} + \frac{(932 - 1297,1)^2}{1297,1} + \frac{(1374 - 1197,4)^2}{1197,4} + \frac{(1000 - 765,0)^2}{765,0}$$

$$\chi^2 = 32,52 + 102,77 + 26,05 + 72,19 = 233,53$$

f. Kesimpulan

Kesimpulan dari hasil perhitungan ini adalah tolak karena $233,53 > 7,81$.

Sehingga dapat diinterpretasikan bahwa distribusi IMT penduduk Negara A pada tahun 2002 tidak mengikuti distribusi kategori populasi

BAGIAN 13

UJI HIPOTESIS

BEDA RERATA INDEPENDENT

(NON-PARAMETRIK)

1. Tujuan pembelajaran

Setelah mengikuti pembelajaran ini mahasiswa mampu menelaah (menghitung, menyajikan, dan menginterpretasikan, serta membandingkan) uji hipotesis beda rerata independen menggunakan Kruskall Wallis, Sign Test dan Wilcoxon Match, serta mampu menelaah (menghitung, menyajikan, dan menginterpretasikan, serta membandingkan) uji hipotesis beda rerata independen menggunakan Median Test dan Mann Whitney

2. Pengantar

Penyakit langka di definisikan sebagai penyakit dengan prevalensi kurang dari 200.000 di sebuah negara A dan mempengaruhi paling sedikit 5 dari 10.000 orang. Obat yang digunakan untuk mengobati penyakit langka sering disebut sebagai obat yatim piatu karena tidak menarik bagi produsen atau pengembang karena keterbatasan potensi profitabilitasnya. Pada tahun 1983, sebuah kongres meloloskan *orphan drug act*, tindakan ini menawarkan pajak dan lainnya yang bersifat pemasaran eksklusif bagi sponsor yang terlibat. Studi penyakit langka ini biasanya melibatkan sampel yang kecil, karena besar populasi yang terbatas atau pasien yang tersedia untuk terlibat tidak ada.

Ketika ukuran sampel kecil ($n \leq 30$) dan hasil distribusi tidak diketahui sehingga tidak dapat diasumsikan terdistribusi normal, maka pengujian alternatif ini disebut uji nonparametric. Uji ini disebut juga uji bebas distribusi dan bebas asumsi (atau asumsi lebih sedikit dibanding tes parametrik). Tes parametrik pada umumnya melibatkan probabilitas distribusi spesifik misalnya distribusi

normal. Karena asumsi yang lebih sedikit, uji nonparametric pada umumnya dianggap kurang kuat dibandingkan dengan uji parametrik. Akan tetapi uji ini memiliki beberapa keunggulan, diantaranya adalah uji ini satu-satunya cara untuk menganalisis data dengan kriteria tertentu. Misalnya hasil yang nominal atau ordinal, hasil peringkat, adanya nilai yang *outlier* sehingga sulit untuk dianalisis dengan uji parametrik (karena asumsi tertentu akan mengalami gangguan jika terdapat nilai *outlier*) serta uji nonparametric relatif sederhana untuk dipraktikkan.

3. Kruskall wallis

Uji nonparametric yang popular untuk membandingkan hasil diantara kelompok yang memiliki jumlah lebih dari 2 (dua) adalah uji Kruskal wallis. Uji ini digunakan untuk membandingkan median antara kelompok k pembanding ($k > 2$) dan terkadang digambarkan sebagai ANOVA dengan bentuk data peringkat. Atau secara singkatnya, dalam beberapa kasus penelitian ingin dibandingkan rerata diantara sampel yang berjumlah lebih dari dua akan tetapi distribusi dari data tidak normal atau data yang terkumpul adalah ordinal. Distribusi yang digunakan pada uji statistik ini adalah chi-kuadrat dengan $df = g - 1$. Pemberian peringkat pada pengamatan dan kemudian membandingkan peringkat rata-rata untuk berbagai kelompok.

Hipotesis nol dan penelitian untuk uji ini dinyatakan sebagai berikut

H_0 : median populasi k adalah sama

H_1 : median populasi k adalah tidak sama

Prosedur tes melibatkan data pengamatan dari sampel k menjadi satu sampel yang digabung, mengamati dari mana sampel per pengamatan berasal dan kemudian diberikan peringkat terendah ke tertinggi, dari 1 sampai N, di mana . Untuk ilustrasi yang nyata, berikut adalah contoh dalam pengujian hipotesis.

a. Contoh kasus

Sebuah studi klinis dirancang untuk menilai perbedaan tingkat albumin pada orang dewasa yang mengikuti diet protein yang

berbeda. Kadar Albumin diukur dalam gram/desiliter darah (g/dL), dan diet rendah protein yang sering disarankan pada penderita penyakit ginjal.

Tabel 13.4 Tingkat Albumin pada tiga jenis diet

5% protein	10% protein	15% protein
3.1	3.8	4.0
2.6	4.1	5.5
2.9	2.9	5.0
	3.4	4.8
	4.2	

Tiga diet dibandingkan, yakni 5% sampai 15% protein. Tingkat albumin peserta yang mengikuti diet ditunjukkan pada tabel 13.4.

- b. Merumuskan hipotesis dan menentukan tingkat signifikansi
- Pertanyaan dalam contoh kasus ini adalah apakah terdapat perbedaan dalam kadar albumin diantara tiga kelompok diet terebut. Berdasarkan referensi, kadar albumin normal berada pada angka 3,4 g/dL dan 5,4 g/dL. Pada tabel dapat dilihat, bahwa diet protein 15% memiliki kadar albumin yang lebih tinggi daripada diet protein 5%. Apakah perbedaan yang diamati ini memiliki perbedaan signifikan secara statistik?

Dalam contoh kasus ini, data yang terkumpul berbentuk data kontinu akan tetapi memiliki besar sampel (n) yang kecil dan seluruh perbandingan memiliki n yang tidak sama, sehingga uji nonparametric menjadi sesuai untuk dilakukan.

Berikut hipotesis yang akan diuji pada tingkat signifikansi 5%

$$H_0 = \text{median pada tiga kelompok adalah sama}$$

$$H_1 = \text{tidak semua median pada tiga kelompok sama}$$

- c. Pemilihan uji statistik yang sesuai

Statistik uji untuk uji Kruskal-Wallis dilambangkan H dan didefinisikan sebagai berikut:

$$H = \left(\frac{12}{N(N+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} \right) - 3(N+1)$$

Di mana k = banyaknya kelompok pembanding, N = besar keseluruhan sampel (total), n_j adalah ukuran sampel dalam kelompok ke- j dan R_j adalah jumlah peringkat dalam kelompok ke- j .

- d. Menetapkan wilayah keputusan

Karena terdapat 3 kelompok diet sehingga $df = k - 1 = 3 - 1 = 2$ dan $\alpha = 0,05$ maka nilai kritisnya adalah 5.656 dan aturan keputusan adalah menolak jika nilai hitung (H) ≥ 5.656

- e. Menghitung statistik uji

Berikutnya untuk melakukan uji, ditetapkan peringkat menggunakan prosedur Kruskal Wallis. Langkah pertama dalam penetapan peringkat adalah mengurutkan data dari yang terkecil sampai yang terbesar.

Perhatikan tabel berikut

Tabel 13.5 Menetapkan Peringkat pada Data Albumin

Besar sampel (diurutkan terkecil ke terbesar)						Ranking		
5% protein	10% protein	15% protein	5% protein	10% protein	15% protein	5% protein	10% protein	15% protein
3.1	3.8	4.0	2.6			1		
2.6	4.1	5.5	2.9	2.9		2.5	2.5	
2.9	2.9	5.0	3.1			4		
	3.4	4.8		3.4			5	
	4.2			3.8			6	
					4.0			7
					4.1			8
					4.2			9
						4.8		10
						5.0		11
						5.5		12

Berikut ditampilkan penyelesaian menggunakan bantuan Ms Excel

5% protein	10% protein	15% protein	5% protein	10% protein	15% protein	5% protein	10% protein	15% protein
3.1	3.8	4	2.6			1		
2.6	4.1	5.5	2.9	2.9		2.5	2.5	
2.9	2.9	5	3.1			4		
	3.4	4.8		3.4			5	
	4.2			3.8			6	
					4			7
				4.1			8	
				4.2			9	
					4.8			10
					5			11
					5.5			12
N	12	n1	n2	n3	R1	R2	R3	
N+1	13		3	5	4	7.5	30.5	40
N(N+1)	156				R ₁ ²	R ₂ ²	R ₃ ²	
3(N+1)	39				56.25	930.25	1600	
12/N(N+1)	0.0769				18.75	186.05	400	
							604.8	
							46.523	
							7.5231	

Gambar 13.6 Penyelesaian menggunakan bantuan MS Excel

$$H = \frac{12}{12(12+1)} \left(\frac{7.5^2}{3} + \frac{30.5^2}{5} + \frac{40^2}{4} \right) - 3(12+1)$$

= ... (dapat dicoba hitung dan dicocokan dengan penyelesaian dari MS Excel)

Dalam seluruh tes statistik yang dilakukan, harus ditentukan apakah tes yang diamati mendukung hipotesis nol atau hipotesis penelitian. Untuk melihat hal ini maka perlu nilai hitung H dibandingkan dengan H tabel. Dengan tingkat signifikansi 5% maka dari tabel di bawah ini dapat diketahui nilai tabel H adalah,

TABLE 8. Critical Values for the Kruskal–Wallis Test, $\alpha = .05$ and $\alpha = .01$

Three groups			Four groups							
n_1	n_2	n_3	$\alpha = .05$	$\alpha = .01$	n_1	n_2	n_3	n_4	$\alpha = .05$	$\alpha = .01$
5	4	2	5.273	7.205	4	3	3	2	6.795	8.333
5	4	3	5.656	7.445	4	3	3	3	6.984	8.659
5	4	4	5.657	7.760	4	4	1	1	5.945	7.909

Gambar 13.7 Tabel Kritis Kruskal-Wallis

f. Menarik kesimpulan

Berdasarkan ukuran sampel (3,5 dan 4) serta tingkat signifikansi (0,05), nilai kritis berdasarkan gambar adalah 5.656. Dengan demikian, ditolak karena $7.52 > 5.656$. Secara signifikan, terbukti bahwa terdapat perbedaan media pada kadar albumin diantara tiga kelompok diet yang berbeda.

4. Sign test

Sign test atau dikenal juga sebagai uji tanda merupakan uji nonparametric paling sederhana untuk data yang berpasangan. Pendekatan yang digunakan adalah menganalisis tanda-tanda perbedaan skor. statistik uji untuk *sign test* adalah jumlah tanda positif atau jumlah tanda negatif, yang mana yang lebih kecil. (Sullivan, 2018)

Sebuah contoh kasus untuk mengilustrasikan hal tersebut adalah suatu penyelidikan klinis untuk menilai efektivitas obat baru yang dirancang untuk mengurangi pengulangan perilaku pada anak yang terkena autisme. Obat ini akan dianggap efektif jika anak-anak menunjukkan lebih sedikit perilaku berulang dengan pengobatan yang dibandingkan dengan kasus tanpa pengobatan. Sebanyak 8 anak diamati dalam penelitian ini selama jangka waktu 3 jam pengamatan. Kemudian pengamatan dilakukan kembali dalam kurun waktu 1 minggu setelahnya. Perilaku berulang dinilai pada skala 0 sampai 100, misalnya skor 0 menunjukkan bahwa selama periode pengamatan anak tidak melakukan pengulangan perilaku, sedangkan skor 100 menunjukkan bahwa anak terus menerus terlibat dalam perilaku berulang

Data tersaji dalam tabel berikut

Tabel 13.6 Data Skor Perilaku Anak Austime

Anak	Sebelum perlakuan	1 minggu setelah perlakuan	Perbedaan (sebelum - sesudah)	Tanda
1	85	75	10	+
2	70	50	20	+
3	40	50	-10	-
4	65	40	25	+
5	80	20	60	+
6	75	65	10	+
7	55	40	15	+
8	20	25	-5	-

Dari tabel di atas, jika Hipotesis nol benar, maka yang diharapkan terjadi adalah terlihat tanda positif dan tanda negatif masing-masing memiliki proporsi 50%. Jika hipotesis penelitian benar, maka yang harusnya terlihat adalah lebih banyak tanda positif (+).

statistik uji untuk *sign test* adalah jumlah tanda positif atau jumlah tanda negatif, mana yang lebih kecil. Pada contoh di tabel 13.6, terdapat dua tanda negatif dan enam tanda positif.

ditolak jika tanda positif atau negatif kurang dari atau sama dengan nilai kritis. Akan tetapi gagal ditolak jika jumlah tanda positif atau negatif lebih besar dari nilai kritis.

a. Contoh kasus

Perawatan kemoterapi baru diusulkan untuk penderita kanker payudara. Pengamatan menyelidiki kemampuan pasien dalam mentolerir pengobatan dan kualitas hidup sebelum dan sesudah menerima pengobatan kemoterapi baru. Kualitas hidup (QOL) diukur pada skala ordinal dengan kategori sebagai berikut

- 1 = buruk
- 2 = cukup
- 3 = baik
- 4 = sangat baik

5 = unggul

Pertanyaan peneliti adalah apakah ada perbedaan QOL setelah pengobatan kemoterapi dibandingkan dengan sebelumnya. Data tersaji pada tabel berikut

Tabel 13.7 kualitas hidup sebelum dan sesudah pengobatan kemoterapi

Pasien	QOL sebelum	QOL sesudah
1	3	2
2	2	3
3	3	4
4	2	4
5	1	1
6	3	4
7	2	4
8	3	3
9	2	1
10	1	3
11	3	4
12	2	3

- b. Merumuskan hipotesis dan menentukan tingkat signifikan

H_0 : perbedaan median adalah sama dengan nol

H_1 : perbedaan median tidak sama dengan nol

$\alpha = 0.05$

- c. Pemilihan statistik uji yang sesuai

Statistik uji untuk yang digunakan adalah uji tanda di mana akan dilihat apakah tanda positif atau negatif yang lebih kecil.

- d. Mengatur aturan keputusan

Nilai kritis dapat ditemukan pada gambar tabel di bawah ini.

Two-Sided Test α	.10	.05	.02	.01
One-Sided Test α	.05	.025	.01	.005
<i>n</i>				
1				
2				
3				
4				
5	0			
6	0	0		
7	0	0	0	
8	1	0	0	0
9	1	1	0	0
10	1	1	0	0
11	2	1	1	0
12	2	2	1	1
13	3	2	1	1

Gambar ... Nilai Kritis untuk Uji Tanda

Dalam menentukan nilai kritis yang seusia, dibutuhkan informasi besar sampel (pada contoh di atas $n = 12$) dan tingkat signifikansi yang digunakan (5% atau 0.05). Maka aturan keputusan yang dapat dibentuk adalah H_0 ditolak jumlah tanda positif atau negatif lebih kecil dari 2 (≤ 2)

e. Hitung statistik uji

Menggunakan bantuan MS Excel, akan diketahui berapakah jumlah tanda positif atau negatif dalam kasus ini. Perhatikan tabel berikut.

Tabel 13.8 selisih nilai kualitas hidup sebelum dan sesudah pengobatan kemoterapi

Pasien	QOL sebelum	QOL sesudah	Sesudah - sebelum
1	3	2	-1
2	2	3	1
3	3	4	1
4	2	4	2

Pasien	QOL sebelum	QOL sesudah	Sesudah – sebelum
5	1	1	0
6	3	4	1
7	2	4	2
8	3	3	0
9	2	1	-1
10	1	3	2
11	3	4	1
12	2	3	1

Dalam tabel, terdapat dua nilai nol (0). Sehingga ditetapkan secara acak tanda negatif (-) untuk pasien nomor 5 dan tanda positif (+) untuk pasien nomor 8. Maka statistik uji menetapkan jumlah tanda negatif adalah 3.

f. Kesimpulan

gagal ditolak karena $3 > 2$. Secara statistik, tidak dimiliki bukti yang signifikan untuk menunjukkan ada perbedaan kualitas hidup pasien setelah pengobatan kemoterapi dibandingkan dengan sebelum pengobatan.

Nilai P juga dapat dihitung menggunakan distribusi binomial, dengan diketahui $n = 12$ dan $p = 0.5$. nilai p untuk dua sisi adalah $p = 2xP(x \leq 3)$ yang setara dengan $p = P(x \leq 3) + P(x \geq 9)$

Nilai probabilitas dua sisi adalah probabilitas pengamatan sedikit atau banyak tanda positif (atau negatif). Di sini tanda yang diamati adalah 3 tanda negatif dan 9 tanda positif ($n=12$).

Berikut perhitungan nilai probabilitas (p-value) menggunakan formula Binomial

$$\begin{aligned}
 p &= 2xP(x \leq 3) = 2(P(0) + P(1) + P(3)) = \\
 p(x \text{ sukses}) &= \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} = p(0) \\
 &= \frac{12!}{0!(12-0)!} \cdot 0.5^0 \cdot (1-0.5)^{12-0} = 1 \times 1 \times \frac{1}{4096} = 0.0002
 \end{aligned}$$

Gunakan Langkah tersebut dengan bantuan kalkulator (dalam buku ini digunakan kalkulator online yakni <https://www.desmos.com/scientific>

Gambar Langkah penyelesaian jika $P(0)$


$$0.5 - 0.5$$
$$0.5^0$$
$$0.5^{12}$$
$$= 0.5 \quad (8)$$
$$= 1$$
$$= \frac{1}{4096} \quad (9)$$



Gunakan fungsi a^b pada kalkulator untuk mencari hasil dari $5^0 \cdot (1 - 0.5)^{12-0}$

Ingat $0! = 1$. Maka

$$p = 2xP(x \leq 3) = 2(P(0) + P(1) + P(2) + P(3))$$
$$= 2 \times (0,0002 + 0,0029 + 0,0161 + 0,0537) = 2x(0,0729)$$
$$= 0,1458$$

Karena p-value adalah 0,1458 melebihi tingkat signifikansi (0,05) sehingga secara signifikan tidak ada perbedaan kualitas hidup pasien setelah pengobatan kemoterapi dilakukan.

5. Wilcoxon match paired test

Apabila ditemukan kondisi data tidak berdistribusi normal, kemudian data hasil adalah peringkat, bukan interval rasio, data berasal dari anggota yang berpasangan (before – after) atau dari populasi yang sama, maka salah satu uji statistik yang dapat digunakan adalah wilcoxon match paired test atau Wilcoxon signed rank test. Coba perhatikan dua (2) ilustrasi berikut

ILUSTRASI 1

Seorang pengamat ingin mengetahui obat baru untuk gejala ADHD (*attention deficit hyperactivity disorder*) dalam mengurangi jumlah gangguan perilaku yang terjadi di lingkungan sekolah. Sejumlah siswa yang berusia 8-10 tahun dari beberapa daerah yang dicocokkan dari segi jenis kelamin, usia dan jumlah gangguan selama 30 hari terakhir digabung untuk belajar Bersama. Satu anggota dari setiap pasangan menerima obat baru dan lainnya menjadi kelompok control. Jumlah gangguan perilaku anak direkam selama 1 minggu. Apakah terdapat perbedaan kelompok yang menerima obat dan kelompok kontrol?

ILUSTRASI 2

Peneliti di sebuah perguruan tinggi tertarik untuk mengetahui apakah obat flu mempengaruhi waktu reaksi. Waktu reaksi peserta diukur sebelum dan satu jam setelah konsumsi obat flu. Dengan asumsi data tidak berdistribusi normal, gunakan data yang tersedia dan tentukan apakah ada perubahan waktu reaksi yang signifikan pada tingkat = 0.05

Perhatikan dan baca dengan saksama kedua ilustrasi di atas. Pada ilustrasi pertama diberikan gambaran terdapat dua kelompok yang memiliki kesamaan karakter dari segi jenis kelamin, usia, dan jumlah gangguan. Diberikan nama berbeda pada tiap kelompok, yakni kelompok pertama adalah kelompok dengan pengobatan (kelompok kasus) sedangkan pada kelompok berikutnya adalah kelompok control (kelompok tanpa pemberian obat). *Wilcoxon match paired* test berada pada ilustrasi pertama (1), karena terdapat kata kunci “dicocokkan” atau *match paired* yang artinya berpasangan yang memiliki kecocokan. Pada ilustrasi kedua, merupakan kasus *before – after treatment*. Kasus sebelum dan sesudah di mana peserta yang diukur adalah peserta yang sama.

Dalam buku *Fundamental Statistical Principles for the Neurobiologist* (Scheff, Stephen W, 2016) uji statistik ini diberikan nama Wilcoxon Matched Pairs Signed Rank Test atau Wilcoxon

Signed Rank Test. Berikut akan dilanjutkan dengan prosedur dari uji ini

a. Contoh kasus

Sejumlah anak mendaftar dalam penelitian terkait efektivitas obat baru untuk mengurangi perilaku berulang pada anak-anak yang terkena gejala autisme. Data tersaji dalam tabel berikut

Tabel ... Perilaku Berulang Sebelum dan Sesudah Perawatan

Anak	Sebelum perlakuan	Setelah perlakuan selama 1 minggu
------	-------------------	-----------------------------------

1	85	75
2	70	50
3	40	50
4	65	40
5	80	20
6	75	65
7	55	40
8	20	25

Perilaku berulang diamati selama 3 jam dan diukur keduanya sebelum dan sekali lagi setelah minum obat baru. Pertanyaan dari penelitian ini adalah apakah terdapat peningkatan perilaku berulang setelah pengobatan jika dibandingkan dengan sebelumnya?

Tabel ... Perilaku Berulang Sebelum dan Sesudah Perawatan

Anak	Sebelum perlakuan	Setelah perlakuan selama 1 minggu	Perbedaan skor (before - after)
1	85	75	10
2	70	50	20
3	40	50	-10
4	65	40	25
5	80	20	60

Anak	Sebelum perlakuan	Setelah perlakuan selama 1 minggu	Perbedaan skor (before – after)
6	75	65	10
7	55	40	15
8	20	25	-5

- b. Merumuskan hipotesis dan menentukan tingkat signifikansi
 Hipotesis uji Wilcoxon peringkat bertanda memiliki kesamaan dengan uji tanda (*sign tets*) di mana hipotesis menyangkut median populasi dari perbedaan skor. Hipotesis penelitian dapat dirumuskan pada satu atau dua sisi. Pada contoh ini adalah tes satu sisi

H_0 : perbedaan median adalah nol

H_1 : perbedaan median adalah positif

$\alpha = 0,05$

- c. Pemilihan statistik uji yang sesuai

Statistik uji untuk uji peringkat bertanda Wilcoxon adalah W , didefinisikan sebagai $W+$ dan W -paling kecil, di mana diperoleh dari masing-masing jumlah ranking tanda + (positif) dan - (negatif)

- d. Menentukan keputusan uji

Nilai kritis W dapat ditemukan pada gambar tabel di bawah ini

	Two-Sided Test α	.10	.05	.02	.01
One-Sided Test α		.05	.025	.01	.005
<i>n</i>					
5		1			
6		2	1		
7		4	2	0	
8		6	4	2	0
9		8	6	3	2
10		11	8	5	3
11		14	11	7	5
12		17	14	10	7
13		21	17	13	10
14		26	21	16	13
15		30	25	20	16

Gambar.... Nilai Kritis W

Ukuran sampel adalah 8 dan nilai kritis yang ditetapkan adalah 0.05 untuk uji satu sisi, maka

$$H_0 \text{ ditolak jika } W \leq 6$$

e. Perhitungan

Langkah perhitungan pada uji ini adalah memberikan urutan/ peringkat pada perbedaan skor. Lihat tabel berikut

Tabel Nilai Peringkat Absolut dari perbedaan skor

Perbedaan skor (before - after)	Susunan perbedaan skor terkecil – terbesar	Peringkat	Tanda peringkat
10	-5	1	-1
20	10	3	3
-10	-10	3	-3
25	10	3	3
60	15	5	5
10	20	6	6
15	25	7	7
-5	60	8	8

Statistik uji untuk uji Wilcoxon signed rank adalah W didefinisikan sebagai $W+$ dan $W-$ yang merupakan jumlah dari peringkat positif dan negatif. Jika hipotesis nol benar, $W+$ dan $W-$ akan serupa. Jika hipotesis penelitian benar, terdapat peringkat yang lebih tinggi dan bertanda positif.

Dalam kasus ini, $W+ = 32$ dan $W-(1+3) = 4$. $W+$ jauh lebih besar dibandingkan $W-$. Selanjutnya adalah menentukan apakah statistik uji yang diamati mendukung hipotesis nol atau penelitian. Diketahui ukuran sampel (n) = 8 dan tingkat signifikansi untuk uji satu sisi (α) = 0.05. Berdasarkan gambar tabel di atas dengan tingkat signifikansi 0.05 dan $n = 8$ maka nilai tabel W adalah 6.

f. Kesimpulan

Karena nilai W hitung (4) kurang daripada W tabel (6), maka secara statistik dengan tingkat signifikansi 0.05 maka ditolak. Dapat ditarik kesimpulan bahwa perbedaan median adalah positif atau ada peningkatan yang signifikan dalam pengulangan perilaku pada anak yang terkena autisme setelah perawatan dibandingkan sebelum perawatan.

6. Median test

Merupakan statistik yang sangat sederhana yang dapat digunakan untuk menentukan dengan cepat apakah terdapat perbedaan antara dua sampel yang independen (bebas atau tidak berpasangan) bahkan dengan ukuran besar sampel per kelompok yang tidak sama. Seperti *sign test* atau uji tanda, statistik ini hanya menghitung berapa banyak pengamatan yang terjadi dalam kategori tertentu. Prosedur dari median test dapat diilustrasikan pada contoh berikut ini

a. Contoh kasus/data

Peneliti muda hendak mengetahui apakah siswa laki-laki sekolah menengah pertama di perkotaan dan pedesaan berbeda tingkat Kesehatan mentalnya. Besar sampel (data = n) yang diambil adalah 12 siswa laki-laki dari pedesaan dan 16 laki-laki dari perkotaan yang diambil secara acak. Siswa tersebut diberikan tes untuk mengukur Kesehatan mentalnya. Data disajikan sebagai berikut

Tabel ... skor Kesehatan mental pada siswa sekolah menengah pertama

Sekolah			
Perkotaan	Pedesaan	Perkotaan	Pedesaan
35	29	25	50
26	50	27	37
27	43	45	34
21	22	46	31
27	42	33	

Sekolah			
Perkotaan	Pedesaan	Perkotaan	Pedesaan
38	47	26	
23	42	46	
25	32	41	

Dengan taraf signifikansi sebesar 5%, apakah dapat disimpulkan bahwa ada perbedaan pada dua kelompok siswa tersebut?

b. Asumsi

Asumsi yang mendasari uji ini adalah sampel dipilih secara independen dan acak dari populasi masing-masing, populasi memiliki bentuk yang sama hanya berbeda dalam lokasi, variable yang diukur adalah kontinu. Tingkat pengukuran paling tidak harus ordinal. Kedua kelompok sampel tidak harus memiliki ukuran yang sama.

c. Hipotesis

$$H_0: M_u = M_R$$

$$H_a: M_u \neq M_R$$

MU adalah skor median populasi sampel siswa perkotaan, dan MR adalah skor median dari populasi sampel siswa pedesaan.

d. Statistik uji

Statistik uji yang digunakan adalah X^2

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

e. Distribusi dari statistik uji

Ketika H_0 benar, maka X^2 terdistribusi disekitar X^2 dengan $(r-1)$ $(c-1) = (2-1)(2-1) = (1)(1) = \text{degree of freedom (df)}$ atau derajat kebebasan adalah sebesar 1

f. Peraturan keputusan

Tolak H_0 jika nilai X^2 hasil perhitungan ≥ 3.841

d.f.	.995	.99	.975	.95	.9	.1	.05	.025	.01
1	0.00	0.00	0.00	0.00	0.02	2.71	3.84	5.02	6.63
2	0.01	0.02	0.05	0.10	0.21	4.61	5.99	7.38	9.21
3	0.07	0.11	0.22	0.35	0.58	6.25	7.81	9.35	11.34
4	0.21	0.30	0.48	0.71	1.06	7.78	9.49	11.14	13.28
5	0.41	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09

Gambar tabel kritis Chi-Square

g. Perhitungan statistik uji

Berikut prosedur perhitungan statistik uji

- 1) Gabungkan skor Kesehatan mental siswa pedesaan dan perkotaan
- 2) Urutkan skor dari terkecil hingga terbesar, kemudian gabungan skor ini hitunglah nilai tengahnya. Median dari skor keseluruhan data adalah $(33+34)/2 = 33.5$
- 3) Pisahkan skor yang berada di atas median untuk daerah perkotaan dan pedesaan demikian juga median bawah, gunakan bantuan tabel pada MS Excel

median	33.5	kelompok	kategori daerah	posisi median			urban	rural
	21	1	perkotaan	dibawah	above median		6	8
	22	2	pedesaan	dibawah	below median		10	4
	23	1	perkotaan	dibawah		0		
	25	1	perkotaan	dibawah		0		
	25	1	perkotaan	dibawah		0		
	26	1	perkotaan	dibawah		0		
	26	1	perkotaan	dibawah		0		
	27	1	perkotaan	dibawah		0		
	27	1	perkotaan	dibawah		0		
	29	2	pedesaan	dibawah		1		
	31	2	pedesaan	dibawah		1		
	32	2	pedesaan	dibawah		1		
	33	1	perkotaan	dibawah		0		
	34	2	pedesaan	diatas		1		
	35	1	perkotaan	diatas		0		
	37	2	pedesaan	diatas		1		

Tabel bantu di atas berguna untuk mengurutkan skor kesehatan mental dengan cepat (diurutkan dari terkecil ke terbesar), memberikan tanda daerah (perkotaan; pedesaan), dan posisi median (di atas; di bawah, *above*; *below* dari angka 33.5).

Selanjutnya, dari tabel di atas maka buatlah tabel kontingensi berikut

Tabel ... skor Kesehatan mental siswa laki-laki sekolah menengah pertama

	Perkotaan	Pedesaan	Total
Jumlah skor yang berada di atas median	6	8	14
Jumlah skor yang berada di bawah median	10	4	14
Total	16	12	28

Selanjutnya hitunglah nilai χ^2

$$\chi^2 = \sum_{i=1}^k \left[\frac{(O_i - E_i)^2}{E_i} \right] = \frac{(6 - 8)^2}{8} + \frac{(8 - 6)^2}{6} + \frac{(10 - 8)^2}{8} + \frac{(4 - 6)^2}{6} \\ = 2.33$$

- Keputusan uji
Karena $2.33 < 3.841$, maka gagal ditolak.
 - Kesimpulan
Tidak ada atau tidak terdapat perbedaan skor Kesehatan mental antara siswa yang sekolah di pedesaan maupun perkotaan, atau dengan kata lain, dua kelompok sampel diperkirakan diambil dari populasi dengan median yang sama
7. Mann whitney
- Merupakan salah satu tes nonparametric yang juga popular untuk membandingkan hasil diantara dua kelompok independen atau tidak berpasangan. Seringkali uji ini disebut tes mann whitney U atau uji whitney Wilcoxon atau uji jumlah peringkat Wilcoxon. Fungsi uji ini adalah untuk mengetahui apakah dua sampel cenderung berasal dari populasi yang sama (yakni apakah dua populasi memiliki bentuk yang sama). Beberapa pengamat bahkan menginterpretasikan tes ini sebagai perbandingan median diantara

dua populasi. Hipotesis penelitian dan hipotesis nol untuk uji dua sisi dinyatakan sebagai berikut

H_0 : kedua populasi adalah sama

H_1 : kedua populasi adalah tidak sama

Prosedur untuk uji ini melibatkan pengumpulan pengamatan dari dua sampel menjadi satu sampel yang digabung, kemudian diberikan peringkat terendah hingga tertinggi dari 1 hingga n. berikut adalah contoh kasus untuk uji mann whitney

- Pendekatan baru untuk perawatan prenatal diusulkan untuk ibu hamil yang tinggal di wilayah pedesaan. Program baru ini melibatkan kunjungan ke rumah selama kehamilan (kunjungan ini diluar jadwal pelayanan puskesmas setempat). Uji coba dilakukan dengan mengambil secara acak 15 wanita hamil. Kemudian dirancang evaluasi apakah Wanita yang berpartisipasi dalam program ini melahirkan bayi yang lebih sehat daripada Wanita yang tidak mengikuti program. Hasilnya adalah skor APGAR yang diukur setelah melahirkan dengan kisaran nilai 0 hingga 10, di mana skor 7 atau lebih tinggi dianggap normal (sehat), 4 – 6 rendah, dan 0 – 3 sangat rendah. Datanya tersaji sebagai berikut.

Perawatan biasa (rutin)	8	7	6	2	5	8	7	3
Perawatan program baru	9	8	7	8	10	9	6	

Apakah secara statistik terbukti perbedaan skor APGAR pada Wanita yang menerima perawatan baru dibandingkan dengan perawatan pada umumnya sebelum melahirkan?

- Merumuskan hipotesis dan menentukan tingkat kemaknaan

H_0 : kedua populasi adalah sama

H_1 : kedua populasi tidak sama

$\alpha = 0.05$

- Pemilihan uji statistik yang sesuai

Karena skor APGAR tidak terdistribusi secara normal dan sampel juga tergolong dalam kategori kecil maka digunakan statistik U.

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2$$

Di mana R_1 dan R_2 adalah jumlah peringkat dari masing-masing kelompok 1 dan kelompok 2.

d. Aturan keputusan

Nilai kritis yang sesuai dengan ukuran sampel ($n_1 = 8$ dan $n_2 = 7$) dan tingkat signifikansi dua sisi ($\alpha=0.05$) dapat dilihat pada gambar tabel di bawah ini.

Two-Sided Test $\alpha = 0.05$

n_2	n_1																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2								0	0	0	1	1	1	1	1	1	2	2	2	2
3				0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	
4			0	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	13	
5		0	1	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20	
6	1	2	3	5	6	8	10	11	13	14	16	17	19	21	22	24	25	27		
7	1	3	5	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34		
8	0	2	4	6	8	10	13	15	17	19	22	24	26	29	31	34	36	38	41	
9	0	2	4	7	10	12	15	17	20	23	26	28	31	34	37	39	42	45	48	
10	0	3	5	8	11	14	17	20	23	26	29	33	36	39	42	45	48	52	55	

Nilai kritis untuk uji ini adalah **10**, maka aturan keputusan adalah

tolak H_0 jika nilai $U \leq 10$

e. Hitung uji statistik

Langkah pertama adalah menetapkan peringkat 1 hingga 15 dengan mengurutkan nilai terkecil ke terbesar. Hal tersebut akan disajikan dalam tabel berikut.

TabelPenetapan Peringkat untuk Data Skor Perawatan

Keseluruhan sampel (yang Ranking
diurutkan dari angka
terkecil ke terbesar)

Perawatan biasa	Program baru	Perawatan biasa	Program baru	Perawatan biasa	Program baru
8	9	2		1	
7	8	3		2	

		Keseluruhan sampel (yang diurutkan dari angka terkecil ke terbesar)		Ranking	
Perawatan biasa	Program baru	Perawatan biasa	Program baru	Perawatan biasa	Program baru
6	7	5		3	
2	8	6	6	4.5	4.5
5	10	7	7	7	7
8	9	7		7	
7	6	8	8	10.5	10.5
3		8	8	10.5	10.5
			9		13.5
			9		13.5
			10		15

Dalam kelompok perawatan biasa jumlah peringkatnya adalah 45.5 dan dalam grup program baru jumlah *ranking*-nya adalah 74.5. Selanjutnya menghitung nilai U

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1 = 8(7) + \frac{8(9)}{2} - 45.5 = 46.5$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2 = 8(7) + \frac{7(8)}{2} - 74.5 = 9.5$$

Sehingga diperoleh statistik uji U bernilai 9.5

f. Kesimpulan

ditolak karena $9.5 < 10$. Secara statistik terbukti signifikan pada $\alpha = 5\%$ menunjukkan bahwa skor APGAR pada populasi tidak sama pada Wanita yang menerima perawatan prenatal biasa dibandingkan dengan program asuhan yang baru selama kehamilan.

8. Evaluasi

a. Seorang pelatih personal tertarik untuk membandingkan ambang anaerobic pada atlet. Ambang anaerobic diartikan sebagai titik di mana otot tidak dapat memperoleh lebih banyak oksigen untuk mempertahankan aktivitas atau batas atas aerobic saat Latihan. Ukuran ini juga berkaitan dengan detak

jantung maksimum. Tabel berikut adalah ambang anaerobic untuk pelari jarak jauh, pesepeda jarak jauh, perenang jarak jauh dan pemain ski lintas alam

Tabel 1 Ambang Anaerobik

Pelari jarak jauh	Pesepeda jarak jauh	Perenang jarak jauh	Pemain ski lintas alam
185	190	166	201
179	209	159	195
192	182	170	180
165	178	183	187
174	181	160	215

Apakah terdapat perbedaan ambang anaerobic diantara kelompok atlet yang berbeda? (gunakan $\alpha = 0.05$)

- b. Sebuah studi dijalankan untuk mengevaluasi efektivitas dari program Latihan yang diklaim mampu menurunkan tekanan darah sistolik pada pasien dengan kategori prehipertensi. Sampel diambil sejumlah 15 pasien prehipertensi, setiap pasien kemudian berpartisipasi dalam program Latihan olahraga di mana mereka mempelajari Teknik dan pelaksanaan yang tepat. Pasien diberi instruksi untuk melakukan program Latihan 3x per minggu selama 6 minggu. Setelah 6 minggu, tekanan darah sistolik diukur kembali. Data sebagai berikut

Tabel 2 Tekanan darah sistolik sebelum dan sesudah mengikuti program Latihan

pasien	Tekanan darah sistolik sebelum program Latihan	Tekanan darah sistolik sesudah program latihan
1	125	118
2	132	134
3	138	130
4	120	124
5	125	105
6	127	130

pasien	Tekanan darah sistolik sebelum program Latihan	Tekanan darah sistolik sesudah program latihan
7	136	130
8	139	132
9	131	123
10	132	128
11	135	126
12	136	140
13	128	135
14	127	126
15	130	132

Pertanyaan dari studi ini adalah apakah ada perbedaan tekanan darah sistolik setelah pasien berpartisipasi dalam program Latihan dibandingkan sebelum berpartisipasi? (tingkat signifikansi = 0.01)

- c. Lima belas catatan pada rekam medik pasien dari masing-masing dua rumah sakit ditinjau dan diberi skor yang dirancang untuk mengukur tingkat perawatan. Skor tersebut sebagai berikut.

Rumah sakit A	99	85	73	98	83	88	99	80	74	91	80	94	94	98	80
Rumah sakit B	78	74	69	79	57	78	79	68	59	91	89	55	60	55	79

Apakah dapat disimpulkan pada tingkat signifikansi bahwa kedua median berasal dari populasi yang berbeda?

- d. Sebuah uji klinis dijalankan untuk menilai efektivitas terapi antiretroviral baru untuk pasien dengan HIV. Pasien dipilih secara acak untuk menerima terapi antiretroviral standar dan terapi antiretroviral baru kemudian dipantau selama 3 bulan. Hasil utama diukur yakni jumlah Salinan HIV per militer darah. Hasil viral load (banyaknya virus di dalam darah seseorang yang telah terinfeksi HIV) 30 peserta acak dengan data sebagai berikut.

Terapi antiretroviral standar	7500	8000	2000	550
	1250	1000	2250	6800
	3400	6300	9100	970
	1040	670	400	
Terapi antiretroviral baru	400	250	800	1400
	8000	7400	1020	6000
	920	1420	2700	4200
	5200	4100	Tidak terdeteksi (undetectable)	

Apakah ada bukti secara statistik terdapat perbedaan viral load pada pasien yang menerima jenis terapi antiretroviral yang berbeda?

REFERENSI

- EJD, Agresti, A. and Finlay, B. (1998) *Statistical Methods for the Social Sciences, Journal of the American Statistical Association*. doi: 10.2307/2670147.
- shafer, douglas; zhang, Z. (2012) *Beginning Statistik*. Available at: <https://2012books.lardbucket.org/books/beginning-statistik/>.
- Spiegel, Murray; Schiller, John; Srinivasan, R. A. (2004) *Probabilitas dan Statistik*. ke dua. Jakarta: Penerbit Erlangga.
- Sullivan, L. M. (2018) *Essentials of Biostatistics in Public Health*. Third edit. Burlington, Massachusetts: Jones & Bartlett Learning.

BAGIAN 14

BESAR SAMPEL UNTUK ESTIMASI

1. Tujuan Pembelajaran

Setelah mengikuti materi ini, mahasiswa diharapkan untuk mampu menelaah (menghitung, menyajikan, dan menginterpretasikan, serta membandingkan) besar sampel untuk estimasi (proporsi dan rerata) serta besar sampel untuk uji hipotesis (beda proporsi, beda rerata, dan korelasi); mampu menjelaskan skema dan penggunaan sampling probabilitas dan nonprobabilitas; mampu menjelaskan penggunaan berbagai teknik sampling probabilitas; mampu menjelaskan penggunaan berbagai teknik sampling nonprobabilitas;

2. Besar sampel untuk Estimasi (proporsi dan rerata) dan sampel untuk uji hipotesis

a. Distribusi sampling rata-rata

Merupakan kumpulan dari bilangan-bilangan yang masing-masing merupakan rata-rata hitung dari sampelnya.

Rumus Distribusi Sampling Rata-rata:

$$\begin{array}{ll} \text{Populasi tidak terbatas} & \text{Populasi terbatas} \\ \frac{n}{N} \leq 5\% & \frac{n}{N} > 5\% \end{array}$$

Rata-rata

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

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

Standar Deviasi

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

Nilai Baku

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

- 1) Penentuan ukuran sampel untuk estimasi mean populasi Formula untuk populasi terhingga (terbatas)

$$n = \frac{N\sigma_x^2}{(N-1)\sigma_{\bar{x}}^2 + \sigma_x^2}$$

Formula untuk populasi tak terhingga

$$n = \frac{\sigma_x^2}{\sigma_{\bar{x}}^2}$$

Contoh soal 1

Sebuah perusahaan farmasi ingin melakukan uji pada kekuatan kapsul yang digunakan sebagai pembungkus obat serbuk/bubuk yang akan diedarkan sebagai jenis bahan kapsul baru. Dari beberapa studi sebelumnya, deviasi standar kekuatan kapsul mencapai 300 N/cm^2 . Direktur pemasaran obat menginginkan tingkat akurasi/estimasi kesalahan tidak melewati $\pm 75 \text{ N/cm}^2$ dengan tingkat kepercayaan 95%, tentukan ukuran sampel yang diperlukan untuk menguji kekuatan kapsul obat baru tersebut

Diketahui:

Tingkat akurasi/estimasi kesalahan = $E = 75 \text{ N/cm}^2$

Tingkat kepercayaan = 95%

Skor z untuk tingkat kepercayaan 95%, $z =$

tingkat kepercayaan	α	satu pihak	z	dua pihak	z
95%	5%	0.05	-1.644853627	0.025	-1.95996
90%	10%	0.1	-1.281551566	0.05	-1.64485
80%	20%	0.2	-0.841621234	0.1	-1.28155

Maka untuk kasus ini digunakan uji dua pihak (akan dijelaskan pada uji hipotesis) sehingga $z = 1,96$

Error standar dari mean sampling $\sigma_{\bar{x}} = \frac{E}{z} = \frac{75}{1,95996} = 38,266$

Asumsi deviasi standar adalah 300 N/cm^2 atau disebut σ_x . Maka besar sampel yang dibutuhkan untuk menguji kekuatan kapsul jenis baru ini adalah

$$n = \frac{\sigma_x^2}{\sigma_{\bar{x}}^2} = \frac{300^2}{38,266^2} = 61,46$$

Kesimpulan: dengan tingkat akurasi 95%, maka dibutuhkan 61 – 62 kapsul untuk diuji kekuatannya sebelum beredar pada masyarakat

- 2) Penentuan ukuran sampel untuk estimasi proporsi populasi
Formula untuk populasi tak terhingga

$$n = \frac{\pi(100 - \pi)}{\sigma_p^2}$$

Formula untuk populasi terhingga (terbatas)

$$n = \frac{N\pi(100 - \pi)}{(N - 1)\sigma_p^2 + \pi(100 - \pi)}$$

Contoh soal 2

Fakultas Kesehatan masyarakat hendak menyelenggarakan seminar nasional dalam rangka Dies Natalis FKM tahun 2019, pihak akademik ingin memperkirakan berapa banyak mahasiswa yang berminat untuk berpartisipasi sebagai peserta pada seminar ini. Akademik menginginkan agar perkiraan tersebut berada dalam kisaran $\pm 5\%$, dengan tingkat kepercayaan 95%. Berapa banyak peminatnya?

Diketahui Tingkat akurasi/estimasi kesalahan = $E = 5$

Tingkat kepercayaan estimasi = 95%

Skor z untuk tingkat kepercayaan 95%, $z = 1,96$

Error standar dari mean sampling, $\sigma_p = \frac{E}{z} = \frac{5}{1,96} = 2,55$

Karena perkiraan persentase populasi tidak disebutkan, maka diasumsikan $\pi = 50$

Maka, banyaknya partisipan yang diperkirakan mencapai

$$n = \frac{\pi(100 - \pi)}{\sigma_p^2} = \frac{50(100 - 50)}{2,55^2} = 384,14$$

Kesimpulan: kemungkinan partisipan seminar nasional mencapai angka 384 – 385 mahasiswa

- b. Distribusi sampling proporsi

Merupakan kumpulan atau distribusi semua perbandingan sampelnya untuk suatu peristiwa.

Rumus Distribusi Sampling Proporsi	
Populasi tidak terbatas	Populasi terbatas
$(\frac{n}{N} \leq 5\%)$	$(\frac{n}{N} > 5\%)$
Rata-rata	$\mu_{\bar{x}} = \pi$
Standar Deviasi	$\sigma_{\bar{x}} = \sqrt{\frac{\pi(1-\pi)}{n}}$
Nilai Baku	$z = \frac{\frac{\bar{x}}{n} - \frac{\mu_x}{n}}{\frac{\sigma_{\bar{x}}}{\sqrt{n}}}$
	$z = \frac{\frac{\bar{x}}{n} - \frac{\mu_x}{n}}{\frac{\sigma_{\bar{x}}}{\sqrt{n-1}}}$

- c. Persoalan satu sampel meliputi pendugaan rata-rata populasi dan pengujian hipotesis rata-rata satu populasi

Menentukan ukuran sampel untuk memperkirakan dan menguji hipotesis terkait rata-rata populasi. Ketika mean dinyatakan akan tetapi tidak diketahui di dalam populasi dan varians yang tidak diketahui, serta didefinisikan sebagai jarak dari rata-rata populasi yang sebenarnya (Lemeshow, 1997)

Contoh kasus adalah ingin diketahui harga ecer sebuah obat yang dijual di apotek di sebuah daerah. Diperkirakan penduga dengan jarak 10 sen dari rata-rata harga yang sebenarnya dengan tingkat kepercayaan 95%. Pada penelitian terdahulu, simpangan baku harga eceran diperkirakan sebesar 85 sen. Berapa apotek yang harus dipilih untuk mengetahui harga ecer obat tersebut?

$$n = \frac{z_{1-\alpha/2}^2 \sigma^2}{d^2}$$

Dengan nilai alpa = 5% maka diperoleh nilai z adalah 1,96 dengan nilai d adalah 10 sen atau 0,1

Dari selisih jarak rerata populasi maka jumlah apotek yang dibutuhkan untuk studi ini adalah 277

d. Uji hipotesis rerata 1 populasi

Ukuran sampel yang diperlukan untuk situasi pengujian hipotesis sampel tunggal adalah pada formula berikut

$$n = \frac{\sigma^2 [z_{1-\alpha} + z_{1-\beta}]^2}{[\mu_0 - \mu_a]^2}$$

Contoh kasus sebagai berikut

Berdasarkan hasil survei berat rata-rata pria di atas 55 tahun dengan penyakit jantung adalah 90 Kg. namun muncul dugaan, bahwa berat rata-rata pria dengan kondisi demikian sedikit lebih rendah. Berapa besar sampel yang diperlukan untuk menguji signifikansi 5 % dengan kekuatan 90%? Apakah bobot rata-rata tidak berubah? Atau dari 90 Kg menjadi 85 Kg dengan perkiraan standar deviasi 20 Kg?

standar deviasi	σ	20	400	
1 - alpa	α	0.05	0.95	
z 1 - alpa		1.644854		
rerata populasi	μ_0	90		$n = \frac{\sigma^2 [z_{1-\alpha} + z_{1-\beta}]^2}{[\mu_0 - \mu_a]^2}$
dugaan rerata populasi	μ_a	85		
beta		10	0.1	
1 - beta		0.9		
z 1 - beta		1.281552		
n		8.563847		
		3425.539		
		25	137.0216	

Besar sampel yang dibutuhkan untuk membuktikan dugaan bahwa pria dengan usia 55 tahun memiliki penyakit jantung adalah 137 atau 138 pasien

e. Persoalan dua sampel

Untuk memperkirakan perbedaan antara dua rata-rata populasi dan pada pengujian hipotesis dalam dua kelompok.

Terdapat selisih antara dua mean (rerata) populasi.

Ahli gizi memperkirakan perbedaan asupan kalori saat makan siang antara anak-anak yang sekolah pada program makan siang panas di sekolah dan anak-anak yang tidak mengikuti program tersebut. Studi lain memperkirakan terdapat standar penyimpangan asupan kalori pada anak sekolah dasar adalah 75 kalori, dan terdapat perbedaan kalori sebanyak 20 kalori dari populasi dengan kepercayaan 95%.

alpa	α	0.05				
	alpa / 2	0.025				
	1 - alpa/2	0.975				
z 1 - alpa/2		1.959964				
stdev	σ	75	5625	11250		
d		20				
pembilang		3.841459	43216.41			
penyebut			400			
n				108.041		

$$n = \frac{z_{1-\alpha/2}^2 [2\sigma^2]}{d^2}$$

Sebanyak 109 siswa setiap sekolah dapat dilibatkan jika terdapat kasus studi membandingkan asupan kalori pada saat makan siang di sekolah

- f. Uji hipotesis untuk rerata dua populasi

$$n = \frac{2\sigma^2 [z_{1-\alpha/2} + z_{1-\beta}]^2}{(\mu_1 - \mu_2)^2}$$

Sebuah penelitian dirancang untuk mengetahui apakah menurunkan asupan natrium dalam makanan dapat memberikan efek pada tekanan darah sistolik. Dari studi percontohan, diketahui bahwa standar deviasi tekanan darah sistolik dalam kelompok dengan natrium tinggi adalah 12 mmHg sedangkan pada kelompok diet rendah natrium adalah 10,3 mm Hg. Jika alpa dan beta masing-masing adalah 0,05 dan 0,10, Berapa sampel yang dibutuhkan dari setiap kelompok jika ingin mendeteksi perbedaan tekanan darah sebesar 2 mmHg antara dua kelompok

standar deviasi	σ	125.045	250.09					
alpa	α	0.05						
alpa / 2		0.025						
1 - alpa /2		0.975						
Z1 - alpa/2		1.959964						
beta	β	0.1						
1 - beta		0.9						
Z 1 - beta		1.281552						
μ_1								
μ_2								
$\mu_1 - \mu_2$		2						
pembilang		10.50742	2627.801					
penyebut				4				
					656.9504			

Berdasarkan perhitungan, 658 akan dibutuhkan dari masing-masing kelompok (diet rendah dan diet sedang natrium)

3. Skema dan penggunaan Sampling probabilitas dan non probabilitas

Dalam suatu penelitian survei, tidak perlu untuk meneliti semua individu dalam suatu populasi, sebab di samping memakan **biaya** yang banyak, juga membutuhkan **waktu** yang lama. Dengan meneliti sebagian dari populasi, diharapkan hasil yang diperoleh akan dapat menggambarkan sifat populasi yang bersangkutan. Untuk dapat mencapai tujuan ini, maka cara-cara pengambilan sebuah sampel harus memenuhi syarat-syarat tertentu. Sebuah sampel harus dipilih sedemikian rupa sehingga setiap satuan unsur **mempunyai kesempatan dan peluang yang sama untuk dipilih**, dan besarnya peluang tersebut tidak boleh sama dengan nol. Di samping itu, pengambilan sampel secara acak (random) harus menggunakan teknik yang tepat sesuai dengan ciri-ciri populasi dan tujuan penelitian.

Dalam menentukan teknik pengambilan sampel yang akan diterapkan dalam suatu penelitian, seorang peneliti harus memperhatikan hubungan antara biaya, tenaga, dan waktu di satu pihak, serta tingkat presisi di pihak lain. Jika jumlah biaya, tenaga, dan waktu sudah dibatasi sejak semula, seorang peneliti harus berusaha mendapatkan teknik pengambilan sampel yang menghasilkan presisi tertinggi. Perlu disadari bahwa tingkat presisi yang tinggi tidak mungkin dapat dicapai dengan biaya, tenaga, dan **waktu** yang terbatas. Di samping itu, perlu diperhatikan pula

masalah “efisiensi” dalam memilih teknik pengambilan sampel.

Setelah membahas berapa besar sampel yang ideal untuk berbagai kondisi, dalam tahapan berikutnya diperlukan teknis atau prosedur di lapangan bagaimana mengambil sampel dari populasi yang telah tersedia. Kondisi yang berbeda akan menyebabkan kebutuhan akan pengambilan sampel juga berbeda. Terdapat dua Teknik umum yang dapat digunakan dalam penelitian observasi ataupun eksperimen. Yakni sampling probabilitas dan non probabilitas. Kedua Teknik ini juga banyak digunakan pada penelitian Kesehatan khususnya Kesehatan masyarakat.

Sampling probabilitas diperuntukkan untuk populasi yang memiliki sifat homogen. Seluruh populasi yang menjadi sasaran penelitian akan memiliki peluang yang sama untuk diambil menjadi bagian dari sampel ataupun subjek penelitian. Selain itu, sampling probabilitas juga membutuhkan daftar populasi ataupun daftar unit (klaster). Salah satu daftar yang dapat digunakan misalnya adalah kartu keluarga yang berada pada Rukun Tetangga. Data ini bisa menjadi dasar populasi yang *eligible* untuk sebuah survei.

Pada penelitian dengan desain tertentu, sampling probabilitas bisa jadi bukan menjadi pilihan dalam Teknik sampling. Sehingga para peneliti atau akademisi dapat memilih sampling non probabilitas. Jika dihadapi kondisi tingkat respons yang sangat rendah, seperti kasus penyakit baru atau langka, sampling jenis ini juga bisa menjadi pilihan (Fu, 2016)

4. Jenis Teknik sampling probabilitas

Secara umum, dalam penelitian sosial dan Kesehatan terdapat 4 teknik sampling probabilitas yang dapat digunakan yakni simple random sampling, *systematic sampling*, *stratified sampling* dan *cluster sampling*. Untuk *cluster sampling*, beberapa penelitian bahkan menggunakan *two stage cluster sampling* hingga *multistage sampling*. Sub bab berikut ini akan menjelaskan teori probabilitas dan bagaimana contoh penerapan di lapangan serta kemungkinan

kendala dan hambatan yang diuraikan berdasarkan pengalaman penulis dan mahasiswa yang telah melalui proses penelitian.

a. Simple random sampling

Dinyatakan simple (sederhana) karena pengambilan sampel anggota populasi dilakukan secara acak tanpa memperhatikan strata yang ada dalam populasi itu. Margono, (2004) menyatakan bahwa simple random sampling adalah teknik untuk mendapatkan sampel yang langsung dilakukan pada unit sampling. Dengan demikian setiap unit sampling sebagai unsur populasi yang terpencil memperoleh peluang yang sama untuk menjadi sampel atau untuk mewakili populasi. Cara demikian dilakukan bila anggota populasi dianggap homogen. Teknik ini dapat dipergunakan bilamana jumlah unit sampling di dalam suatu populasi tidak terlalu besar. Misal, populasi terdiri dari 500 orang mahasiswa program S1 (unit tampling).

Dalam rangka memilih suatu sampel acak sederhana, perlu dilakukan hal sebagai berikut

- 1) Membuat daftar (atau kerangka) dari n unit penghitungan (*frame sampling*)
- 2) Menggunakan proses acak (misalnya dengan tabel bilangan acak) untuk mendapatkan n angka antara 1 dan n yang akan menetapkan n individu mana yang terpilih ke dalam sampel

Definisi

Sampel acak sederhana adalah sampel di mana setiap sampel dari sejumlah NC_n sampel yang mungkin mempunyai kesempatan yang sama untuk terpilih, yaitu $1/NC_n$

Keuntungan pengambilan sampel acak sederhana adalah sebagai berikut

- 1) mudah dipahami
- 2) cara ini memberikan dasar probabilistik terhadap banyak teori statistik
- 3) cara ini memberikan dasar perbandingan bagi metode lainnya

Kelemahan pengambilan sampel acak sederhana adalah

- 1) seluruh N unit penghitungan dalam populasi harus ditetapkan dan diberi nomor sebelum pengambilan sampel dilakukan. proses ini kemungkinan besar sangat mahal dan makan waktu sehingga menjadi tidak realistik untuk melaksanakannya dalam praktik
- 2) individu yang terpilih mungkin sangat tersebar. ini mengakibatkan kunjungan kepada tiap individu yang terpilih mungkin merupakan proses yang sangat banyak makan waktu dan mahal
- 3) sub-klaster tertentu dalam populasi, secara kebetulan mungkin terpilih semua ke dalam sampel

b. Systematic sampling

Metode ini dapat menghemat waktu dan usaha dan dalam beberapa situasi lebih efisien daripada pengambilan sampel acak sederhana. Sebagai contoh, misalnya sebuah sampel berukuran n , dari kartu rekam medis pasien yang dirawat sepanjang tahun lalu di suatu balai pengobatan puskesmas, diperlukan suatu untuk survei pemasukan gizi. Dengan sampel sistematik, ini diperoleh dengan cara membuat n daerah yang masing-masing terdiri dari k rekam medis. Di dalam daerah pertama, pilih satu angka acak antara 1 dan k yang merupakan unsur kartu pertama yang terpilih. Rekam medis berikutnya ditentukan dengan secara berturut-turut menambahkan konstanta k kepada angka acak awal I jadi sampel adalah terdiri dari kartu ke- i , ke $(i+k)$, disajikan sebagai berikut

- 1) misalnya populasi (N) 100 dan sampel (n) 80, maka .
- 2) nomor sampel acak pertama adalah 1,25, selanjutnya sebagai berikut
- 3) (angka acak berada diantara 1 dan 1,25 (k) sehingga $i = 1$)
- 4) $I + 3 k = 1 + 3 \times 1,25 = 4,75$; $I + 4 k = 1 + 4 * 1,25$

k	i	1	n	80
1	1,25	2,25	i + k	
2	1,25	3,5	i + 2k	populasi 100,00
3	1,25	4,75	i + 3k	d 0,05
4	1,25	6	i + 4k	d^2 1,25
5	1,25	7,25		n 80,00
6	1,25	8,5		k 1,25
7	1,25	9,75		
8	1,25	11		angka acak 1
9	1,25	12,25		
10	1,25	13,5		
11	1,25	14,75		
12	1,25	16		
13	1,25	17,25		
14	1,25	18,5		
15	1,25	19,75		
16	1,25	21		
17	1,25	22,25		
18	1,25	23,5		
19	1,25	24,75		
44	1,25	56		
45	1,25	57,25		
46	1,25	58,5		
47	1,25	59,75		
48	1,25	61		
49	1,25	62,25		
50	1,25	63,5	i + (n-1) k	

c. Stratified sampling

Dalam praktik sering dijumpai populasi yang tidak homogen. Makin heterogen suatu populasi, makin besar pula perbedaan sifat antara lapisan-lapisan tersebut. Presisi dan hasil yang dapat dicapai dengan penggunaan suatu metode pengambilan sampel, antara lain dipengaruhi oleh derajat keseragaman dari populasi yang bersangkutan. Untuk dapat menggambarkan secara tepat mengenai sifat-sifat populasi yang heterogen, maka populasi yang bersangkutan dibagi ke dalam lapisan-lapisan (stratum) yang seragam dan dari setiap lapisan diambil sampel secara acak. Dalam sampel berlapis, peluang untuk terpilih satu

strata dengan yang lain mungkin sama, mungkin pula berbeda (proporsional atau non proporsional).

Ada dua syarat yang harus terpenuhi untuk dapat mempergunakan metode pengambilan sampel acak berlapis, yaitu (a) ada kriteria jelas yang akan dipergunakan sebagai dasar untuk menstratifikasi populasi, dan (b) diketahui dengan tepat jumlah satuan-satuan elementer dari tiap lapisan dalam populasi itu.

Besarnya sampel yang diambil dari tiap-tiap strata dapat berimbang dan dapat pula tidak berimbang. Dalam pengambilan sampel yang berimbang, unsur-unsur satuan yang diambil dari setiap strata berbanding lurus dengan jumlah satuan-satuan elementer dalam strata yang bersangkutan. Kalau peneliti akan mempergunakan metode tidak berimbang, maka ia dapat menentukan sendiri jumlah unsur-unsur sampel yang akan diambilnya.

Stratified random sampling biasa digunakan pada populasi yang mempunyai susunan bertingkat atau berlapis-lapis. Teknik ini digunakan bila populasi mempunyai anggota/unsur yang tidak homogen dan berstrata secara proporsional.

Suatu organisasi yang mempunyai pegawai dari berbagai latar belakang pendidikan, maka populasi pegawai itu berstrata. Misalnya jumlah pegawai yang lulus S1 = 45, S2 = 30, STM = 800, ST = 900, SMEA = 400, SD = 300. Jumlah sampel yang harus diambil meliputi strata pendidikan tersebut yang diambil secara proporsional jumlah sampel.

d. Cluster sampling

Teknik ini disebut juga *cluster random sampling*. Teknik ini digunakan bilamana populasi tidak terdiri dari individu-individu, melainkan terdiri dari kelompok-kelompok individu atau *cluster*.

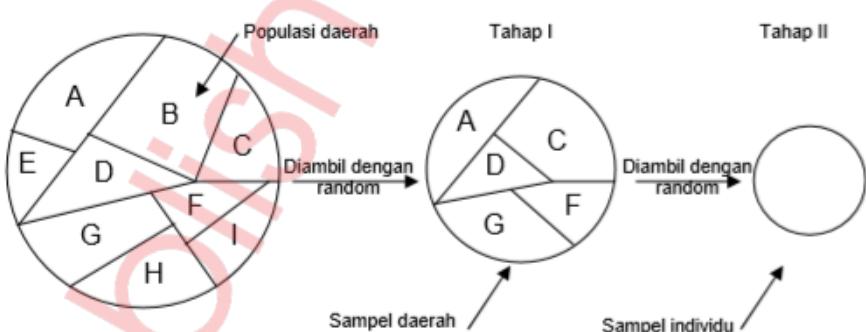
Teknik sampling daerah digunakan untuk menentukan sampel bila objek yang akan diteliti atau sumber data sangat luas, misalnya penduduk dari suatu negara, propinsi atau

kabupaten. Untuk menentukan penduduk mana yang akan dijadikan sumber data, maka pengambilan sampelnya berdasarkan daerah populasi yang telah ditetapkan.

sebagai contoh, di Indonesia terdapat 27 propinsi, dan sampelnya akan menggunakan 10 propinsi, maka pengambilan 10 propinsi itu dilakukan secara random. Tetapi perlu diingat, karena propinsi-propinsi di Indonesia itu berstrata maka pengambilan sampelnya perlu menggunakan *stratified random sampling*.

Contoh lainnya dikemukakan bila penelitian dilakukan terhadap populasi pelajar SMU di suatu kota. Untuk random tidak dilakukan langsung pada semua pelajar-pelajar, tetapi pada sekolah/kelas sebagai kelompok atau *cluster*.

Teknik sampling daerah ini sering digunakan melalui dua tahap, yaitu tahap pertama menentukan sampel daerah, dan tahap berikutnya menentukan orang-orang yang ada pada daerah itu secara sampling juga. Teknik ini dapat diilustrasikan di bawah ini.



5. Jenis Teknik sampling non-probabilitas

Sampling jenis ini akan dibutuhkan pada kondisi-kondisi tertentu. Salah satunya dipaparkan pada sub bab sebelumnya, bahwa jika terjadi kasus penyakit yang langka ataupun prinsip random tidak dapat ditegakkan dengan maksimal. Pemilihan

teknis sampling di lapangan tergantung pada beberapa kondisi diantaranya yakni waktu, biaya dan tenaga yang tersedia serta data yang ada untuk digunakan sebagai dasar pengambilan sampel. Beberapa ahli mengatakan bahwa, prinsip random merupakan salah satu syarat penggunaan statistik inferensial, jika demikian, apakah berarti jika pemilihan sampel non random dilarang mutlak menggunakan statistik inferensial? Tentu tidak, silahkan digunakan akan tetapi kesimpulan yang ditarik pada akhir laporan hanyalah berlaku untuk keseluruhan sampel, bukan populasi.

Berikut adalah beberapa jenis Teknik sampling non probabilitas

a. Sampling kuota

Menyatakan bahwa sampling kuota adalah teknik untuk menentukan sampel dari populasi yang mempunyai ciri-ciri tertentu sampai jumlah (kuota) yang diinginkan.

Dalam teknik ini jumlah populasi tidak diperhitungkan akan tetapi diklasifikasikan dalam beberapa kelompok. Sampel diambil dengan memberikan jatah atau quorum tertentu terhadap kelompok.

Pengumpulan data dilakukan langsung pada unit sampling. Setelah jatah terpenuhi, pengumpulan data dihentikan. Sebagai contoh, akan melakukan penelitian terhadap pegawai golongan II, dan penelitian dilakukan secara kelompok. Setelah jumlah sampel ditentukan 100, dan jumlah anggota peneliti berjumlah 5 orang, maka setiap anggota peneliti dapat memilih sampel secara bebas sesuai dengan karakteristik yang ditentukan (golongan II) sebanyak 20 orang.

b. Sampling aksidental

Sampling aksidental adalah teknik penentuan sampel berdasarkan kebetulan, yaitu siapa saja yang secara kebetulan bertemu dengan peneliti dapat digunakan sebagai sampel, bila dipandang orang yang kebetulan ditemui itu cocok sebagai sumber data.

Dalam teknik ini pengambilan sampel tidak ditetapkan lebih dahulu. Peneliti langsung mengumpulkan data dari unit

sampling yang ditemui. Misalnya penelitian tentang pendapat umum mengenai pemilu dengan mempergunakan setiap warga negara yang telah dewasa sebagai unit sampling. Peneliti mengumpulkan data langsung dari setiap orang dewasa yang dijumpainya, sampai jumlah yang diharapkan terpenuhi.

c. Sampling Purposive

Sampling purposive adalah teknik penentuan sampel dengan pertimbangan tertentu.

Pemilihan sekelompok subjek dalam *purposive sampling*, didasarkan atas ciri-ciri tertentu yang dipandang mempunyai sangkut paut yang erat dengan ciri-ciri populasi yang sudah diketahui sebelumnya. Dengan kata lain unit sampel yang dihubungi disesuaikan dengan kriteria-kriteria tertentu yang diterapkan berdasarkan tujuan penelitian. Misalnya akan melakukan penelitian tentang disiplin pegawai, maka sampel yang dipilih adalah orang yang ahli dalam bidang kepegawaian saja.

d. Sampling Jenuh

Sampling jenuh adalah teknik penentuan sampel bila semua anggota populasi ~~digunakan~~ sebagai sampel. Hal ini sering dilakukan bila jumlah populasi relatif kecil, kurang dari 30 orang. Istilah lain sampel jenuh adalah sensus, di mana semua anggota populasi dijadikan sampel.

e. Snowball Sampling

Teknik penentuan sampel yang mula-mula jumlahnya kecil, kemudian sampel ini disuruh memilih teman-temannya untuk dijadikan sampel. Begitu seterusnya, sehingga jumlah sampel semakin banyak.

Ibarat bola salju yang menggelinding, makin lama semakin besar. Pada penelitian kualitatif banyak menggunakan sampel *purposive* dan *snowball*.

6. Evaluasi Formatif

- a. Wilayah generalisasi yang terdiri atas objek/subjek yang mempunyai kuantitas dan karakteristik tertentu yang ditetapkan oleh peneliti untuk dipelajari dan kemudian ditarik kesimpulannya disebut...
 - a. Populasi
 - b. Sampel
 - c. Hipotesis
 - d. Subjek penelitian
- b. Suatu populasi dengan jumlah individu dalam kelompok tidak mempunyai jumlah yang tetap, ataupun jumlahnya tidak terhingga, disebut populasi...
 - a. Finit
 - b. Infinit
 - c. Homogen
 - d. Heterogen
- c. Populasi yang tidak dapat ditemukan batas-batasnya, sehingga tidak dapat dinyatakan dalam bentuk jumlah secara kuantitatif disebut populasi...
 - a. Terhingga
 - b. Tak terhingga
 - c. Homogen
 - d. Heterogen
- d. Populasi yang unsur-unsurnya memiliki sifat atau keadaan yang bervariasi, sehingga perlu ditetapkan batas-batasnya, baik secara kualitatif maupun secara kuantitatif disebut populasi...
 - a. Terhingga
 - b. Tak terhingga
 - c. Homogen
 - d. Heterogen
- e. Sejumlah populasi yang secara kuantitatif dapat dinyatakan dengan tegas disebut populasi...
 - a. Terhingga
 - b. Tak terhingga
 - c. Populasi yang tersedia
 - d. heterogen

- f. Bagian dari populasi, sebagai contoh yang diambil dengan menggunakan cara-cara tertentu disebut...
- Populasi
 - Sampel
 - Hipotesis
 - Finit
- g. Berikut ini yang menjadi penyebab diperlukannya sampel adalah...
- Populasi terlalu kecil.
 - Penelitian bermaksud mengadakan generalisasi dari hasil-hasil kepenelitiannya, dalam arti mengenakan kesimpulan-kesimpulan kepada objek, gejala, atau kejadian yang lebih luas
 - Tuntutan metode penelitian
 - Keterbatasan kemampuan peneliti
- h. Berikut ini yang menjadi alasan penggunaan sampel, kecuali...
- Ukuran populasi.
 - Masalah biaya
 - Karena peneliti menginginkannya
 - Masalah ketelitian.
- i. Jika diketahui populasi penelitian adalah 5.000.000 orang guru SMA pada awal tahun 1985, dengan karakteristik; masa kerja 2 tahun, lulusan program Strata 1. Populasi yang demikian disebut populasi...
- Infinit
 - Tak terbatas.
 - Terbatas
 - Homogen.
- j. Accessible population adalah...
- Sejumlah populasi yang secara kuantitatif dapat dinyatakan dengan tegas
 - Populasi yang tidak dapat ditemukan batas-batasnya, sehingga tidak dapat dinyatakan dalam bentuk jumlah secara kuantitatif.

- c. Populasi yang unsur-unsurnya memiliki sifat yang sama, sehingga tidak perlu dipersoalkan jumlahnya secara kuantitatif.
- d. Populasi yang unsur-unsurnya memiliki sifat atau keadaan yang bervariasi, sehingga perlu ditetapkan batas-batasnya, baik secara kualitatif maupun secara kuantitatif.
- k. Cara untuk menentukan sampel yang jumlahnya sesuai dengan ukuran sampel yang akan dijadikan sumber data sebenarnya, dengan memperhatikan sifat-sifat dan penyebaran populasi agar diperoleh sampel yang representatif disebut...
 - a. Sampling
 - b. Teknik sampling
 - c. Populasi
 - d. Sampel
- l. Teknik sampling yang memberikan peluang yang sama bagi setiap unsur (anggota) populasi untuk dipilih menjadi anggota sampel disebut...
 - a. Probability sampling
 - b. Nonprobability sampling
 - c. Aksidental sampling
 - d. Purposive sampling
- m. Teknik yang tidak memberi peluang/kesempatan yang sama bagi setiap unsur atau anggota populasi untuk dipilih menjadi sampel disebut...
 - a. Probability sampling
 - b. Nonprobability sampling
 - c. Aksidental sampling
 - d. Purposive sampling
- n. Berikut ini yang termasuk ke dalam teknik probability sampling, kecuali...
 - a. simple random sampling
 - b. disproportionate stratified random sampling
 - c. sampling kuota
 - d. proportionate stratified random sampling

- o. Berikut ini yang termasuk pada teknik non probability sampling, kecuali...
 - a. area (cluster) sampling (sampling menurut daerah).
 - b. purposive sampling
 - c. snowball sampling
 - d. sampling jenuh
- p. adalah teknik penentuan sampel berdasarkan urutan dari anggota populasi yang telah diberi nomor urut disebut...
 - a. area (cluster) sampling
 - b. purposive sampling
 - c. snowball sampling
 - d. sampling sistematis
- q. Pengambilan sampel anggota populasi dilakukan secara acak tanpa memperhatikan strata yang ada dalam populasi itu disebut...
 - a. simple random sampling
 - b. disproportionate stratified random sampling
 - c. sampling kuota
 - d. proportionate stratified random sampling
- r. Teknik penentuan sampel yang mula-mula jumlahnya kecil, kemudian sampel ini disuruh memilih teman-temannya untuk dijadikan sampel disebut...
 - a. area (cluster) sampling
 - b. purposive sampling
 - c. snowball sampling
 - d. sampling sistematis
- s. Teknik penentuan sampel bila semua anggota populasi digunakan sebagai sampel disebut...
 - a. area (cluster) sampling
 - b. purposive sampling
 - c. snowball sampling
 - d. sampling jenuh.
- t. Teknik penentuan sampel berdasarkan kebetulan, yaitu siapa

saja yang secara kebetulan bertemu dengan peneliti dapat digunakan sebagai sampel, bila dipandang orang yang kebetulan ditemui itu cocok sebagai sumber data disebut...

- a. area (cluster) sampling
- b. purposive sampling
- c. snowball sampling
- d. Aksidental sampling

Referensi

- Fu, C. W. D. J. T. S. Y. (2016). *Handbook of Survey Methodology* (Issue January). SAGE Publications Ltd.
- Lemeshow, H. & K. (1997). *Besar sampel dalam penelitian kesehatan*. Gajah Mada University Press.
- Margono. (2004). *Metodologi penelitian pendidikan*. Rineka Cipta.