Tedi Hartoyo
Bambang Suhartawan
Deasy Wahyuni

Get Press Indonesia

PENGANTAR METODE STATISTIK

ISBN 978-634-255-022-9
9 786342 550229

PENGANTAR METODE STATISTIK

Tedi Hartoyo
Bambang Suhartawan
Deasy Wahyuni

GETPRESS INDONESIA

PENGANTAR METODE STATISTIK

Penulis :
Tedi Hartoyo
Bambang Suhartawan
Deasy Wahyuni

ISBN : 978-634-255-022-9

Editor : Mila Sari, M. Si.
Desain Sampul dan Tata Letak : Tri Putri Wahyuni, S. Pd.

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

Cetakan pertama, September 2025

KATA PENGANTAR

Puji dan syukur kami panjatkan ke hadirat Tuhan Yang Maha Esa, yang telah memberikan rahmat dan karunia-Nya sehingga buku berjudul *Pengantar Metode Statistik* ini dapat diselesaikan. Isi buku ini mencakup pembahasan tentang konsep dasar statistika, penyajian dan pengolahan data, probabilitas dan distribusi peluang, pengambilan sampel dan distribusi sampling, dan uji hipotesis statistik. Kami berharap semoga buku ini dapat menjadi referensi dan memberikan manfaat serta inspirasi bagi semua pembaca.

Padang, Agustus 2025

Penulis

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

BAB 1
KONSEP DASAR STATISTIKA

Oleh Tedi Hartoyo

1.1 Pendahuluan

Kegiatan penelitian wajib dilakukan oleh seorang mahasiswa di tingkat akhir dan peneliti pastinya membutuhkan data-data dari hasil penelitiannya, disebut dengan data observasi. Biasanya untuk mengambil suatu keputusan dalam penelitian, data hasil observasi dianalisis dengan menggunakan alat bantu yaitu ilmu statistik (Statistika).

Gomez and Gomez (1984) menyatakan, bahwa Ilmu Statistik atau tepatnya Metode Statistika adalah ilmu yang mempelajari cara-cara untuk mengumpulkan, mengatur, mengolah, menyajikan, menganalisis, serta akhirnya interpretasi yaitu menarik kesimpulan tentang data (bahan keterangan) yang berupa angka-angka, berdasarkan metodologi tertentu. Kegunaan sangat luas, mulai dari penelitian ilmiah, pengambilan keputusan bisnis, hingga pemantauan kebijakan publik. Statristik dasar digunakan untuk memahami data, mengidentifikasi pola, dan menarik kesimpulan yang valid.

Dengan mengetahui dasar-dasar statistika dan kegunaan statistika, seseorang dapat menggunakan data secara efektif untuk memahami fenomena di sekitarnya, membuat keputusan yang tepat, dan mengambil tindakan yang efektif.

Selanjutnya statistika adalah cabang ilmu matematika terapan yang terdiri dari teori dan metoda mengenai bagaimana cara mengumpulkan, mengukur, mengklarifikasi, menghitung, menjelaskan, mensisntesis, menganalisis, dan menafsirkan data yang diperoleh secara matematis. Dengan demikian, didalamnya terdiri sekumpulan prosedeur mengenai bagaimana cara:

1) Mengumpulkan data
2) Meringkas data
3) Mengolah data
4) Menyajikan data
5) Menarik kesimpula dan interpretasi data berdasarkan kumpulan data hasil analisisnya.

Statistika memiliki kegunaan yang sangat luas, mulai dari merangkum data, menguji hipotesis, hingga membuat prediksi. Kegunaan ini memungkinkan pengambilan keputusan yang tepat, seperti di bidang pertanian, ekonomi (bisnis), kesehatan, Pendidikan dan social (Gomez dan Gomez, 1984). Kegunaan statistik lebih mendetil lagi dimanfaatkan untuk:

1) Merangkum dan Menganalisis Data
2) Menguji Hipotesi
3) Memprediksi
4) Analisis Data
5) Pengambilan Keputusan
6) Pemantauan Kinerja
7) Kualiti Konbtrol
8) Pertanian
9) Ekonomi
10) Pendidikan
11) Kesehatan

Juga akan dibahas mengenai pengertian dan perbedaan:

1) Statistik dan Parameter
2) Populasi dan sampel
3) Metode statistik
4) Statistik deskriptif dan inferensia
5) Statistika parametrik dan statistika non parametrik
6) Statistik univariat dan multivariat

Selanjutnya dalam menganalisis data (analisis statistik), ketepatan pemilihan metode uji adalah sangat penting untuk mendapat kesimpulan yang benar atau valid. Dalam dunia penelitian dua pendekatan utama untuk menganalisis data adalah Metode Uji Statistika Parametrik dan Metode Uji Statistika Non-Parametrik. Kedua metode ini memiliki prinsip, asumsi, serta kelebihan dan kekurangan yang berbeda, sehingga memahami dulu karakteristik metodenya merupakan langkah awal sebelum menjalankannya dalam penelitian.

Uji Parametrik (Statistika Parametrik) dan Uji Non Parametrik (Statistika Non Parametrik) mempunyai beberapa perbedaan. Uji Parametrik secara umum digunakan saat data yang dianalisis berasal dari populasi yang diasumsikan memiliki distribusi normal, sedangkan untuk Uji Non Parametrik tidak diperlukan data yang berdistribusi normal (Sudrajat S.W, 1984; dan Siegel, S., 1988). Pemahaman perbedaan tersebut memungkinkan peneliti untuk memilih metode yang lebih sesuai dengan sifat data mereka.

1.2 Dasar-dasar Statistika

Walpolle (1993) menyatakan, bahwa dasar di dalam statistika meliputi pengumpulan data, pengolahan data, penyajian data, analisis data dan interpretasi data.

1.2.1 Pengumpulan Data

Pengumpulan data adalah proses sistematis untuk mengambil atau mengumpulkan data dan mengukur informasi dari berbagai sumber tentang variabel yang diminati, dengan tujuan untuk menjawab pertanyaan penelitian, mengujim hipotesis atau mengevaluasi hasil. Data bisa diperoleh dari lapangan atau labolatorium.

Pengumpulan data dilakukan pada awal seseorang mau melakukan penelitian atau pengumpulan data untuk keperluan analisis. Data bisa berupa nilai (kuantitatif) atau kategori (kualitatif), data ini berkaitan selanjutnya dengan analisis yang akan diguakan. Sebelum ke analisis data perlu diolah atau dibuatkan dalam bentuk tabulasi (*main table*).

Pengumpulan data melibatkan penentuan tujuan penelitian, memilih metode pengumpulan data, memastikan akurasi dan keandalan data, menganalisis data. Pentingnya pengumpulan data yang akurat dan relevan untuk menjaga integritas penelitian dan memastikan hasil yang valid, dapat membuat keputusan bisnis yang tepat, mengembangkan produk yang lebih baik, meningkatkan oprasi bisnis. Juga data memungkinkan identifikasi tren, pola dan peluang yang mungkin tidak terlihat secara langsung.

1.2.2 Pengolahan Data

Pengolahan data adalah proses mengubah data dari data mentah menjadi informasi yang lebih bermanfaat dan mudah dipahami. Ini melibatkan serangkaian kegiatan sampai penyajian dalam bentuk yang siap untuk digunakan.

Tujuan utama untuk mendapatkan data yang dapat digunakan dalam pengambilan keputusan, analisis dan ntujuan lainnya. Tujuan selanjutnya mengubah dari

data mentah menjadi informasi yang berguna, mendukung pengambilan keputusan, meningkatkan efisiensi bisnis, mempermudah analisis, menyajikann informasi dengan cara yang lebih mudah.

1.2.3 Penyajian Data

Penyajian data adalah proses menyusun dan menyajikan informasi atau data agar mudah dipahami dan dianalisis. Proses ini biasa dilakukan dalam bentuk visual seperti table, diagram atau grafik, sehingga data lebih mudah dimengerti dibandingkan dengan hanya dalam bentuk mentah.

Tujuan penyajian data adalah memudahkan pemahaman data oleh audiens, mempermudah analisis data, mendukung pengambilan keputusan, memudahkan komunikasi informasi data.

1.2.4 Analisis Data

Analisis data adalah proses sistematis dalam memeriksa, membersihkan, mengubah dan memodelkan data untuk menemukan informasi yang berguna, menarik kesimpulan, dan mendukung pengambilan keputusan.

Analisis data melibatkan langkah-langkah:

1) Pengumpulan data dari berbagai sumber,
2) Pembersihan data untuk mengatasi data yang tidak lengkap, salah atau tidak relevan,
3) Transformasi data untuk mengubah data ke format yang sesuai untuk analisis lebih lanjut,
4) Analisis data untuk menemukan pola, tren, dan hubunhan dalam data,

5) Interpretasi data untuk menafsirkan hasil analisis untuk mendapatkan wawasan dan menarik kesimpulan dan

6) Visualisasi data untuk menyajikan data yang sudah dianalisis dalam bentuk yang mudah difahami, seperti table atau grafik.

Manfaat analisis data yaitu data yang dianalisis memberikan dasar yang kuat untuk menarik kesimpulan, meningkatkan efisiensi, pengembangan produk dan layanan, peningkatan keuntungan. Analisis data dapat diterapkan dalam berbagai bidang, bisnis, ilmu pengetahuan, ilmu sosial.

1.2.5 Interpretasi Data

Interpretasi data adalah proses memberikan makna atau penjelasan pada data yang sudah dikumpulkan dan dianalisis. Interpretasi data meliputi menganalisis data, menjelaskan data, membuat keputusan, dan menarik kesimpulan

1.3 Kegunaan Statistika

Beberapa kegunaan statistika sebagai berikut (Walpole, 1984; Steel and Torrie, 1981; dan Nasution A,H, 1980):

1.3.1 Merangkum dan Menganalisis Data

Statistika digunakan untuk merangkum dan menggambarkan data mentah dalam bentuk dalam bentuk yang lebih mudah dipahami, seperti rata-rata, median, modus, simpangan baku (*standard deviation*), ragam (*variation*). Misalnya menghitung rata-rata nilai ujian matakuliah tertentu di suatu kelas, ini dapat mengetahui perform akelas secara keseluruhan, dan sebagainya.

1.3.2 Menguji Hipotesis

Statistika adalah alat penting dalam penelitian ilmiah, dan biasanya untuk melakukan pengujian terhadap dugaan sementara (hipotesis) yaitu dengan menggunakan alat-alat uji dalam statistika, seperti uji-t, Uji-Z dan Uji-F. Misalnya untuk membandingkan pendapata petani yang menggunakan pupuk organik dengan petani yang menggunakan pupuk an-organik, biasa dibantu dengan uji banding yaitu uji-t kalua datanya sampel dan uji-Z kalua datanya populasi.

1.3.3 Memprediksi

Statistika dapat digunakan untuk memprediksi tentang kejadian dimasa depan berdasarkan data historis. Misalkan memprediksi cuaca, pertumbuhan penduduk, pertumbuhan tanaman, pertumbuhan ekonomi atau tren pasar, dan sebagainya.

1.3.4 Analisis Data

Statistika menyediakan alat analisis yang bisa digunakan untuk membandingkan yaitu uji statistika (uji-t, uji-Z dan uji-F), dan juga untuk melihat hubungan atau memprediksi yaitu dengan menggunakan analisis regresi sederhana atau berganda. Misalnya menganalisis data penjualan untuk mengidentifikasi tren dan factor yang mempengaruhi penjualan, menganalaisis pengaruh factor-faktor produksi terhadap jumplah produksi padi, menganalisis hubungan antara lama berolahraga dengan jumlah denyut nadi orang dewasa antara 40 tahun sampai dengan 60 tahun, dan sebagainya.

1.3.5 Pengambilan Keputusan

Statistika membantu dalam pengambilan suatu keputusan, berdasarkan bukti data yang objektif. Keputusan ini diperoleh dari hasil pengumpulan data dan pengolahan atau analisis data. Misalnya mengumpulkan data hasil survai untuk memutuskan kebijakann public atau strategi pemasaran, data hasil studi kasus untuk menentukan kelayakan suatu usaha, dan sebagainya.

1.3.6 Pemantauan Kinerja

Pemantauan kinerja dalam statistika melibatkan penggunaan metode statistik untuk mengumpulkan, menganalisis, dan mengevaluasi atau menginterpretasikan data terkait kinerja, baik itu kinerja individu, tim, system atau proses. Tujuannya adalah untuk mengidentifikasi tren, pola dan masalah potensial, serta untuk mengukur kemajuan menuju tujuan yang ditetapkan. Konsep dasar pemantaun kinerja dalam statistic a, yaitu pengumpulan data, analisis deskriptif, analisis inferensia, visualisasi data. Pemantauan kinerja dapat diterapkan diberbagai bidang, termasuk bisnis, IT, dan olahraga.

1.3.7 Kualiti Kontrol

Statistika digunakan dalam pengendalian suatu kualitas produksi untuk memastikan standar kualitas terpenuhi. Misalnya menggunakan data statistik untuk memantau proses produksi dan mengidentifikasi potensi masalah kualitas.

1.3.8 Pertanian, Ekonomi, Pendidikan, Kesehatan, dll.

Statistika biasa digunakan dalam bidang pertanian, yaitu melihat pengaruh pemupukan terhadap pertumbuhan tanaman. Dalam bidang Ekonomi yaitu untuk melihat pengaruh faktor-faktor (usia, jarak tempuh, insentif, dll.) terhadap produktivitas tenaga kerja, dalam bidang Pendidikan yaitu membedakan hasil nilai ujian dengan menggunakan cara pembelajaran dan dalam Kesehatan yaitu menghubungkan lama berolahraga dengan denyut nadi seseorang.

1.4 Penggunaan Istilah dalam Statistika

1.4.1 Metode Statistika

Metode statistik adalah prosedur yang digunakan untuk mengolah, menganalisis, dan menyajikan data. Prosedur ini mencakup berbagai tahap, mulai dari pengumpulan data sampai dengan penarikan kesimpulan, pada umumnya metode statistika dibagi dua bagian, yaitu statistika deskriptif dan statistika inferensia (Gomez and Gomez, 1984; Walpole, 1993; dan Steel and Torrie, 1981).

Statistika deskriptif fokus pada penyajian dan ringkasan data, sedangkan statistika inferensia digunakan untuk menarik kesimpulan dan membuat prediksi berdasarkan data yang ada.

1) Statistikia Deskriptif

   Tujuannya yaitu hanya menyajikan, meringkas, menggambarkan data. Metode yang biasa digunakan yaitu menggunakan tabulasi, grafik dan ukurang ringkasan (mean, median, modus).

Contoh mempresentasekanpenjualan suatu produk dalam bentuk grafik atau menghitung rata-rata penjualan per bulan.

2) Statistika inferensial

Tujuannya yaitu untuk menarik suatu kesimpulan, membuat prediksi dan menguji hipotesis berdasarkan data yang ada. Metode yang digunakan misalnya penarikan sampoel untuk menduga populasi, menguji hipotesis dengan menggunakan alat uji (uji-t, uji-Z, uji-F, dll.).

Contoh menguji apakah ada perbedaan signifikan antara rata-rata pendapatan dua kelompok, atau membuat prediksi tentang prilaku konsumen dengan keputusan pembelian dengan cara survai.

1.4.2 Statistik dan Parameter

Statistik adalah ukuran numerik yang menggambarkan karakteristik sampel, sedangkan parameter adalah ukuran numerik yang menggambarkan populasi. Dengan kata lain statistik mengacu pada sebagian dari kelompok, sedangkan parametermengacu pada keseluruhan kelompok (populasi). Contoh parameter adalah rata-rat6a usia semua penduduk Indonesia, sedangkan kaqlu statistic yaitu rata-rata usia dari 1000 orang yang dipilih secara acak penduduk Indonesia.

1.4.3 Populasi dan Sampel

Populasi adalah keseluruhan objek atau individu yang menjadi perhatian si peneliti. Populasi merupakan kelompok besar yang menjadi target penelitian. Misalkan, jika penelitian tentang efektivitas metode belajar, maka populasinya bisa jadi seluruh siswa di suatu sekolah atau bahkan di siswa diseluruh daerah.

Sedangkan sampel adalah sebagian dari populasi yang dipilih untuk mewakili karakteristik populasi secara keseluruhan. Sampel harus dipilih secara acak dengan kata lain mempunyai kesempatan yang sama untuk diambil sebagai sampel. Cara pengambilan sampel dapat dilakukan dengan menggunakan kocokan atau menggunakan table acak.

1.4.4 Statistika Parametrik dan Statistika Non Parametrik

Statistika parametrik dan statistika non parametrik adalah dua pendekatan yang berbeda dalam analisis statistik, khususnya dalam uji hipotesis. Perbedaan utama terletak pada asumsi yang mendasari dan jenis data yang digunakan.

Statistika parametrik mengasumsikan data mengikuti distribusi normal dan menggunakan parameter populasi seperti rata-rata dalam analisis, skala ukur yang digunakan yaitu skala rasio dan bisa juga menggunakan skala interval, uji yang biasa digunakan uji-t, uji-Z dan uji-F.

Sedangkan statistika non parametrik tidak membuat asumsi tentang distribusi data dan sering digunakan untuk data yang tidak berdistribusi normal, skala yang digunakan adalah skala nominal dan ordinal dan bisa juga skala interval, uji yang biasa digunakan uji chi-square, uji friedman, uji korelasi spearman, uji Wilcoxon, uji Kruskal-wallis, dan seterusnya.

1.4.5 Statistik Univariat dan Multivariat

Statistik univariat dan statistik multivariat adalah dua pendekatan yang berbeda dalam analisis data. Statistik univariat berfokus pada analisis satu variabel pada satu waktu. Sedangkan statistik multivariat adalah statistic yang melibatkan dua atau lebih variabel secara bersamaan, untuk memahami hubungan antara variabel dan interaksi kompleks diantaranya.

Contoh untuk analisis statistik univarat adalah perhitungan mean, medin, modus, simpangan baku (*standard deviation*) dan pembuatan grafik seperti histogram untuk satu variabel. Analisis ini berguna untukmeringkas dan mendeskripsikan data tunggal, seperti distribusi tinggi badan mahasiswa. Misalnya menghitung nilai rata-rata ujian.

Contoh untuk statistik multivariat yaitu regresi sederhana dan regresi berganda, analisi factor, analisis diskriminan dan analisis kluster. Analisis ini berguna untuk mengidentifikasi pola,hubungan dan perbedaan yang kompleks diantara variabel, serta untuk membuat prediksi dan klasifikasi. Misal menganalisis hubungan antara tingkat pendididkan, pendapatan dan kesehatan seseorang. Menganalisis hubungan antara produksi padi dengan factor-faktor yang mempengaruhi padi (pupuk, luas lahan, tenaga kerja, dll.).

DAFTAR PUSTAKA

Gomes, K.A. and A.A. Gomes. 1984. Statistical Procedures for Agricultural Research, 2 nd edition, an International. Rice Research Intitude book, A.Wiley-Intersci. Publ.,

John Wiley and Sons, New York-Chichester-Brisbane Toronto-Singapore.

Nasution, A.H. dan Barizi, 1980. Metode Statistik. PT. Gramedia Jakarta.

Siegel, S., & Castellan, N. J. 1988. Nonparametric Statistics for the Behavioral Sciences. McGraw-Hill.

Steel R.G.D and Torrie J.H. 1981. Principles and Procedures of Statistics. A Biometrical

Sudradjat S.W. M. 1984. Mengenal Ekonometrika Pemula. Penerbit Armico, Bandung.

Walpole R.E. 1975. Introduction to Statistics. New York Macmillan London Collier

Walpole R.E. 1993. Pengantar Statistika. Gramedia Pustaka Utama. Indonesia.

BAB 2
PENYAJIAN DAN PENGOLAHAN DATA

Oleh Bambang Suhartawan

Penyajian dan pengolahan data adalah proses mengubah data mentah menjadi informasi yang mudah dipahami dan bermakna melalui teknik organisasi, visualisasi, dan analisis statistic (In and Lee, 2017). Dalam penyajian data, informasi ditampilkan menggunakan tabel (seperti distribusi frekuensi dan tabel silang) dan grafik (seperti diagram batang, garis, lingkaran, dan histogram) yang dipilih berdasarkan jenis data dan tujuan komunikasi. Pengolahan data melibatkan perhitungan statistik deskriptif seperti ukuran pemusatan (mean, median, modus), ukuran penyebaran (range, varians, standar deviasi), dan ukuran posisi (kuartil, persentil) untuk memberikan gambaran komprehensif tentang karakteristik data (Bhattacharjee, Dhar and Subramanian, 2011). Proses ini harus mengikuti prinsip kesederhanaan, kejelasan, akurasi, dan konsistensi agar dapat mengkomunikasikan temuan secara efektif kepada berbagai audience, sambil menghindari kesalahan interpretasi yang umum seperti confusing correlation dengan causation atau penggunaan skala yang menyesatkan.

2.1 Konsep Dasar Penyajian Data

1. Pengertian Penyajian Data

Penyajian data adalah proses mengorganisir, menata, dan menyusun data mentah menjadi bentuk yang lebih mudah dipahami dan diinterpretasi. Dalam konteks statistik, penyajian data bertujuan untuk mengkomunikasikan informasi dengan cara yang jelas dan efektif kepada pembaca atau audience tertentu (Drees and Rappaz, no date).

Data mentah yang dikumpulkan dari penelitian atau survei seringkali berupa angka-angka yang tidak terstruktur. Melalui proses penyajian data, informasi tersebut ditransformasi menjadi format yang memungkinkan kita untuk melihat pola, tren, dan hubungan yang mungkin tidak terlihat sebelumnya.

2. Tujuan dan Manfaat Penyajian Data

Penyajian data memiliki beberapa tujuan utama. Pertama, untuk menyederhanakan informasi kompleks agar lebih mudah dipahami oleh audience yang beragam. Kedua, untuk mengidentifikasi pola dan tren dalam data yang dapat membantu dalam pengambilan keputusan. Ketiga, untuk membandingkan berbagai kelompok data atau periode waktu secara efektif .

Manfaat konkret dari penyajian data yang baik meliputi kemudahan dalam komunikasi hasil penelitian, peningkatan akurasi interpretasi, dan kemampuan untuk menyajikan argumen yang didukung bukti empiris. Dalam konteks bisnis, penyajian data yang efektif dapat membantu manajemen membuat keputusan strategis berdasarkan fakta, bukan asumsi.

3. Prinsip-prinsip Penyajian Data yang Efektif

Ada beberapa prinsip fundamental yang harus diperhatikan dalam penyajian data. Prinsip kesederhanaan mengharuskan kita untuk menyajikan informasi tanpa elemen yang tidak perlu atau membingungkan. Prinsip kejelasan memastikan bahwa setiap elemen dalam penyajian memiliki makna yang jelas dan dapat dipahami dengan mudah.

Prinsip akurasi menuntut bahwa data yang disajikan harus tepat dan tidak menyesatkan. Hal ini termasuk penggunaan skala yang appropriate dan tidak memanipulasi visual untuk mendukung agenda tertentu. Prinsip konsistensi memastikan bahwa format, warna, dan style yang digunakan uniform di seluruh presentasi.

4. Jenis-jenis Data dan Karakteristiknya

Data dapat dikategorikan berdasarkan berbagai kriteria. Berdasarkan sifatnya, data dibedakan menjadi data kualitatif (nominal dan ordinal) dan data kuantitatif (diskrit dan kontinu). Data nominal seperti jenis kelamin atau agama tidak memiliki urutan, sedangkan data ordinal seperti tingkat pendidikan memiliki urutan yang jelas.

Data kuantitatif diskrit berupa bilangan bulat seperti jumlah anak dalam keluarga, sementara data kontinu dapat berupa nilai pecahan seperti tinggi badan atau berat badan. Pemahaman karakteristik data ini penting karena menentukan metode penyajian yang paling sesuai.

2.2 Penyajian Data dalam Bentuk Tabel

1. Struktur dan Komponen Tabel Statistik

Tabel statistik yang baik memiliki struktur yang jelas dan mudah dibaca. Komponen utama meliputi judul tabel yang menjelaskan isi tabel secara singkat namun informatif, stub (kolom paling kiri) yang berisi kategori atau kelompok data, dan kolom-kolom yang berisi data numerik dengan header yang jelas (Remshard and Queenborough, 2023).

Setiap tabel harus dilengkapi dengan catatan kaki jika diperlukan untuk menjelaskan sumber data, definisi variabel, atau keterangan khusus lainnya. Penggunaan garis horizontal dan vertikal harus konsisten untuk memudahkan pembacaan data.

2. Tabel Distribusi Frekuensi

Tabel distribusi frekuensi menunjukkan seberapa sering setiap nilai atau kelompok nilai muncul dalam dataset. Untuk data kategorik, tabel ini cukup sederhana dengan dua kolom utama: kategori dan frekuensi. Untuk data numerik, sering diperlukan pengelompokan data ke dalam interval kelas.

Dalam membuat tabel distribusi frekuensi, penting untuk mempertimbangkan jumlah kelas yang optimal. Terlalu sedikit kelas akan kehilangan detail, sedangkan terlalu banyak kelas akan membuat tabel sulit dibaca (Sopan *et al.*, 2013). Aturan umum menggunakan 5-15 kelas, tergantung ukuran dataset.

3. Tabel Silang (*Crosstab*)

Tabel silang atau crosstab digunakan untuk menampilkan hubungan antara dua atau lebih variabel kategorik. Struktur tabel ini menempatkan satu variabel

di baris dan variabel lainnya di kolom, dengan sel-sel menunjukkan frekuensi kombinasi kategori tersebut.

Tabel silang sangat berguna untuk menganalisis asosiasi antara variabel. Misalnya, untuk melihat hubungan antara tingkat pendidikan dan status pekerjaan, atau antara jenis kelamin dan preferensi produk. Analisis ini sering menjadi langkah awal sebelum melakukan uji statistik yang lebih kompleks.

4. Tabel Kontingensi

Tabel kontingensi adalah bentuk khusus dari tabel silang yang digunakan dalam analisis statistik untuk menguji independensi antara dua variabel kategorik. Tabel ini tidak hanya menampilkan frekuensi observasi, tetapi juga sering dilengkapi dengan frekuensi yang diharapkan jika variabel-variabel tersebut independen.

Struktur tabel kontingensi memungkinkan kita untuk menghitung berbagai statistik seperti chi-square, yang dapat digunakan untuk menguji apakah ada hubungan signifikan antara variabel-variabel yang dianalisis.

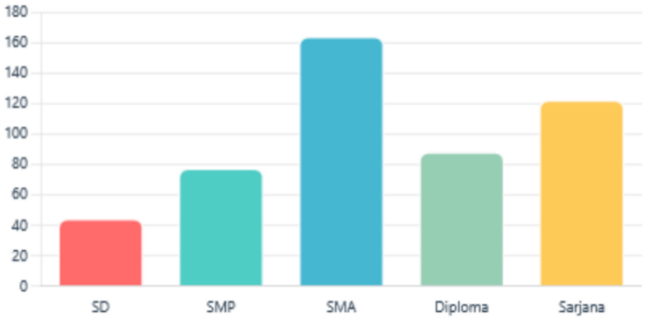
5. Aturan Pembuatan Tabel yang Baik

Tabel yang efektif harus mengikuti beberapa aturan dasar. Pertama, gunakan judul yang deskriptif dan informatif. Kedua, pastikan header kolom dan baris jelas dan tidak ambigu. Ketiga, susun data secara logis, misalnya dari yang terkecil ke terbesar atau berdasarkan abjad. Keempat, gunakan format angka yang konsisten, termasuk jumlah desimal dan pemisah ribuan. Kelima, hindari menggunakan terlalu banyak garis atau border yang dapat mengalihkan perhatian dari data itu sendiri. Terakhir, berikan sumber data dan catatan kaki yang diperlukan.

2.3 Penyajian Data dalam Bentuk Grafik

1. Diagram Batang (*Bar Chart*)

Diagram batang efektif untuk membandingkan data kategorik atau data numerik diskrit. Tinggi atau panjang batang menunjukkan nilai atau frekuensi dari setiap kategori. Diagram batang dapat disusun secara vertikal (*column chart*) atau horizontal (*bar chart*), tergantung pada panjang label kategori dan preferensi desain. Sebagai contoh Grafik perbandingan tingkat Pendidikan.



Tingkat Pendidikan	Jumlah Responden	Persentase
SD	45	9%
SMP	78	15.6%
SMA	165	33%
Diploma	89	17.8%
Sarjana	123	24.6%

Gambar 2. 1 Perbandingan Tingkat Pendidikan

Untuk data yang memiliki urutan natural seperti tingkat pendidikan atau kelompok umur, batang harus disusun sesuai urutan tersebut. Untuk data nominal tanpa urutan alami, batang dapat disusun berdasarkan nilai dari tertinggi ke terendah untuk memudahkan perbandingan.

2. Diagram Garis (*Line Chart*)

Diagram garis paling cocok untuk menampilkan data time series atau data yang menunjukkan perubahan seiring waktu. Sumbu X biasanya menunjukkan waktu (hari, bulan, tahun), sedangkan sumbu Y menunjukkan nilai variabel yang diukur. Garis yang menghubungkan titik-titik data membantu kita melihat tren secara visual, Contoh :

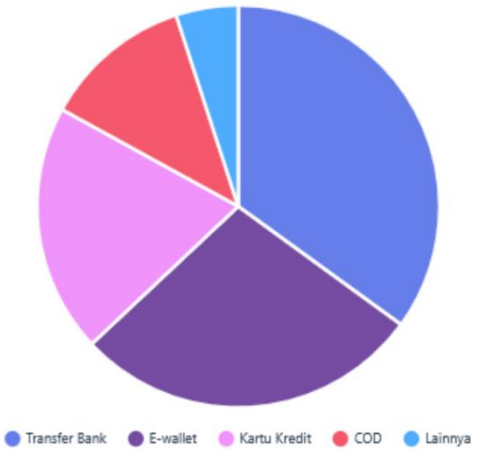
Interpretasi: *Grafik menunjukkan tren penjualan yang meningkat dengan fluktuasi seasonal. Puncak penjualan terjadi pada bulan November-Desember (periode liburan) dan penurunan pada awal tahun.*

Gambar 2. 2 Grafik Perubahan Penjualan

Diagram garis juga dapat digunakan untuk membandingkan tren beberapa variabel atau kelompok dengan menggunakan garis yang berbeda warna atau style. Ini sangat berguna dalam analisis komparatif antar periode atau antar kelompok.

3. Diagram Lingkaran (*Pie Chart*)

Diagram lingkaran menunjukkan proporsi atau persentase dari keseluruhan. Setiap segmen mewakili kategori dengan ukuran proporsional terhadap nilainya. Diagram ini paling efektif ketika jumlah kategori tidak terlalu banyak (idealnya 5-7 kategori) dan ketika kita ingin menekankan bagian dari keseluruhan, contoh :



Metode Pembayaran	Jumlah	Persentase
Transfer Bank	350	35%
E-wallet	280	28%
Kartu Kredit	200	20%
COD	120	12%
Lainnya	50	5%

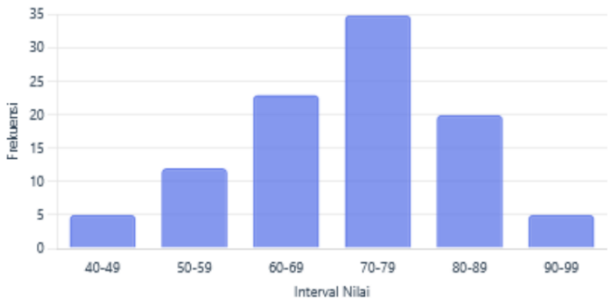
Gambar 2. 3 Diagram Metode Pembayaran

Penggunaan warna yang kontras dan label yang jelas sangat penting dalam diagram lingkaran. Segmen terkecil sebaiknya ditempatkan di posisi yang mudah dibaca, dan pertimbangkan untuk menggabungkan kategori-kategori kecil menjadi kategori "lainnya" jika diperlukan.

4. Histogram dan Poligon Frekuensi

Histogram menunjukkan distribusi data numerik kontinu dengan membagi data ke dalam interval kelas. Tidak seperti diagram batang, batang-batang dalam histogram saling bersentuhan karena mewakili data

kontinu. Lebar setiap batang mewakili interval kelas, sedangkan tingginya mewakili frekuensi, contoh :



Interval Nilai	Frekuensi	Frekuensi Relatif
40-49	5	5%
50-59	12	12%
60-69	23	23%
70-79	35	35%
80-89	20	20%
90-99	5	5%

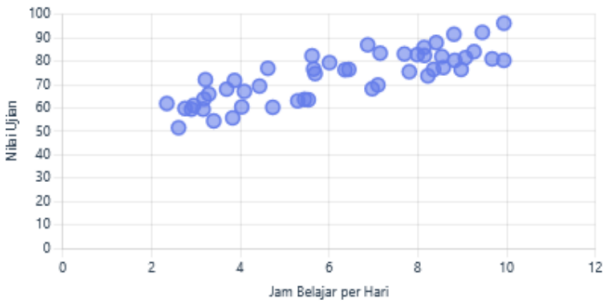
Gambar 2. 4 Histogram Frekuensi Nilai

Poligon frekuensi adalah garis yang menghubungkan titik tengah puncak setiap batang histogram. Grafik ini memberikan gambaran yang lebih halus tentang bentuk distribusi data dan memudahkan perbandingan beberapa distribusi dalam satu grafik.

5. Diagram Pencar (*Scatter Plot*)

Diagram pencar menunjukkan hubungan antara dua variabel numerik. Setiap titik dalam grafik mewakili satu observasi dengan koordinat X dan Y yang sesuai dengan nilai kedua variabel. Pattern titik-titik dapat

mengindikasikan jenis hubungan: linear positif, linear negatif, atau tidak ada hubungan, contoh :



Korelasi: *Terlihat hubungan positif yang cukup kuat antara jam belajar dan nilai ujian. Semakin banyak jam belajar, cenderung semakin tinggi nilai yang diperoleh.*

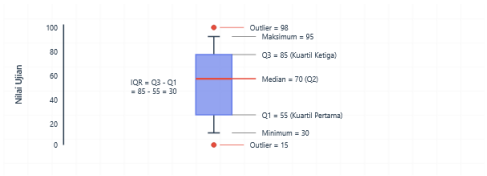
Gambar 2. 5 Perbedaan Jam Belajar terhadap Nilai Ujian

Diagram pencar sangat berguna dalam analisis korelasi dan regresi. Outlier atau nilai ekstrem juga mudah diidentifikasi dalam diagram ini, yang penting untuk analisis data yang akurat.

6. *Box Plot* dan *Stem-and-Leaf Plot*

Box plot (diagram kotak) memberikan ringkasan visual dari distribusi data menggunakan lima angka: minimum, kuartil pertama, median, kuartil ketiga, dan maksimum. Kotak menunjukkan 50% data tengah, sedangkan garis atau "*whiskers*" menunjukkan sebaran data keseluruhan.

Stem-and-leaf plot menggabungkan keunggulan tabel frekuensi dan histogram. Bagian "*stem*" menunjukkan digit utama, sedangkan "*leaf*" menunjukkan digit terakhir. Plot ini mempertahankan nilai data asli sambil menunjukkan distribusinya secara visual, contoh :



Statistik	Nilai	Penjelasan
Minimum	30	Nilai terendah (bukan outlier)
Q1 (Kuartil Pertama)	55	25% siswa mendapat nilai di bawah 55
Median (Q2)	70	50% siswa mendapat nilai di bawah 70
Q3 (Kuartil Ketiga)	85	75% siswa mendapat nilai di bawah 85
Maksimum	95	Nilai tertinggi (bukan outlier)
IQR	30	Rentang 50% data tengah (Q3 - Q1)
Outlier	15, 98	Nilai ekstrem di luar batas normal

Gambar 2. 6 Box Plot Nilai Ujian Mahasiswa

2.4 Pengolahan Data Deskriptif

1. Ukuran Pemusatan Data (Mean, Median, Modus)

Mean atau rata-rata aritmatika adalah jumlah semua nilai dibagi dengan jumlah observasi. Mean sensitif terhadap nilai ekstrem dan paling cocok untuk data yang terdistribusi normal (Rjr and Brownlow, 1979). Median adalah nilai tengah ketika data diurutkan, yang lebih robust terhadap outlier. Modus adalah nilai yang paling sering muncul dalam dataset.

Pemilihan ukuran pemusatan yang tepat tergantung pada distribusi data dan tujuan analisis. Untuk data yang skew atau memiliki outlier, median seringkali lebih representatif daripada mean. Untuk data kategorik, modus adalah satu-satunya ukuran pemusatan yang meaningful.

2. Ukuran Penyebaran Data (*Range*, Varians, Standar Deviasi)

Range adalah selisih antara nilai maksimum dan minimum, memberikan gambaran kasar tentang sebaran data. Varians mengukur rata-rata kuadrat deviasi dari mean, sedangkan standar deviasi adalah akar kuadrat dari varians dan memiliki satuan yang sama dengan data asli.

Standar deviasi yang kecil menunjukkan data yang clustered di sekitar mean, sedangkan standar deviasi yang besar menunjukkan data yang tersebar luas. Dalam distribusi normal, sekitar 68% data berada dalam satu standar deviasi dari mean.

3. Ukuran Posisi Data (Kuartil, Desil, Persentil)

Kuartil membagi data menjadi empat bagian yang sama. Q1 (kuartil pertama) adalah nilai di bawah mana 25% data berada, Q2 adalah median, dan Q3 adalah nilai di bawah mana 75% data berada. Desil membagi data menjadi sepuluh bagian, sedangkan persentil membagi menjadi seratus bagian.

Ukuran posisi ini berguna untuk memahami distribusi data dan membuat perbandingan. Misalnya, nilai ujian di persentil ke-90 berarti lebih baik dari 90% peserta ujian lainnya.

4. Ukuran Bentuk Distribusi (*Skewness* dan Kurtosis)

Skewness mengukur asimetri distribusi data. Skewness positif menunjukkan distribusi memiliki ekor yang panjang di sisi kanan (*right-skewed*), sedangkan *skewness* negatif menunjukkan ekor panjang di sisi kiri (*left-skewed*). Distribusi simetris memiliki skewness mendekati nol.

Kurtosis mengukur "ketajaman" puncak distribusi dibandingkan dengan distribusi normal. Kurtosis tinggi menunjukkan distribusi dengan puncak yang tajam dan ekor yang tebal, sedangkan kurtosis rendah menunjukkan distribusi yang lebih datar.

2.5 Teknik Pengelompokan Data

1. Penentuan Kelas Interval

Pengelompokan data diperlukan ketika bekerja dengan dataset besar atau data kontinu. Jumlah kelas interval harus optimal: tidak terlalu sedikit sehingga kehilangan informasi, dan tidak terlalu banyak sehingga sulit diinterpretasi. Aturan Sturges menyarankan jumlah kelas = 1 + 3.3 log(n), di mana n adalah jumlah data.

Lebar interval kelas sebaiknya sama untuk semua kelas agar memudahkan interpretasi. Lebar interval = (nilai maksimum - nilai minimum) / jumlah kelas. Batas bawah kelas pertama sebaiknya nilai bulat yang mudah diingat dan sedikit di bawah nilai minimum data.

2. Pembuatan Distribusi Frekuensi Berkelompok

Distribusi frekuensi berkelompok menunjukkan berapa banyak data yang jatuh dalam setiap interval kelas. Setiap kelas harus *mutually exclusive* (tidak tumpang tindih) dan *collectively exhaustive* (mencakup semua data). Format penulisan interval harus konsisten, misalnya 10-19, 20-29, dan seterusnya.

Titik tengah kelas digunakan sebagai representasi nilai dalam interval tersebut untuk perhitungan statistik lebih lanjut. Titik tengah = (batas bawah + batas atas) / 2.

3. Frekuensi Relatif dan Frekuensi Kumulatif

Frekuensi relatif menunjukkan proporsi atau persentase data dalam setiap kelas terhadap total data. Frekuensi relatif = frekuensi kelas / total data. Jumlah semua frekuensi relatif harus sama dengan 1 atau 100%.

Frekuensi kumulatif menunjukkan jumlah akumulatif data dari kelas pertama hingga kelas tertentu. Frekuensi kumulatif berguna untuk menentukan persentil dan membuat kurva ogive.

4. Kurva Ogive

Kurva ogive adalah grafik dari frekuensi kumulatif. Sumbu X menunjukkan batas atas kelas, sedangkan sumbu Y menunjukkan frekuensi kumulatif. Kurva ogive memiliki bentuk S dan berguna untuk menentukan median, kuartil, dan persentil secara grafis.

Ada dua jenis ogive: ogive "kurang dari" yang dimulai dari 0 dan naik ke total frekuensi, dan ogive "lebih dari atau sama dengan" yang dimulai dari total frekuensi dan turun ke 0.

2.6 Analisis Data Kategorik

1. Penyajian Data Nominal dan Ordinal

Data nominal seperti jenis kelamin, agama, atau warna tidak memiliki urutan alami. Penyajian data nominal biasanya menggunakan diagram batang atau lingkaran, dengan urutan kategori berdasarkan frekuensi (dari tertinggi ke terendah) atau abjad.

Data ordinal memiliki urutan yang bermakna seperti tingkat pendidikan atau rating kepuasan. Dalam penyajian data ordinal, urutan kategori harus dipertahankan sesuai tingkatannya. Diagram batang

lebih cocok daripada diagram lingkaran untuk data ordinal.

2. Tabel Frekuensi untuk Data Kategorik

Tabel frekuensi untuk data kategorik sederhana dengan kolom kategori, frekuensi absolut, dan frekuensi relatif. Dapat ditambahkan kolom persentase untuk memudahkan interpretasi. Total frekuensi harus sama dengan jumlah observasi, dan total persentase harus 100%.

Untuk data dengan banyak kategori, pertimbangkan untuk menggabungkan kategori-kategori dengan frekuensi kecil menjadi kategori "lainnya" agar tabel lebih ringkas dan mudah dibaca.

3. Diagram untuk Data Kategorik

Selain diagram batang dan lingkaran, data kategorik dapat disajikan menggunakan pictogram atau diagram gambar. Pictogram menggunakan simbol atau gambar yang relevan dengan data untuk menarik perhatian audience. Namun, hati-hati dengan skala agar tidak menyesatkan.

Diagram batang horizontal lebih cocok untuk kategori dengan nama yang panjang, sedangkan diagram batang vertikal cocok untuk kategori dengan nama pendek. Penggunaan warna harus konsisten dan mempertimbangkan *audience* yang mungkin memiliki *color blindness.*

4. Analisis Proporsi dan Persentase

Analisis proporsi berguna untuk membandingkan ukuran relatif berbagai kategori. Proporsi = frekuensi kategori / total frekuensi. Persentase adalah proporsi yang dikalikan 100%. Analisis ini memungkinkan

perbandingan antar dataset dengan ukuran yang berbeda.

Confidence interval untuk proporsi dapat dihitung untuk memberikan gambaran tentang ketidakpastian estimasi, terutama dalam survei dengan sampling. Hal ini penting untuk interpretasi yang akurat.

2.7 Penggunaan Software Statistik

1. Pengenalan *Software* Statistik (Excel, SPSS, R)

Microsoft Excel adalah tools yang paling aksesibel untuk analisis statistik dasar. Meskipun terbatas untuk analisis yang kompleks, Excel cukup untuk penyajian data sederhana dan perhitungan statistik deskriptif. Interface yang familiar membuat Excel mudah dipelajari (SaThierbach *et al.*, 2015).

SPSS (*Statistical Package for the Social Sciences*) adalah *software* khusus statistik dengan *interface point-and-click yang user-friendly*. SPSS cocok untuk peneliti yang tidak memiliki *background programming* namun perlu melakukan analisis statistik yang *sophisticated* (Pradipta, 2012).

R adalah bahasa pemrograman khusus statistik yang sangat *powerful* dan fleksibel. Meskipun memiliki *learning curve* yang curam, R menawarkan kemampuan analisis yang tidak terbatas dan gratis. Python dengan library seperti pandas dan matplotlib juga menjadi alternatif yang populer.

2. Input dan Import Data

Dalam Excel, data dapat diinput langsung atau *diimport* dari berbagai format file seperti CSV, TXT, atau database. Penting untuk memastikan tipe data yang tepat untuk

setiap kolom (*text, number, date*) agar analisis dapat dilakukan dengan benar.

SPSS dapat membaca data dari berbagai format termasuk Excel, CSV, dan database. Definisi variabel (nama, label, tipe data, *missing values*) harus dikonfigurasi dengan benar. R memiliki fungsi untuk membaca hampir semua format data dengan packages yang sesuai.

3. Pembuatan Tabel dan Grafik Otomatis

Excel menyediakan PivotTable untuk membuat tabel silang dan summary statistics dengan mudah. Chart wizard memungkinkan pembuatan berbagai jenis grafik dengan beberapa klik. Namun, customization terbatas dibandingkan software lain.

SPSS memiliki menu khusus untuk descriptive statistics dan berbagai jenis grafik. Output disajikan dalam format yang siap untuk publikasi. R memberikan kontrol penuh atas appearance grafik melalui packages seperti ggplot2, meskipun memerlukan coding.

4. Interpretasi *Output Software*

Output Excel relatif sederhana dan mudah dipahami. Perhatikan format angka dan pastikan hasil perhitungan masuk akal. Excel tidak selalu memberikan warning untuk kesalahan statistik, jadi user harus hati-hati.

Output SPSS lebih komprehensif dengan berbagai statistik dan test results. Penting untuk memahami asumsi di balik setiap test dan cara interpretasi *p-value, confidence interval*, dan *effect size.* R output bervariasi tergantung fungsi yang digunakan, namun umumnya lebih detail dan teknis.

2.8 Interpretasi dan Komunikasi Hasil

1. Cara Membaca dan Menginterpretasi Tabel

Mulai dengan membaca judul tabel untuk memahami konteks data. Perhatikan header kolom dan baris untuk memahami struktur data. Identifikasi pola dalam data: mana yang tertinggi, terendah, dan apakah ada tren yang jelas.

Jangan hanya fokus pada angka besar atau kecil, tetapi perhatikan proporsi dan perbandingan relatif. Cek konsistensi data dan cari anomali yang mungkin mengindikasikan error. Selalu perhatikan catatan kaki dan sumber data untuk konteks yang lengkap.

2. Cara Membaca dan Menginterpretasi Grafik

Mulai dengan membaca judul dan label sumbu untuk memahami apa yang digambarkan. Perhatikan skala sumbu - apakah dimulai dari nol atau tidak, karena ini dapat mempengaruhi interpretasi. Identifikasi tren umum, pola, atau hubungan yang terlihat.

Untuk time series, cari tren jangka panjang, siklus, atau seasonality. Untuk scatter plot, perhatikan arah dan kekuatan hubungan. Untuk bar chart, bandingkan tinggi relatif antar kategori. Selalu pertimbangkan apakah visual tersebut menyesatkan atau tidak.

3. Menyampaikan Temuan Statistik kepada *Audience*

Kenali *audience* Anda dan sesuaikan level teknis presentasi. Untuk *audience* non-teknis, fokus pada insight dan implikasi praktis daripada detail metodologi. Gunakan bahasa yang jelas dan hindari jargon statistik yang tidak perlu.

Mulai dengan *key findings* atau *takeaway* utama, baru kemudian berikan detail *supporting evidence*. Gunakan

visual yang efektif dan *storytelling* untuk membuat presentasi lebih engaging. Siapkan untuk pertanyaan dan berikan konteks yang cukup untuk interpretasi yang akurat.

4. Kesalahan Umum dalam Interpretasi Data

Kesalahan *correlation vs causation* adalah yang paling umum. Hanya karena dua variabel berkorelasi tidak berarti satu menyebabkan yang lain. Selalu pertimbangkan faktor *confounding* dan *alternative explanations.*

Kesalahan sampling bias terjadi ketika sample tidak representatif terhadap populasi target. *Cherry picking data* atau hanya menampilkan hasil yang mendukung argumen tertentu juga merupakan kesalahan serius. *Overgeneralization* dari sample kecil atau spesifik juga harus dihindari.

2.9 Studi Kasus dan Latihan

1. Contoh Kasus Nyata Penyajian Data

Kasus survei kepuasan pelanggan: Sebuah perusahaan e-commerce melakukan survei kepuasan terhadap 1000 pelanggan. Data yang dikumpulkan meliputi rating kepuasan (skala 1-5), kategori produk yang dibeli, metode pembayaran, dan demografis pelanggan.

Proses penyajian dimulai dengan cleaning data dan handling missing values. Kemudian dibuat tabel distribusi frekuensi untuk setiap variabel, tabel silang untuk melihat hubungan antar variabel, dan berbagai grafik untuk visualisasi. Analisis deskriptif dilakukan untuk setiap segmen pelanggan.

2. Latihan Pembuatan Tabel dan Grafik

Latihan 1: Diberikan data mentah penjualan bulanan selama 2 tahun, mahasiswa diminta membuat tabel yang menunjukkan tren penjualan dan grafik yang *appropriate*. Diskusikan pemilihan jenis grafik dan interpretasi tren yang terlihat.

Latihan 2: Data survei preferensi merek dari 500 responden dengan berbagai karakteristik demografis. Buat tabel silang untuk melihat hubungan antara umur dan preferensi merek, serta grafik yang menunjukkan market *share* setiap merek.

3. Analisis Data dari Berbagai Bidang

Bidang kesehatan: Analisis data epidemiologi tentang penyebaran penyakit. Fokus pada penyajian data spasial dan temporal, serta komunikasi risiko kepada publik. Perhatikan sensitivitas data dan *ethical considerations.*

Bidang pendidikan: Analisis data nilai siswa dan faktor-faktor yang mempengaruhi prestasi akademik. Penyajian data harus mempertimbangkan privacy siswa dan tidak stigmatizing. Fokus pada *actionable insights* untuk *improvement.*

Bidang bisnis: Analisis data finansial dan operasional perusahaan. Penyajian data untuk *stakeholders* yang berbeda memerlukan pendekatan yang berbeda. *Executive* summary untuk manajemen senior, *detailed analysis* untuk tim operasional.

4. Evaluasi Kualitas Penyajian Data

Kriteria evaluasi meliputi akurasi data, *appropriateness* metode penyajian, *clarity* dan *readability*, serta *effectiveness* dalam komunikasi *message*. Gunakan

checklist untuk memastikan semua elemen penting sudah *included*.

Peer review process dapat membantu mengidentifikasi bias atau kesalahan interpretasi. *Feedback* dari *audience* target juga *valuable* untuk *improving presentation. Continuous improvement* berdasarkan *lesson learned* dari setiap presentasi.

DAFTAR PUSTAKA

Bhattacharjee, M., Dhar, S.K. and Subramanian, S. (2011) 'Recent Advances in Biostatistics', *Recent Advances in Biostatistics* [Preprint]. Available at: https://doi.org/10.1142/8010.

Drees, B. and Rappaz, B.S. (no date) 'Design Considerations for Tables Design Considerations for Graphical Representations', pp. 17–18.

In, J. and Lee, S. (2017) 'Statistical data presentation', *Korean Journal of Anesthesiology*, 70(3), pp. 267–276. Available at: https://doi.org/10.4097/kjae.2017.70.3.267.

Pradipta (2012) 'PROPENSITY SCORE MATCHING IN SPSS', 早稲田大学大学院商学研究科 商学研究科紀要, 7(2), pp. 57–77.

Remshard, M. and Queenborough, S.A. (2023) 'Design of tables for the presentation and communication of data in ecological and evolutionary biology', *Ecology and Evolution*, 13(7), pp. 1–9. Available at: https://doi.org/10.1002/ece3.10062.

Rjr, Y. and Brownlow, J.D. (1979) :'] N / kSA'.

SaThierbach, K. *et al.* (2015) *Pengantar Statistika*, *Proceedings of the National Academy of Sciences*. Available at: http://dx.doi.org/10.1016/j.bpj.2015.06.056%0Ahttps://academic.oup.com/bioinformatics/article-abstract/34/13/2201/4852827%0Ainternal-pdf://semisupervised-3254828305/semisupervised.ppt%0Ahttp://dx.doi.org/10.1016/j.str.2013.02.005%0Ahttp://dx.doi.org/10.10.

Sopan, A. *et al.* (2013) 'Exploring Data Distributions: Visual Design and Evaluation', *International Journal of Human-Computer Interaction*, 29(2), pp. 77–95. Available at: https://doi.org/10.1080/10447318.2012.687676.

BAB 3

PROBABILITAS DAN DISTRIBUSI PELUANG

Oleh Deasy Wahyuni

3.1 Pengantar

Di dunia nyata, terdapat banyak peristiwa yang tidak dapat dipastikan dengan kepastian mutlak. Misalnya, hasil lemparan dadu, probabilitas hujan besok, atau peluang seorang siswa lulus ujian masuk perguruan tinggi. Untuk mengatasi ketidakpastian ini, kita menggunakan konsep probabilitas, yang merupakan landasan utama analisis statistik modern. (Sianturi, 2023)

Probabilitas adalah ukuran atau derajat kemungkinan suatu peristiwa terjadi. Peristiwa tersebut tidak akan terjadi jika probabilitasnya nol, dan jika probabilitasnya 1 maka peristiwa tersebut pasti terjadi. Probabilitas dapat dihitung dengan metode ini: (Wintat, 2022)

1. Klasik, berdasarkan teori dan hasil yang setara (contoh: melempar koin).
2. Relatif, berdasarkan pengamatan atau eksperimen yang diulang.

3. Subjektif, berdasarkan intuisi atau pengalaman pribadi. (Rocchi & Gianfagna, 2012)

Beberapa istilah penting dalam teori probabilitas meliputi:

- Dalam teori statistik atau teori peluang, Ruang sampel terdiri dari semua hasil yang mungkin dari sebuah eksperimen.
- Acara adalah himpunan bagian dari ruang sampel, dan
- Probabilitas suatu peristiwa (P(E)) adalah ukuran kemungkinan terjadinya peristiwa E.

Probabilitas digunakan untuk menganalisis fenomena acak dan merupakan dasar untuk teknik statistik seperti estimasi parameter, pengujian hipotesis, dan prediksi.

Distribusi probabilitas menggambarkan bagaimana probabilitas terdistribusi di seluruh nilai variabel acak. Variabel acak dibagi menjadi:

1. Variabel Acak Diskrit adalah teknik yang mengambil nilai-nilai spesifik dan dapat dihitung (misalnya jumlah pelanggan harian di sebuah toko).
2. *Continuous Random Variables* adalah teknik yang mengambil nilai dalam interval pada garis bilangan riil (misalnya berat badan seseorang). (Bewersdorff, 2021)

Ada dua bentuk utama distribusi probabilitas:

- Distribusi Probabilitas Diskrit: Contohnya, Distribusi Binomial dan Distribusi Poisson;
- Distribusi Probabilitas Kontinu: Misalnya, Distribusi Normal, Distribusi Eksponensial, dan Distribusi Chi-Kuadrat. (Santosa & Haryono, 2025)

Distribusi probabilitas memberikan informasi penting tentang:

- Polanya distribusi data,
- Rata-rata dan simpangan baku populasi,
- Kemungkinan terjadinya peristiwa langka atau ekstrem,
- Dasar untuk pengambilan keputusan berdasarkan data. (Supriadi & Ningsih, 2022)

3.2 Konsep Dasar Peluang

Probabilitas adalah cabang penting dari matematika dan statistik yang digunakan untuk mengukur kemungkinan terjadinya suatu peristiwa. Konsep ini muncul dari kebutuhan manusia untuk memprediksi dan mengukur ketidakpastian dalam berbagai fenomena kehidupan, seperti permainan dadu, cuacam, hingga analisis risiko dalam ekonomi dan kesehatan. Seiring dengan perkembangan ilmu pengetahuan, konsep probabilitas semakin diperluas dalam bentuk distribusi probabilitas. Distribusi probabilitas menggambarkan bagaimana nilai-nilai variabel acak terdistribusi dan seberapa besar kemungkinan setiap nilai tersebut terjadi. Ada dua jenis distribusi probabilitas, yaitu: Distribusi diskrit dan distribusi kontinu. (Rachmat & Sari, 2022)

3.2.1 Definisi Probabilitas

Dalam statistik, probabilitas adalah ukuran kemungkinan suatu peristiwa terjadi dalam ruang sampel. Nilai paling rendah 0 menunjukkan bahwa

peristiwa tersebut tidak mungkin terjadi, dan nilai tertinggi 1 menunjukkan bahwa peristiwa tersebut pasti terjadi. Secara statistik, probabilitas suatu peristiwa disamakan dengan:

$$P(A) = \frac{Jumlah\ kejadian\ yang\ diingankan}{Jumlah\ seluruh\ kejadian\ dalam\ ruang\ sampel}$$

Contoh:

Sebuah dadu dilempar sekali, probabilitas angka 4 muncul adalah:

$$P(4) = \frac{1}{6}$$

3.2.2 Ruang Sampel dan Acara

1. Ruang Sampel

Ruang sampel mengumpulkan semua hasil yang mungkin dari suatu eksperimen acak. Huruf S adalah simbol ruang sampel dalam statistik.

Contoh:

Ruang Sampel $S = H, T$

Di mana $H = kepala$, $T = ekor$, $S = 1,2,3,4,5,6$

2. Acara

Sebagian dari ruang sampel disebut peristiwa. Peristiwa dapat terdiri dari:

- Satu elemen (peristiwa sederhana).
- Lebih dari satu elemen (peristiwa majemuk).

Contoh:

- Dari lemparan dadu: kemunculan angka genap. $A = 2,4,6$

3.3 Distribusi Peluang

Distribusi probabilitas, yang juga dikenal sebagai distribusi probabilitas variabel acak, merupakan perpanjangan dari materi probabilitas dalam kursus statistik matematika. Kita mengakui keberadaan eksperimen statistik dalam probabilitas. Eksperimen ini dapat menghasilkan bilangan real, atau bilangan real. Variabel acak adalah identifikasi atau deskripsi numerik (dalam bilangan real) dari hasil suatu eksperimen. (Supriadi & Ningsih, 2022)

Ada dua jenis distribusi probabilitas (Nurhidayat, 2023)

- distribusi diskrit untuk variabel acak dengan nilai diskrit; d
- distribusi kontinu untuk variabel acak yang nilainya dapat diambil dari rentang nilai kontinu.

3.3.1 Variabel Acak Diskrit

Variabel acak diskrit hanya mengambil nilai-nilai tertentu yang terpisah. Contoh nilai variabel acak diskrit adalah0, 1, 2, 3, … (Sitopu & Siswadi, 2022)

3.3.2 Fungsi Probabilitas Massa

(PMF) PMF adalah fungsi yang memberikan probabilitas untuk setiap nilai variabel acak diskrit$X$. Fungsi ini memuaskan:

$$- P(X = x) \geq 0 untuk\ semua\ x$$

$$\sum_x P(X = x) = 1$$ - (Karlsson & Haimes, 1988)

3.3.3 Contoh Distribusi Diskrit yang Umum

a) Distribusi Binomial

Distribusi probabilitas diskrit yang disebut distribusi binomial menunjukkan jumlah keberhasilan dalam n percobaan yang saling independen, di mana setiap percobaan hanya memiliki dua pilihan: keberhasilan (sukses) atau kegagalan (gagal). (Pebriyanto et al., 2021)

Distribusi ini digunakan ketika suatu eksperimen dilakukan berulang kali dalam kondisi yang sama, dengan probabilitas keberhasilan yang tetap dalam setiap percobaan.

Distribusi ini digunakan untuk menghitung probabilitas keberhasilan N percobaan independen yang hanya memiliki dua opsi: keberhasilan atau kegagalan. Probabilitas keberhasilan pada setiap percobaan adalah P, sedangkan kegagalan adalah$q = 1 - p$ dengan persamaan:

$$P(X = 3) = (\frac{n}{k})p^k q^{n-k}, k = 0,1,2, \dots, n$$

Contoh: Lempar 5 koin, probabilitas 3 kepala adalah:

$$P(X = 3) = \binom{5}{3} 0.5)^3 (0.5)^2 10 \times 0.125 \times 0.25 = 0.3125$$

Manfaat/Penggunaan Distribusi Binomial

Distribusi binomial banyak digunakan dalam statistik, bisnis, kesehatan, dan ilmu sosial untuk:

1. Menghitung probabilitas keberhasilan

   Contoh: Probabilitas mendapatkan tepat 3 pelanggan yang membeli dari 10 orang yang ditawarkan produk.

2. Analisis kualitas

   Contoh: Menghitung probabilitas menemukan 2 produk cacat dari 20 produk dalam inspeksi.

3. Prediksi dalam pengambilan keputusan

   Contoh: Menentukan strategi pemasaran berdasarkan peluang keberhasilan suatu kampanye.

4. Dasar teoritis dalam distribusi lain

   Distribusi binomial merupakan dasar untuk distribusi lain, seperti distribusi normal (melalui aproksimasi distribusi binomial ketika nnn besar dan ppp mendekati 0,5).

b) Distribusi Poisson

Distribusi Poisson adalah distribusi probabilitas diskrit yang digunakan untuk menggambarkan jumlah peristiwa dalam interval waktu, jarak, area, atau volume tertentu, dengan asumsi bahwa peristiwa-peristiwa tersebut:

1. Terjadi secara acak,
2. Terjadi secara independen satu sama lain,
3. Jumlah rata-rata peristiwa (λ) per interval konstan.

Distribusi ini cocok jika Anda ingin mengetahui jumlah kali suatu peristiwa terjadi dalam interval waktu tertentu, bukan keberhasilan dalam sejumlah percobaan tertentu (seperti dalam distribusi binomial). Distribusi ini digunakan untuk menghitung probabilitas jumlah peristiwa acak yang terjadi dalam interval waktu atau ruang tertentu menggunakan rata-rataΔ . Dengan persamaan:

$$P(X = k) = \frac{A^k e^{-A}}{k!}, k = 0,1,2, \dots$$

Contoh:

Rata-rata 3 panggilan telepon per jam di kantor, probabilitas masker menerima tepat 2 panggilan dalam satu jam adalah:

$$P(X = 2) = \frac{3^2 e^{-3}}{2!} = \frac{9e^{-3}}{2} \approx 0.224$$ (Baru et al., 2022)

Manfaat / Penggunaan Distribusi Poisson

Distribusi Poisson banyak digunakan di berbagai bidang untuk memodelkan frekuensi peristiwa dalam interval waktu atau ruang, termasuk:

1. Manajemen Operasional
    - Memprediksi jumlah pelanggan yang datang ke bank dalam satu jam.
    - Menganalisis jumlah panggilan masuk ke pusat panggilan per menit.
2. Kesehatan dan Epidemiologi
    - Jumlah pasien yang dirawat di ruang gawat darurat dalam sehari.
    - Jumlah kasus penyakit langka di suatu wilayah.

3. Kualitas dan Produksi
    - Jumlah cacat pada sepotong kain sepanjang 100 meter.
    - Jumlah kerusakan mesin dalam satu minggu.
4. Lalu Lintas dan Transportasi
    - Jumlah kecelakaan lalu lintas di satu jalan per bulan.
    - Jumlah kendaraan yang melintasi jalan tol dalam waktu tertentu.
5. Ilmu Pengetahuan Alam
    - Jumlah partikel radioaktif yang terdeteksi dalam satu menit.

c) Distribusi Geometrik

Distribusi Geometrik adalah distribusi probabilitas diskrit yang mewakili probabilitas keberhasilan pertama terjadi pada percobaan ke-x dalam serangkaian percobaan Bernoulli yang independen, di mana setiap percobaan hanya memiliki dua kemungkinan: keberhasilan atau kegagalan (Yuzan et al., 2019) . Digunakan untuk mengukur jumlah percobaan hingga keberhasilan pertama terjadi. Persamaan untuk probabilitas keberhasilan pada percobaan ke-−k :

$$P(X = k)1 - p)^{k-1}p, k = 1,2,3 \ldots$$

Contoh: Jika probabilitas keberhasilan per percobaan adalah 0,2, probabilitas keberhasilan pertama terjadi pada percobaan ke-4 adalah:

$$P(X = 4) = (1 - 0.2)^3 \times 0.2 = 0.8^3 \times 0.2 = 0.1024$$

Manfaat / Penggunaan Distribusi Geometrik

Distribusi geometris berguna dalam berbagai situasi ketika kita tertarik pada jumlah percobaan hingga keberhasilan pertama, seperti:

1. Manajemen Operasional
    - Menentukan jumlah panggilan telepon yang harus dilakukan hingga seseorang menerima jawaban.
2. Industri dan Kualitas
    - Menghitung berapa banyak produk yang harus diperiksa hingga cacat pertama ditemukan.
3. Bisnis dan Pemasaran
    - Menganalisis berapa kali promosi perlu ditawarkan sebelum seorang pelanggan membeli.
4. Ilmu Komputer
    - Menentukan jumlah upaya login yang diperlukan hingga pengguna berhasil masuk.
5. Statistik Eksperimental
    - Mengetahui eksperimen mana yang memberikan hasil sukses pertama dalam uji coba laboratorium.

3.4 Distribusi Probabilitas Kontinu

3.4.1 Variabel Acak Kontinu

Setiap nilai dapat diterima untuk variabel acak kontinu dalam rentang waktu tertentu (misalnya 5-10 menit antara kedatangan bus).

3.4.2 Fungsi Kepadatan Probabilitas (PDF)

PDF (x) memenuhi:

- $f(x) \geq 0 untuk\ semua$
- Luas area di bawah kurva PDF sama dengan 1: (Sugito & Hoyyi, 2013)

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

Probabilitas acak berada dalam interval$[a, b]$ dihitung sebagai:

$$P(a \leq X \leq b) \int_{b}^{a} f(x)dx$$

3.4.3 Contoh Distribusi Kontinuum Umum

a) Distribusi Normal (Gaussian)

- Distribusi simetris berbentuk lonceng ini sangat penting dalam statistik karena banyak fenomena alam mengikuti distribusi ini.
- Parameter: rata-rataμ , simpangan baku . σ

- Fungsi PDF:

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

Contoh:

Jika rata-rata tinggi siswa adalah 165 cm dengan simpangan baku 7 cm, probabilitas bahwa seorang siswa memiliki tinggi antara 160 cm dan 170 cm dihitung dengan menemukan luas di bawah kurva normal pada interval tersebut (biasanya menggunakan tabel z).

b) Distribusi Eksponensial

- Digunakan untuk menghitung waktu antara peristiwa dalam proses Poisson.
- Fungsi PDF:

$$f(x) = Ae^{-Ae}, x \geq 0$$

Contoh: Waktu tunggu rata-rata antara kedatangan bus adalah 10 menit ($A = \frac{1}{10}$). Probabilitas bahwa waktu tunggu kurang dari 5 menit adalah (Novitasari Mara & Kusnandar, 2013) :

$$P(X \leq 5) = \int_0^5 \frac{1}{10} e^{-\frac{x}{10}} dx = 1 - e^{-0.5} = 1 - e^{-0.5}$$
$$\approx 0.393$$

3.5 Statistik Distribusi Probabilitas

3.5.1 Harapan (Rata-rata) Nilai rata-rata dari suatu variabel acak (Budiwanto, 2017)

- Diskrit:

$$E\ (X) = \sum_{x} xP(X = x)$$

- Diskrit

$$E(X) = \int_{-\infty}^{\infty} -xf(x)dx$$

3.5.2 Varians dan Simpangan Baku

Varians mengukur seberapa jauh data tersebar dari rata-rata:

-

$$Var(X) = E[(X - \mu)^2] = \sum_{X} (X - \mu)^2 P(X = x)$$

- Kontinu:

$$Var(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx$$

Simpangan baku adalah akar kuadrat dari variansi: (Limbongan, 2021)

$$\sigma = \sqrt{Var(X)}$$

3.6 Distribusi Binomial

Sebuah eksperimen dengan hanya dua kemungkinan hasil, yaitu A dan bukan A. Jika probabilitas terjadinya peristiwa A adalah$P(A) = \pi$ dan nilai$\pi = P(A)$ tetap untuk setiap percobaan, maka eksperimen yang diulang ini disebut eksperimen Bernoulli (Nurhaliza et al., 2024) . Selain itu, eksperimen Bernoulli dilakukan N kali secara independen. Misalkan X adalah peristiwa A dan$(X - X)$ menghasilkan bukan A. Karena$\pi = P(A)$, maka probabilitas terjadinya peristiwa A sebanyak X kali dari N percobaan dapat dihitung dengan persamaan berikut:

$$VIII\ (6) \ldots\ldots \boxed{p(x) = P(X = x) = \binom{n}{k}\pi^x(1-\pi)^{N-x}}$$

$$dimana\ x = 0,1,2, \ldots N, 0 < \pi < 1, dan$$

$$VIII\ (7) \cdots\cdots \boxed{\binom{N}{n} = \frac{N!}{x!\,(N-X)!}}$$

$$dengan\ N! = 1 \times 2 \times 3 \ldots X\ (N - 1) \times N\ dan\ 0! = 1! = 1(N!\ dibaca\ N\ faktorial).$$

Persamaan VIII (7) menunjukkan koefisien binomial, dan Persamaan VIII (6) menunjukkan hubungan distribusi dengan variabel acak diskrit yang disebut distribusi binomial. Jelas bahwa, menurut Persamaan VIII (1), hubungan tersebut memenuhi sifat dasar distribusi binomial$\sum p(x) = 1$. Penjumlahan dilakukan untuk seluruh variabel acak X yang mungkin terjadi di ruang sampel, sesuai dengan istilah distribusi yang digunakan. $x = 0,1, \ldots, N$

Distribusi binomial memiliki beberapa parameter penting; dalam konteks ini, parameter yang akan kita gunakan adalah *nilai harapan* distribusiμ dan simpangan bakuσ . Persamaannya adalah:

$$VIII\ (8) \cdots\cdots \boxed{\sigma = \frac{\mu = N\pi}{\sqrt{N\pi(1-\pi)}}}$$

Dengan kata lain, parameter tersebut dilihat dari peristiwa A.

Contoh:

1) Probabilitas wajah G muncul 6 kali saat melempar koin 10 kali adalah:

   $P(X = 6)\ (\binom{10}{6})(\frac{1}{2})6(\frac{1}{2})4\ (210)(\frac{1}{2})10\ \ 0{,}2050.$

   Misalkan X adalah jumlah kali wajah G muncul dalam suatu eksperimen.

2) Sebuah eksperimen dilakukan dengan melempar 10 dadu homogen secara bersamaan. Tentukan probabilitas bahwa dadu bernilai 6 muncul 8 kali?

   $\pi = P(6) = 1/6$ Tidak Dikenal Dalam konteks ini, nilai-nilai N=10N = 10N= 10 dan X=8X = 8X= 8, di mana X mewakili jumlah dadu yang menunjukkan angka 6 di sisi atas. Kemudian, perhitungan peluang dapat dilakukan sebagai berikut:

   $P(X = 8) = (\binom{10}{8})(\frac{1}{6})8(\frac{5}{6})2 = 0{,}000015.$

   Ini berarti bahwa dalam undian menggunakan 10 dadu, peluang angka 6 muncul 8 kali akan terjadi sekitar 15 kali dalam setiap satu juta percobaan**.**

3) Sebuah sampel acak terdiri dari 30 bola dengan 10% jenis bola bolling diambil. Berapa probabilitas bahwa sampel tersebut akan mengandung sebanyak 30 bola bolling (total)?
   a. 30 bola (total)
   b. 1 bola
   c. 2 bola
   d. minimum 1 bola
   e. maksimum 2 bola
   f. Tentukan jumlah rata-rata bola dengan jenis bola yang terdapat dalam sampel.

Solusi

a. Misalkan X mewakili jumlah bola bolling dalam sampel.

   $Maka\ \pi\ =\ \ peluang$ bola bolling $=\ 0{,}10$

   Semua termasuk dalam bola bolling. $X\ =\ 30$

   $$P(X\ =\ 30)\ =\ \binom{30}{30}(0{,}10)^{30}(0{,}90)^{0}\ =\ 10^{-30}$$

   Nilai probabilitasnya sangat kecil sehingga dapat dianggap hampir sama dengan nol.

b. Satu item yang termasuk dalam kategori A berarti jumlah item kategori A dalam sampel adalah satu, atau X=1

   $(X\ =\ 1)\ =\ (30\ 1\ )(0{,}10)1(0{,}90)29\ =\ 0{,}1409$

   Probabilitas mengandung bola bolling adalah 0,1409.

c. Diketahui bahwa$X$ = 2, sehingga:

   $(X\ =\ 2)\ =\ (30\ 2\ )(0.10)2(0{,}90)28\ =\ 0{,}2270.$

d. Setidaknya ada satu bola bolling,$X = 1,2,3,\cdots,30$. Oleh karena itu:

$$(X = 1) + P(X = 2) + \cdots + \quad P(X = 30) = 1,$$

sehingga:

$$1 - P(X = 0)\ Sekarang\ (X = 0) = (30\ 0\ )(0{,}10)0(0{,}90)30 = 0{,}0423.$$

1 - 0,0423 = 0,9577

e. Ada paling banyak 2 bola Bolling, artinya$X = 0,1,2$. Perlu ditemukan

$$(X = 0) + P(X = 1) + P(X = 2).$$

Sehingga = 0.0423 + 0.1409 + 0.2270 = 0.4102.

f. Menggunakan Persamaan VII(8), kita mendapatkan:$\mu = 30(0{,}1) = 3$. Dalam setiap sampel 30 objek, diperkirakan terdapat 3 objek yang termasuk dalam bola Bolling.

3.7 Distribusi Hipergeometrik

Misalkan populasi berukuran N mengandung D objek yang termasuk dalam suatu kategori tertentu. Sebuah sampel berukuran n dipilih dari populasi tersebut. Berapa probabilitas bahwa terdapat x objek dalam sampel yang termasuk dalam kategori tertentu? Persamaan distribusi hipergeometrik berikut menentukan jawabannya: (Kaharuddin, 2018)

$$VII(10).\ldots\ldots\ldots..p(x) = \frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}}$$

Di mana$x = 0, 1, 2, \ldots, n$ dan faktor di sisi kanan ditentukan oleh Persamaan . $VIII(7)$

Notasi untuk Rata-rata dalam distribusi hipergeometrik adalah$\mu = nD/N$. Contoh: Sebuah kelompok terdiri dari 50 orang dan 3 di antaranya lahir pada tanggal 1 Januari. 5 orang akan dipilih secara acak. Berapa probabilitas memilih 5 orang yang:

a) tidak lahir pada tanggal 1 Januari.

b) kurang dari satu orang lahir pada tanggal 1 Januari?

Penyelesaian:

a) misalkan x = jumlah orang dan n = 5 orang yang lahir pada tanggal 1 Januari. Maka dengan , , persamaan$N = 50D = 3VIII(10)$ menjadi:

$$p(0) = \frac{\binom{3}{0}\binom{47}{5}}{\binom{50}{5}} = 0{,}724$$

Kemudian probabilitasnya adalah 0,724, bahwa semua lima orang tidak lahir pada tanggal 1 Januari.

b) Jumlah maksimum 1 orang yang lahir pada tanggal 1 Januari berarti $x = 0\ dan\ x = 1.12$

$p(0)$ sudah dihitung di atas.

$$p(1) = \frac{\binom{3}{1}\binom{47}{4}}{\binom{50}{5}} = 0{,}253.$$

Oleh karena itu, probabilitas maksimum 1 orang lahir pada tanggal 1 Januari adalah 0,724 + 0,253 = 0,977.

3.8 Distribusi Normal

Semua distribusi yang dibahas sebelumnya memiliki variabel acak diskrit. Pembahasan selanjutnya adalah tentang distribusi dengan variabel acak kontinu, yaitu distribusi normal atau distribusi Gauss, yang merupakan salah satu distribusi paling luas digunakan dan penting dalam statistik (Li & Ren, 2013) . Jika variabel acak kontinu F memiliki fungsi kepadatan probabilitas (PDF), maka fungsi tersebut dapat digunakan untuk menghitung rata-rata, varians, dan probabilitas.

$$fungsi\ densitas\ pada\ X\ =\ x$$

Hubungan antara variabel acak dan probabilitas dapat diekspresikan dalam persamaan berikut**:**

$$VIII\ (13)\ldots\ldots f(x)\ =\ \frac{1}{\sigma\sqrt{2_\pi}}e^{\frac{1}{2(\frac{x-\mu}{\sigma})}}.$$

Dimana:

$\pi\ =\ 3{,}1416.$

$$e = 2{,}7183$$

$$\mu = rata - rata$$

σ = simpangan baku.

Dan nilai x,$-\infty < x < \infty$, sehingga variabel acak X berdistribusi normal. Sifat-sifat distribusi normal adalah:

1) Grafik terletak pada *sumbu x*.

2) Bentuknya simetris, $x\ =\ \mu$

3) Memiliki mode, sehingga kurva tersebut merupakan kurva unimodal yang mencapai puncaknya pada titik tersebut = μ *sebesar* $\frac{0{,}3989}{\sigma}$.

4) Grafik mendekati sumbu horizontal x secara asimtotik, baik ke kiri maupun ke kanan, tanpa pernah menyentuh sumbu$x = \mu + 3\sigma$ ke kanan dan**x**$= \mu - 3\sigma$ ke kiri.

5) Luas total di bawah grafik fungsi distribusi ini selalu sama, yang mewakili probabilitas total sebesar 1.

Untuk setiap pasanganμ danσ , sifat-sifat ini selalu terpenuhi, tetapi hanya bentuk kurva yang berbeda. Kurva menjadi platiskurik saat ukuran semakin besar, dan untukσ , kurva leptokurtik berkembang saat ukuran semakin kecil.

3.9 Distribusi Student (T)

Distribusi lain dengan variabel acak kontinu, selain distribusi normal, adalah distribusi Student atau distribusi t**.** Distribusi t, atau secara lengkap distribusi t-Student, adalah distribusi probabilitas yang digunakan dalam statistik inferensial, terutama ketika ukuran sampel kecil (biasanya kurang dari 30) dan simpangan baku populasi tidak diketahui**.** Fungsi kepadatan adalah:

$$\text{VIII(19)} \ldots\ldots f\ (t\ ) = k\ (1 + 2tn\text{ -1})\ 1/2n\ .$$

Untuk t,$-\infty < t < \infty$, dan K adalah bilangan konstan yang nilainya bergantung pada n, sehingga luas di bawah kurva sama. Dalam distribusi t ini, terdapat bilangan$(n - 1)$. Derajat kebebasan akan disingkat sebagai dk.$VIII(19)$ Jika jumlah derajat kebebasan adalah dk, maka dikatakan bahwa populasi *terdistribusi* t dengan$dk = (n - 1)$.

Distribusi ini memiliki bentuk yang mirip dengan grafik distribusi normal standar, simetris terhadap$t = 0$. Untuk$n \geq 30$, seperti yang dinyatakan dalam Persamaan VIII(16), distribusi t mendekati distribusi normal standar. Derajat kebebasan ditampilkan pada kolom dk di sebelah kiri, dan nilai odds ditampilkan pada baris atas (Pane & M., 2023)

3.10 Distribusi Chi Kuadrat

Distribusi Chi Kuadrat adalah distribusi dengan variabel acak kontinu. Distribusi chi-kuadrat (χ^2) adalah distribusi probabilitas yang digunakan untuk menguji hipotesis tentang varians populasi dan hubungan antara variabel kategorikal**.** Distribusi ini terbentuk dari jumlah kuadrat dari sejumlah variabel acak normal standar yang saling independen (z). Simbol yang digunakan untuk chi-kuadrat adalah (χ^2) (dibaca: ci kuadrat). Persamaan distribusi chi-kuadrat adalah:

$$VIII(20)\ldots\ldots\ldots\ldots f(u) = K.u^{1/2v-1}e^{1/2u}$$

$$VIII(21)\ldots\ldots\ldots\ldots x^2 = \sum \frac{(O_i - E_i)^2}{E_i}.$$

Dengan

$O_i = frekuensi\ observasi$

$E_i = frekuensi\ harapan$

$u = (\chi^2), u > 0, v = derajat\ kebebasan$, K = bilangan tetap yang bergantung pada v, dan e = 2.7183. Grafik distribusi chi-kuadrat umumnya berupa kurva positif yang

miring ke kanan. Kemiringan ini berkurang jika derajat kebebasan v semakin besar. (Fauzy, 2017)

Manfaat / Penggunaan Distribusi Chi-Square

Distribusi ini sangat penting dalam statistik, terutama dalam:

1. Uji Kesesuaian Data 2.

- Menguji apakah data yang diamati sesuai dengan distribusi teoretis yang diharapkan.
- Contoh: Apakah distribusi jenis kelamin di kelas sesuai dengan distribusi yang diharapkan?

2. Uji Kemandirian

- Menguji hubungan antara dua variabel kategorikal.
- Contoh: Apakah ada hubungan antara jenis kelamin dan preferensi produk?

3. Uji Homogenitas

- Menguji apakah dua atau lebih populasi memiliki proporsi yang sama dalam suatu kategori.
- Contoh: Apakah distribusi tingkat pendidikan sama di berbagai wilayah?

4**.** Uji Varians Populasi

- Menguji apakah varians dari suatu populasi sama dengan nilai yang dihipotesiskan.
- Digunakan dalam analisis variansi (ANOVA) sebagai dasar untuk pengujian.

3.11 Distribusi F

Distribusi F memiliki variabel acak kontinu. Distribusi F adalah distribusi probabilitas yang digunakan terutama dalam analisis variansi (ANOVA) dan untuk membandingkan dua variansi dari dua populasi yang independen. Persamaan fungsi kepadatan distribusi F adalah sebagai berikut:

$$VIII(21)\ldots\ldots\ldots\ldots f(F) = K \cdot \frac{F^{1/2(v_{1-2})}}{(1+\frac{v_1 F}{v_2})^{1/2(v_1+v_2)}}.$$

Dengan variabel acak F yang memenuhi batas-batas$F > 0$, K = konstanta yang nilainya bergantung padav_1 , danv_2 sedemikian sehingga luas di bawah kurva sama dengan satu,v_1 = dk pembilang dan$v_2 = dk$ penyebut. Distribusi F memiliki dua derajat kebebasan. Grafik distribusi F tidak simetris dan umumnya positif.

Dalam daftar distribusi F, nilai F untuk probabilitas 0,01 dan 0,05 dengan derajat kebebasan adalahv_1 danv_2 . Probabilitas ini sama dengan luas daerah yang diarsir di ujung kanan, dengan$dk = v_1$ di baris atas dan$dk = v_2$ di kolom kiri. Untuk setiap dk = ,v_2 berada di baris atas untuk probabilitas$p = 0{,}05$ dan di baris bawah untuk$p = 0{,}01$. (Pratikno et al., 2020)

Manfaat/ Penggunaan Distribusi F

Distribusi F sangat berguna dalam pengujian hipotesis, terutama dalam konteks berikut:

1. Analisis Variansi (ANOVA)

- Untuk menguji apakah rata-rata dari tiga atau lebih kelompok berbeda secara signifikan.

- Contoh: Apakah hasil belajar rata-rata berbeda di antara beberapa metode pembelajaran?

2. Uji Perbandingan Dua Varians (Uji F)

- Untuk menguji apakah dua populasi memiliki varians yang sama**.**
- Contoh: Apakah variasi tinggi badan di dua sekolah berbeda secara signifikan?

3. Regresi Linier Berganda

- Distribusi F digunakan untuk menguji signifikansi model regresi secara keseluruhan**.**
- Contoh: Apakah variabel independen secara bersama-sama mempengaruhi variabel dependen?

Contoh:

Untuk derajat kebebasan ($v_{1,2}$) = (24, 8),$p = 0{,}05$ maka$F = 3{,}12$ sedangkan untuk$p = 0{,}01$ kita mendapatkan$F = 5{,}28$. Oleh karena itu,$F_{0{,}005(24{,}8)} = 3{,}12$ dan$F_{0{,}001(24{,}8)} = 5{,}28$.

Daftar yang diberikan hanya untuk peluang p = 0,01 dan p = 0,05, tetapi sebenarnya masih mungkin mendapatkan nilai F dengan peluang 0,99 dan 0,95. Untuk persamaan ini digunakan:

$$VIII(22)\ldots\ldots\ldots \boldsymbol{F}_{(1-p)(v1,v2)} = \frac{\mathbf{1}}{\boldsymbol{F}_{p(v1,v2)}}$$

Dalam persamaan di atas, perhatikan pertukaran antara p dan (1 - p) serta pertukaran antara derajat kebebasan$(v_1 v_2)$ menjadi . $(v_1 v_2)$

3.12 Aplikasi pada Distribusi Normal

Dalam statistik induktif, sangat penting untuk mengetahui bentuk distribusi data yang mendasari model, karena hal ini memainkan peran penting dalam proses analisis data selanjutnya, seperti dalam teori estimasi parameter dan pengujian hipotesis. Teori pengujian hipotesis umumnya didasarkan pada asumsi bahwa populasi yang diteliti mengikuti distribusi normal. Jika asumsi ini tidak terpenuhi, maka kesimpulan yang diperoleh dari teori tersebut tidak valid atau tidak dapat diterapkan secara sah.

Oleh karena itu, sebelum suatu teori digunakan dan kesimpulan ditarik berdasarkan teori yang mengasumsikan normalitas, perlu dilakukan verifikasi terhadap asumsi normalitas tersebut. Berdasarkan data sampel yang tersedia, analisis dapat dilakukan untuk memberikan informasi apakah distribusi data dalam populasi mendekati distribusi normal atau tidak.

Langkah pertama adalah menyusun distribusi frekuensi data sampel populasi. Selanjutnya, dibuat distribusi frekuensi relatif kumulatif. Berbeda dengan distribusi frekuensi biasa, pada tahap ini batas-batas kelas interval ditentukan terlebih dahulu untuk menyusun distribusi kumulatif.

Frekuensi kumulatif relatif kemudian digambar pada kertas probabilitas normal, yang merupakan grafik khusus yang dirancang untuk mengevaluasi normalitas data. Sumbu horizontal (x) menunjukkan skala batas atas kelas, sedangkan sumbu vertikal (y) menunjukkan persentase kumulatif data. Skala pada sumbu vertikal umumnya dimulai dari 0,01% hingga 99,99%. Oleh karena itu, kelas interval

yang memiliki persentase kumulatif 0% dan 100% tidak perlu dimasukkan ke dalam grafik.

Selanjutnya, kertas probabilitas normal digunakan untuk memplot titik-titik yang ditentukan oleh pasangan batas atas kelas dan frekuensi kumulatif relatif. Periksa pola distribusi titik dengan cermat. Jika titik-titik tersebut berada di dekat atau di dekat garis lurus, maka data tersebut terdistribusi secara normal. Mengenai data itu sendiri, Data didistribusikan secara normal atau hampir normal, atau dapat diapproksimasi oleh distribusi normal.

Mengenai populasi dari mana data sampel diambil, dikatakan bahwa populasi tersebut memiliki distribusi normal atau hampir normal (atau dapat dianggap sebagai distribusi normal). Jika lokasi titik data menyimpang dari sekitar garis lurus, maka data atau populasi dari mana sampel tersebut tidak terdistribusi normal. Penyelidikan normalitas adalah upaya untuk mengetahui apakah populasi atau data terdistribusi secara normal. (Widiatmika, 2015)

DAFTAR PUSTAKA

Baru, M., Admission, B., Dan, P. T. N., Admission, B., & Dan, P. T. N. (2022). *Distribusi Poisson: Analisis Probabilitas Perkembangan Minat Mahasiswa Baru*. *Juni*. https://doi.org/10.13140/RG.2.2.22789.83683

Bewersdorff, J. (2021). Probabilitas dan Geometri. *Keberuntungan, Logika, dan Kebohongan Putih*, *Mei 2010*, 37-40. https://doi.org/10.1201/9781003092872-8

Budiwanto, S. (2017). Metode Statistik. Dalam *Fakultas Ilmu Olahraga, Universitas Negeri Malang 2017* (Edisi April).

Fauzy, A. (2017). Distribusi Chi Kuadrat. *Jurnal MIPA IKIP Malang*,*25* (1), 103-`04. https://www.researchgate.net/publication/319090524

Kaharuddin, A. (2018). *Distribusi Hipergeometrik*. *5*, 0–16.

Karlsson, P.-O, & Haimes, Y. Y. (1988). Distribusi probabilitas dan pembagiannya. Dalam *Water Resources Research* (Vol. 24, Edisi 1). Springer Singapore. https://doi.org/10.1029/WR024i001p00021

Li, N., & Ren, L. (2013). *Studi ketersediaan sistem komponen seri*.*2* (April), 94-100.

Limbongan, Y. (2021). *Statistik dan Desain Eksperimen*.

Novitasari Mara, M., & Kusnandar, D. (2013). Studi tentang Sifat-sifat Distribusi Normal Bivariat. *Jurnal Ilmiah Matematika dan Statistik serta Aplikasinya (Bimaster)*, *02*(2), 127–132.

Nurhaliza, Depriwana Rahmi, Annisah Kurniati, & Suci Yuniati. (2024). Model Distribusi Binomial dalam Mengukur Probabilitas Keberhasilan Uji Kualitas Layanan Sistem Informasi. *Jurnal Teknologi Industri dan Manajemen Terapan*,*3* (4), 405-410. https://doi.org/10.55826/jtmit.v3i4.506

Nurhidayat, W. D. (2023). *Distribusi Probabilitas Variabel Diskrit Wahyu Dwi Nurhidayat*.*4* (Januari), 1-3. https://www.researchgate.net/publication/367307192

Pane, S., & M., S. K. (2023). Distribusi Probabilitas. *Departemen Statistik UII*,*2* , 1-35. https://medium.com/statistics-uii/distribusi-peluang-dengan-r-b50bd0c7973a

Pebriyanto, A. F., Sartika, D., Ruspandi, I., Zihani, N., Sam, M. A. N., Gifari, M. F., Priatna, R. K., & Pradana, R. A. (2021). Distribusi Binomial sebagai Ukuran Keberhasilan dan Kegagalan Industri Rumahan @One Hand Made Production. *Bulletin of Applied Industrial Engineering Theory*, *2*(2), 94–98.

Pratikno, A. S., Prastiwi, A. A., & Ramahwati, S. (2020). Distribusi Peluang Acak Kontinu, Distribusi Normal, Distribusi Normal Standar, Distribusi T, Distribusi Chi Kuadrat, dan Distribusi F. *Osf Preprints*,*27* (3), 1-5. https://doi.org/10.31219/osf.io/grdnm

Rachmat, M. T., & Sari, R. P. (2022). Analisis Perhitungan dalam Penerapan Konsep Distribusi Peluang Diskrit dan Statistik Deskriptif. *Jurnal Pendidikan Matematika (JUDIKA EDUCATION),5* (1), 35-41. https://doi.org/10.31539/judika.v5i1.3465

Rocchi, P., & Gianfagna, L. (2012). Sebuah Esai tentang Dua Sifat Probabilitas. *Advances in Pure Mathematics,02* (05), 305-309. https://doi.org/10.4236/apm.2012.25041

Santosa, R. G., & Haryono, N. A. (2025). Penggunaan Persamaan Stirling dan Fungsi Pembangkit Moment dalam Proses Pembuktian Distribusi Normal sebagai Pendekatan terhadap Distribusi Binomial. *Epsilon: Jurnal Matematika Murni dan Terapan,18* (2), 240. https://doi.org/10.20527/epsilon.v18i2.14164

Sianturi, R. (2023). Penggunaan Teorema Binomial dalam Menentukan Peluang Suatu Kejadian. *Journal on Education,5* (4), 12922-12936. https://doi.org/10.31004/joe.v5i4.2281

Sitopu, J. W., & Siswadi, S. (2022). Fungsi Pembangkit Moment dari Distribusi Probabilitas Diskrit. *FARABI: Jurnal Matematika dan Pendidikan Matematika,5* (2), 144-153. https://doi.org/10.47662/farabi.v5i2.435

Sugito, S., & Hoyyi, A. (2013). Proses Antrian dengan Kedatangan Berdistribusi Poisson dan Pola Layanan Berdistribusi Umum. *Statistics Media,6* (1), 113-120. https://doi.org/10.14710/medstat.6.1.51-60

Supriadi, A., & Ningsih, Y. L. (2022). Kemampuan Representasi Matematika Siswa pada Materi

Distribusi Probabilitas. *Indiktika: Jurnal Inovasi Pendidikan Matematika*,*4* (2), 14-25. https://doi.org/10.31851/indiktika.v4i2.7678

Syamsul, M., Ramlan, P., Muhammadiyah, U., Rappang, S., Syakurah, R., & Ngkolu, N. W. (2022). *Statistik Kesehatan: Teori dan Aplikasi* (Edisi April 2023).

Widiatmika, K. P. (2015). Statistik Pendidikan. Dalam *Etika Jurnalistik di Koran Kuning: Studi Koran Green Light* (Vol. 16, Edisi 2).

Wintat, P. K. J. H. (2022). Probabilitas Dasar. *Artikel*,*02* (November), 2-5.

Yuzan, S., YANUAR, F., & DEVIANTO, D. (2019). Estimasi Parameter Distribusi Geometrik dengan Metode Bayes. *Jurnal Matematika UNAND*,*8* (3), 85-92. https://doi.org/10.25077/jmu.8.3.85-92.2019

BAB 4

PENGAMBILAN SAMPEL DAN DISTRIBUSI SAMPLING

Oleh Tedi Hartoyo

4.1 Pendahuluan

Metode penelitian adalah cara atau strategi menyeluruh untuk mendapatkan, menemukan atau memperoleh data yang diperlukan. Metode penelitian dapat dibedakan menjadi tiga (Leedy, 1980 *dalam* Kusnaka Adimiharja, 2002). yaitu:

1) Metode Historik
2) Metode Survey
3) Metode Eksperimen

Metode historik adalah pendekatan sistematis untuk menyelidiki dan memahami peristiwa masa lalu dengan mengumpulkan, menganalisis dan menafsirkan bukti-bukti sejarah. Hal ini dengan tujuan untuk membangun narasi yang koheren dan akurat tentang apa yang terjadi dan mengapa terjadidan bagaimana dampaknya sekarang.

Metode historik digunakan jika data yang diperlukan terutama berkaitan dengan masa lalu, sehingga teknik pengumpulan data yang digunakan terutama adalah studi dokumenter, wawancara juga dapat dilakukan jika ada pelaku sejarah yang bersangkutan masih hidup.

Metode survey adalah cara pengumpulan data primer dengan mengajukan pertanyaan kepada responden untuk memperoleh data yang ada pada saat penelitian dilakukan. Data dapat dikumpulkan melalui beberapa teknik, seperti wawancara dan pengamatan atau observasi. Metode servey ini dapat berupa servey deskriptif atau survey analitik.

Survai bertujuan untuk mendapatkan informasi mengenai berbagai aspek, seperti halnya opini, sikap, prilaku atau karakteristik dari responden dengan akurat. Metode ini sering digunakan dealam bidang ilmu sosial, pemasaran, dan lain sebagainya.

Metode Eksperimen adalah suatu cara pendekatan penelitian yang melibatkan manipulasi variabel independent, untuk mengamati pengaruhnya terhadap variabel dependen dalam suatu lingkungan yg terkendali. Tujuannya yaitu untuk mengidentifikasi hubungan sebab akibat antara kedua variabel tersebut.

Metode ini digunakan jika data yang diinginkan sengaja ditambahkan atau disorong munculnya. Dorongan atau rancangan untuk kemunculan data tersebut merupakan variabel bebas atau disebut juga perlakuan (treatment), jadi dalam eksperimen akan dicari hubungan sebab akibat antara variabel bebas dan variabel terikat.

Metode penelitian perlu dibedakan dari teknik pengambilan sample atau teknik pengumpulan data, yang merupakan teknik yang lebih spesifik untuk memperoleh data. Teknik pengambilan sampel adalah proses pengambilan sebagian dari sebuah populasi untuk dijadikan sebagai sampel dalam penelitian, yang diharapkan dapat mewakili keseluruhan populasi, dan sampel tersebut mempunyai peluang yang sama untuk terambil. Sedangkan Teknik pengumpulan data adalah metode atau cara yang digunakan untuk mengumpulkan informasi atau fakta yang

relevan dalam suatu penelitian. Teknik ini penting karena menentukan validitas data.

4.2 Teknik Penarikan Sampel

Teknik penarikan atau pengambilan sampel adalah proses pengambilan sebagian dari sebuah populasi untuk dijadikan sampel dalam penelitian, yang diharapkan dapat mewakili keseluruhan populasi. Jadi dalam hal ini perlu untuk mengetahui apa itu populasi dan juga sampel.

4.2.1 Populasi (*Population atau Universe*)

Sugiyono (2007), Kusnaka A (2002), Husen Umar (2000), Masri S dan Efendi E (1995) menyatakan, bahwa populasi adalah keseluruhan atau totalitas dari objek atau individu yang memiliki karakteristik yang sama dan menjadi subjek penelitian. Dalam kontek penelitian, populasi bisa berupa manusia, tumbuhan, hewan, benda, atau bahkan gejala atau peristiwa.

1) Populasi menurut objek nya yaitu objek psikologis bisa merupakan objek kongkrit (*tangable*) maupun objek abstrak (*untagable*)
2) Populasi menurut ukurannya atau ukuran populasi adalah banyaknya objek psikologis dalam populasi dan biasa secara umum diberi lambang N
3) Populasi dapat dibagi menjadi dua bagian, yaitu :
   - Populasi terhingga (*Finite population*)
   - Populasi tak terhingga (*Infinite population*)
4) Survey biasa digunakan dalam penelitian sosial yang meneliti tingkah laku (*behaviour*) dan populasi nya adalah populasi tehingga
5) Populasi sasaran (*target population*) adalah populasi yang nantinya akan menjadi atau mencakup kesimpulan penelitian. Populasi sasaran bisa berbentuk kelompok orang, hewan, atau objek yang menjadi focus utama dalam suatu penelitian

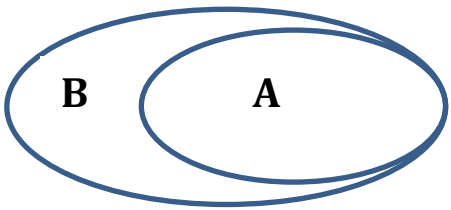
atau survai. Ini adalah keseluruhan kelompok yang ingin dipelajari oleh peneliti atau analis, dan darimana sampel sampel akan diambil. Populasi sasaran memiliki karakteristik yang jelas dan spesifik, yang membedakannya dari populasi umum.

Contoh: Rencana percobaan dengan populasi sasaran (*target population*), yaitu:

- Anak usia 5 - 10 tahun
- Jenis Kelamin laki - laki
- Berada di Tasikmalaya
- Kecamatan A
- Desa B

Maka sampling tidak sesukar dalam survey dengan populasi tak terhingga (*infinite*), kesimpulan bersifat umum, dan berlaku secara umum.

6) Populasi yang dipelajari (*study population*) adalah keseluruhan objek atau subjek penelitian yang memiliki karakteristik tertentu dan menjadi wilayah generalisasi dari hasil penelitian. Pengertiannya bisa juga seprti ini, bahwa populasi yang dipelajari adalah populasi yang mana tidak merupakan populasi secara keseluruhan.



Keterangan :

A = *Study Population*

A dan B = *Target Population*

7) Sensus adalah kegiatan mengumpulkan, mencatat, dan menghitung informasi secara sistematis tentang anggota suatu populasi, terutama dalam kontek demografi (kependudukan), ekonomi, atau social, yang dilakukan secara menyeluruh dan berkala oleh pemerintah atau Lembaga resmi.

Tujuan utamanya adalah untuk mendapatkan data yang akurat mengenaijumlah penduduk, karakteristik social ekonomi, dan seterusnya, yang penting untuk perencanaan pembanguana dan kebijakan publik. Atau sensus yaitu penelitian yang sifatnya menyeluruh, dimana setiap objek di dalam populasi diperiksa. Sensus tidak dapat dilakukan pada penelitian yang sifatnya menghancurkan (*destructive)*, sifat yang seperti itu biasanya menggunakan sampel.

4.2.2 Sampel atau Satuan Sampel (*Sampling Unit*)

Sugiyono (2007) dan Hermanto W.(1995) menyatakan, bahwa satuan sampel (*Sampling unit*) adalah individua atau unit dalam populasi yang menjadi dasar pengukuran atau pengamatan dalam suatu penelitianatau pengambilan sampel. Dalam katan lain, satuan sampel adalah elemen-elemen yang termasuk dalam sampel yang dipilih dari populasi yang lebih besar untuk mewakili karakteristik populasi tersebut.

Contoh: Apabila Indonesia dibagi menjadi 34 provinsi dan dalam penelitian ini provinsi yang akan dipilih, maka provinsi menjadi satuan sampel (*samling unit*).

4.2.3 Kerangka Sampel (*Sampling Frame*)

Kerangka sampel (*Sampling Frame*) adalah daftar atau kumpulan lengkap dari semua unit atau elemen yang ada dalam suatu populasi yang akan menjadi dasar

pengambilan sampel. Atau kerangka sampel adalah satuan sampel yang ada dalam sebuah populasi.

Contoh: Nama - nama provinsi yang 34 provinsi itu terdaftar dalam bentuk daftar.

4.2.4 Galat Baku (*Standart Error of the Estimator*) atau σ

Galat baku atau simpangan baku adalah akar daripada ragam atau varian (σ^2), dimana formulasi dari ragam atau varian adalah sebagai berikut :

$$\sigma^2 = \frac{\sum(Xi)^2 - (\sum Xi)^2 / N}{N}$$

Nilai dari simpangan baku (σ) biasa disebut juga ukuran presisi (*measure of precision*) yang menggambarkan bagaimana sebaran nilai statistik sekitar parameternya.

Presisi adalah tingkat ketepatan yang ditentukan oleh perbedaan hasil yang diperoleh dari sampel dibandingkan dengan hasil yang diperoleh dari pencacahan lengkap.

4.2.5 Rencana Sampel (*Sampling Plan*) dan Rancangan Sampel (*Sampling Design*)

Rencana sampel (*sampling plan*) adalah rencana terperinci yang menjelaskan bagaimana cara memilih dan mengambil sampel dari suatu populasi untuk keperluan penelitian atau analisis. Rencana sampel adalah sebuah gambaran garis besar yang menyangkut:

1) Penentuan populasi sasaran dan populasi penelitian

2) Penentuan bentuk dan ukuran satuan sampel
3) Penentuan ukuran sampel (n)
4) Penentuan cara memilih satuan sampel

Rancangan sampel (*sampling design*) adalah rencana sampel dan metode penaksiran atau analisis. Hal ini dilakukan untuk mendapatkan hasil yang dapat digeneralisasikan atau diterapkan pada seluruh populasi.

4.2.6 Tipe - Tipe Sampel

Sampel dengan proses memilih sebagai satuan sampel, dapat dilihat dari:

1). Dipandang dari cara memilihnya

- Sampel dengan pengembalian

- Sampel tanpa pengembalian

Catatan:

Sampel dengan pengembalian jarang digunakan, karena tidak efisien.

2). Dipandang dari besarnya peluang pemilihan

- Sampel non peluang (*Non Probability Sampling*)

- Sampel peluang (*Probability Sampling*)

Catatan:

Sampel non peluang sangat sederhana, tapi ada kekurangannya bahwa hasil penelitian tidak dapat digeneralisasikan ke seluruh populasi, karena bias pemilihan sampel, potensi bias lebih tinggi dibandingkan dengan sampel berpeluang. Secara umum sampel tidak berpeluang cocok untuk penelitian awal atau penelitian kualitataif, tetapi tidak

direkomendasikan untuk penelitian yang membutuhkan generalisasi statistik yang kuat.

4.3 Sampel Tidak Berpeluang (*Non Probability Sampling*)

Beberapa sampel yang termasuk Sampel tidak Berpeluang (*Non Probability Sampling*), yaitu (M. Nazir, 1999):

1) *Haphazard Sampling/ Portuitous Sampling/ Accidental Sampling*/ Sampel Seadanya
   Haphazard sampling atau sampel sembarangan adalah metode pengambilan sampel tidak berpeluang yang tidak mengikuti aturan atau prosedur tertentu dalam pemilihan unit sampel. Sampel diambil secara kebetulan atau seadanya, tanpa upaya untuk mencapai representasi yang akurat dari seluruh populasi. Metode ini biasa digunakan dalam situasi tertentu, seperti survai awal atau penelitian eksplorasi.
2) *Voluntary Sampling*/ Sampling Sukarela
   Voluntary sampling atau pengambilan sampel sukarela adalah metode pengambilan sampel, dimana individu secara sukarela berpartisipasi dalam penelitian, setelah diundang atau diminta untuk melakukannya. Partisipasi dalam metode ini memilih sendiri untuk ikut serta, biasanya dengan menanggapi kuesioner atau suervei yang disebarkan. Misalnya Penelitian Kedokteran, dll.
3) *Purposive Sampling/ Judgment Sampling/ Expert Sampling*
   Purposive sampling adalah Teknik pengambilan sampel dalam penelitian dimana peneliti memilih anggota sampel secara sengaja berdasarkan kriteria atau karakteristik tertentu yang dianggap relevan dengan tujuan penelitian. Ini bukan pengambilan sampel acak, akan tetapi berdasarkan pertimbangan dan penilaian

peneliti. Misal peneliti ingin meneliti efektivitas program pelatihan, mungkin akan memilih peserta yang mengikuti pelatihan tersebut, bukan orang yang belum pernah mengikuti pelatihan. (Penelitian atas dasar tertentu)

4) *Snowboll Sampling*/ Beruntun
Snowboll sampling adalah Teknik pengambilan sampel tidak berpeluang dimana peneliti memulai dengan sekelompok kiecil individu yang memenuhi kriteria penelitian dan kemudian meminta mereka untuk merekomendasikan individu lain yang memenuhi kriteria yang sama. Proses ini berlanjut , seperti bola salju yang menggelinding dan semakin besar, hingga sampel yang cukup terkumpul. (Penelitian tentang berkaitan dengan pemasaran produk tertentu, dll.)

5) Quota Sampling
*Quota samplin*g adalah Teknik pengambilan sampel tidak berpeluang dimana peneliti memilih sampel berdasarkan proporsi karakteristik tertentu yang telah ditentukan sebelumnya dari populasi, hingga mencapai quota yang ditettapkan. Misal peneliti ingin melibatkan 100 responden dengan perbandingan jenis kelamin 50:50. Maka peneliti akan menetapkan 50 persen laki-laki dan 50 persen wanita. (Penelitian pemasaran dan pengumpulan pendapat)

4.4 Sampel Berpeluang (*Probability Sampling*)

Beberapa sampel yang termasuk Sampel Berpeluang (*Probability Sampling*), yaitu (M.Nazir, 1999):

1) Simple Random Sampling (SRS)
2) Systimatic Sampling (SS)
3) Stratified Random Sampling (SRS)
4) Cluster Random Sampling (CRS)

4.4.1 *Simple Random Sampling* (SRS)

Simple Random Sampling (SRS) adalah proses memilih satuan sampel dari populasi sedemikian rupa, sehingga setiap satuan sampel dalam populasi mempunyai peluang yang sama besar untuk terpillih ke dalam sampel dan peluang itu diketahui sebelum pemilihan dilakukan. Sejalan dengan pendapat Sugiyano (2019), bahwa simple random sampling adalah pengambilan anggota sampel dari populasi dilakukan secarab acak tanpa memperhatikan strata yang ada dalam populasi itu.

Terdapat du acara dapat dilakukan untuk dalam simple random sampling, yaitu system kocokan dan menggunakan table acak. Berikut langkah-langkah untuk mendapatkan sampel acak:

- Buat kerangka sampel atau *Sampling Frame* adalah daftar lengkap dari semua satuan sampel, apapun bentuknya yang ada dalam populasi.
- Selanjutnya: Angka Random yaitu angka yang dipilih melalui mekanisme tertentu sedemikian rupa sehingga angka dari nol sampai dengan sembilan mempunyai peluang yang sama untuk terpilih.
- Menentukan Ukuran Sampel (n)

$n_0 = \{ (Z_{(1-\alpha)})(S) / \delta \}^2$

$n = \{ (n_0 / (1+(n_0 / N)) \}$

Keterangan : S = Simpangan baku variabel dalam populasi

δ = *bound of error* yang bisa di tolerir

$Z_{(1-\alpha)}$ = diperoleh dari tabel normal pada α Tertentu

4.4.2 *Systematic Sampling* (SS)

Systematic sampling adalah metode pengambilan sampel dimana elemen-elemen dipilih dari suatu populasi dengan interval tetap setelahn memilih titik awal secara acak. Proses pemilihan secara sistematik, ada dua cara pendekatan :

1). Pendekatan *Linear Systematic Selection* (LSS) digunakan kalau ukuran populasi (N) kelipatan tepat dari ukuran sampel (n)

2). Pendekatan *Circular Systematic Selection* (CSS), kerangka sampel merupakan sebuah lingkaran

Systematic Sampling digunakan apabila bisa disusun kerangka samplingnya yang lengkap dan keadaan variable relative homogen dan tersebar diseluruh populasi.

4.4.3 *Stratified Random Sampling* (SRS)

Dalam ilmu sosial tidak berhadapan dengan variable yang keadaanya homogen, dan dalam keadaan heterogenitas akan menyebabkan harus diambilnya ukuran sampel yang besar, apabila ingin dicari presisi tertentu.

Untuk meningkatkan presisi, peneliti harus membentuk sub -sub populasi yang relative homogen dan tujuan dari stratifikasi untuk memperoleh populasi yang relative homogen (presisi tinggi)

Menentukan Ukuran Sampel (n)

$$n = \{(\sum(N_{i^2}\ S_{i^2}/W_i)\}\ \{(\delta/Z_{(1-\propto)})^2 N^2 + \sum N_i S_{i^2}\}$$

Keterangan : S = Simpangan baku variable dalam populasi

δ = *Bound Of Error* yang bisa di tolerir

$Z_{(1-\alpha)}$ = Diperoleh dari tabel normal pada α tertentu

W = n_i / n

4.4.4 *Cluster Random Sampling* (CRS)

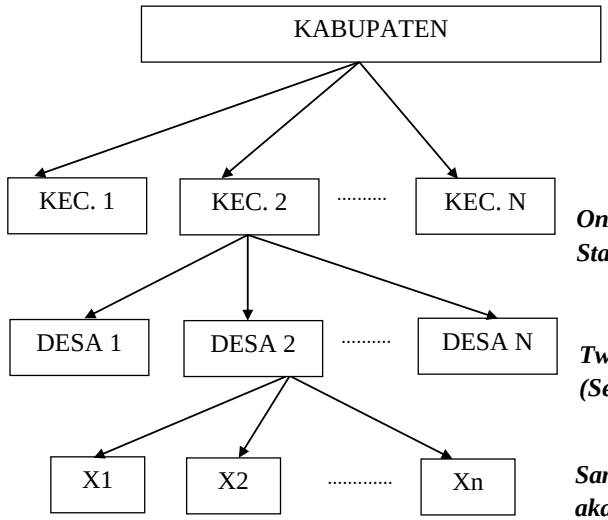
Cluster random sampling adalah metode pengambilan sampel dimana populasi dibagi menjadi beberapa kelompok (cluster), lalu secara acak dipilih beberapa cluster untuk dijadikan sampel. Jadi yang terpilih menjadi sampel bukan individu, melainkan keseluruhan kelompok atau cluster yang terpilih.

Satuan sampel bisa merupakan individu atau kelompok individu. Contoh Seorang peneliti ingin memeriksa:

- Tiap batang rokok (sampel adalah batang rokok)
- Keadaan setiap bungkus rokok (sampel adalah bungkus rokok)
- Slove rokok (sample adalah slove rokok)

Cluster adalah suatu satuan sampel yang di dalamnya berisi satuan - satuan sampel yang lebih kecil.

Contoh :



Menentukan Ukuran Sampel (n)

$n = \{(n_0) / (\bar{M}^2 + n0/N)\}$

$n_0 = \{(Z_{(1-\alpha)} S)/\delta\}^2$

Keterangan : S = Simpangan baku variable dalam populasi

δ = Bound Of Error yang bisa di tolerir

$Z_{(1-\alpha)}$ = Diperoleh dari tabel normal pada α tertentu

N = Satuan sampel primer (SSP)

M = Satuan sampel sekunder (SSS)

DAFTAR PUSTAKA

Husen Umar. 2000. *Metode Penelitian untuk Skripsi dan Thesis Bisnis.* PT. Raja Grafindo Persada. Jakarta

Sugiyono. 2007. Metode Penelitian Kuantitatif Kualitatif dan R&D. Penerbit Alfabeta. Bandung.

Harun Al-Rasyid. 1997. Teknik Penarikan Sampel dan Penyusunan Skala. Program Pasca Sarjana UNPAD. Bandung.

Kusnaka Adimihardja. 2002. Metode Penelitian Sosial. PT.Remaja Rosdakarya. Bandung.

Masri Singarimbun dan Sofian Effendi. 1995. Metode Penelitian Survai. PT. Pustaka LP3ES. Jakarta.

M. Nazir. 1999. Metode Penelitian. Ghalia Indonesia. Jakarta.

Hermanto Wasito. 1995. Pengantar Metodologi Penelitian. PT. Gramedia PustakaUtama. Jakarta.

BAB 5
UJI HIPOTESIS STATISTIK

Oleh Tedi Hartoyo

5.1 Pendahuluan

Hipotesis adalah pernyataan sementara yang diajukan sebagai jawaban atas suatu masalah penelitian, yang perlu diuji kebenarannya melalui penelitian (Steel and Torrie, 1981). Dalam kata lain, hipotesis adalah dugaan atau asumsi awal yang dibuat oleh sipeneliti, sebelum melakukan penelitian untuk kemudian dibuktikan dan pada akhirnya akan ada penerimaan atau penolakan dugaan tersebut. Hipotesis bukan merupakan kesimpulan akhir, melainkanjawaban sementara yang perlu diuji kebenarannya.

Hipotesis seringkali melibatkan hubungan antar dua variabel atau lebih, juga uji banding sebuah, dua buah atau lebih populasi atau perlakuan. Hipotesis terdiri dari dua bagian yaitu sebagai berikut:

1) yang pertama, hipotesis utama dalam penelitian, yang merujuk pada teori atau litelature yang ada, pengalaman dan observasi. Hipotesis pertama diberi simbol H_0.
2) yang kedua, hipotesis tandingan, yang merupakan kebalikan dari hipotesis pertama. Hipotesis kedua diberi simbol H_1.

Di dalam penelitian untuk membantu membuktikan hipotesis yaitu dengan menggunakan bantuan analisis statistik atau metode statistika. Tepatnya pemilihan metode uji adalah sangat penting untuk mendapat kesimpulan yang benar atau valid. Dalam dunia penelitian dua pendekatan utamanya adalah Uji Parametrik dan Uji Non-Parametrik. Kedua metode ini memiliki prinsip, asumsi, serta kelebihan dan kekurangan yang berbeda, sehingga memahami dulu karakteristik metodenya merupakan langkah awal sebelum menjalankannya dalam penelitian.

Ada beberapa perbedaan uji parametrik dan non parametrik. Sejalan dengan pendapat Siegel and Castellan (1988) dan Sudrajat S.W. (1984) menyatakan, bahwa Uji parametrik secara umum dapat digunakan jika data berasal dari populasi yang mengikuti distribusi normal. Uji non-parametrik, di sisi lain, lebih fleksibel, uji dapat digunakan bahkan jika data tidak memenuhi asumsi distribusi tertentu. Pemahaman perbedaan tersebut memungkinkan peneliti untuk memilih metode yang lebih sesuai dengan sifat data mereka.

Bab ini akan membahas konsep dasar, perbedaan, serta aplikasi dari kedua jenis uji tersebut dalam konteks umum statistik inferensial.

5.2 Uji Parametrik

5.2.1 Definisi dan Konsep Dasar

Steel and Torrie, 1981; Gomez and Gomez, 1984 dan Walpole, 1993) menyatakan, bahwa uji parametrik adalah metode uji statistik yang memerlukan asumsi, bahwa distribusi populasi dari mana sampel diambil, harus berdistribusi normal. Umumnya, uji ini harus memenuhi syarat bahwa data mengikuti distribusi

normal dan mempunyai varians yang homogen. Karena karakteristiknya yang berbasis distribusi, uji parametrik cenderung lebih kuat dalam mendeteksi perbedaan atau hubungan dibandingkan dengan metode non-parametrik

5.2.2 Asumsi Uji Parametrik

Sebelum menggunakan uji parametrik, terdapat beberapa asumsi yang harus dipenuhi, di antaranya:

a. Distribusi Normal: Data harus berdistribusi normal atau mendekati normal.
b. Homogenitas Varians: Varians antar kelompok harus sama.
c. Skala Pengukuran: Data harus diukur dalam skala interval atau rasio.
d. Independensi: Observasi harus independen satu sama lain.

Ketika semua asumsi ini terpenuhi, maka uji parametrik menjadi pilihan yang lebih optimal karena memberikan hasil uji yang akurat dan efisien.

5.2.3 Jenis-Jenis Uji Parametrik

1) Uji-t (*t-test*) dan Uji-Z

Uji-t dan Uji-Z mempunyai formulasi yang sama, yang membedakan hanya simbol untuk simpangan baku nya. Uji-t digunakan kalua data yang diolah adalah sampel dan Uji-Z digunakan kalua data yang diolah adalah populasi. Tabel yang digunakan juga berbeda yaitu Tabel-t Student dan Tabel Z (Walpole, 193).

Uji-t dan uji-Z dapat dimanfaatkan untuk membandingkan rata-rata sebuah populasi atau sebuah kelompok, atau membandingkan dua buah populasi atau dua buah kelompok data. Kelompok yang dibandingkan

bisa berupa sampel atau populasi. Jadi perbedaan terletak pada jumlah data yang diamati dan pengetahuan tentang varians populasi. Uji-t digunakan jika ukuran sampel kecil dan varians populasi tidak diketahui, sementara uji-Z digunakan ketika sampel berukuran besar dan varians populasi diketahui (Walpole, 1984; Gomez and Gomez, 194).

Sebagai contoh, seorang peneliti ingin mengetahui apakah metode pengajaran yang baru lebih efektif dibandingkan metode yang lama, dengan membandingkan hasil ujian dua kelompok siswa. Dalam kasus ini, uji-t independen dapat digunakan apakah terdapat perbedaan yang signifikan antara kedua kelompok siswa tersebut.

a) Digunakan untuk membandingkan rata-rata sebuah populasi (perlakuan).

Hipotesis: $H_0 : \mu = \mu_x$

$H_1 : \mu \neq \mu_x$

Pengujian Statistik sebagai berikut:

- Formulasi uji untuk Data Sampel

$$t_{hitung} = (X - \mu_x)/S_x$$

$$S = \sqrt{\{\sum(X_i - X)^2\}/(n-1)} \quad \text{dan} \quad S_x = S/n$$

Table yang digunakan nuntuk pembanding yaitu Tabel-t Student, pada taraf nyata α dan derajat bebas (n-1)

Keputusan: Jika $t_{hit.} < t_{tab.}$, maka H_0 akan diterima

Jika $t_{hit.} \geq t_{tab.}$, maka H_0 ditolak

- Formulasi uji untuk Data Populasi

$$Z = (X - \mu_x)/\sigma_x$$

$$\sigma_x = \sqrt{\{\sum(X_i - \bar{X})^2\}/N}$$

Table yang digunakan untuk membandingkan yaitu Tabel-Z untuk mendapatkan nilai peluang (p)

Keputusan: Jika p(peluang) ≥ α , maka H_0 diterima

Jika p(peluang) < α , maka H_0 ditolak

Keterangan:
X = Rata-rata pengamatan atau observasi
μ_x = Nilai dugaan
S_x = Simpangan baku sampel
δ_x = Simpangan baku populasi
n = Sampel
N = Populasi

Contoh: Menguji hasil produksi tanaman tertentu dengan menggunakan pupuk anjuran di suatu daerah, dan syarat yang disesuiakan dengan ketentuan (anjuran)

b) Digunakan untuk membandingkan rata-rata dua populasi (perlakuan).

- Dua buah populasi berpasangan

Hipotesis: $H_0 : \mu_A = \mu_B$

$H_1 : \mu_A \neq \mu_B$

Pengujian Statistik sebagai berikut:

$t_{hitung} = \{(X_A - X_B) - (\mu_A - \mu_B)\}/S_{(XA - XB)}$, untuk data sampel

$t_{hitung} = (X_A - X_B)/S_{(XA - XB)}$

$t_{hitung} = \bar{d}_i / S_d$

$$S_d = \sqrt{\{\sum(d_i - \bar{d})^2\}/n(n-1)}$$

Contoh Tampilan Data sebagai berikut:

X	Y	$d_i = (X - Y)$
X_1	Y_1	d_1
X_2	Y_2	d_2
.	.	.
.	.	.
.	.	.
Jumlah		$\sum d_i$

$$\bar{d}_i = (\sum d_i) / n$$

Table yang digunakan nuntuk pembanding yaitu Tabel t-Student, pada taraf nyata α dan derajat bebas (n-1)

Keputusan: Jika $t_{hit.} < t_{tab.}$, maka H_0 diterima

Jika $t_{hit.} \geq t_{tab.}$, maka H_0 ditolak

Keterangan:

$\bar{d}$ = Rata-rata selisih pengamatan atau observasi

S = Simpangan baku sampel

S_d = Simpangan baku rata-rata

n = Sampel

Contoh: Membandingkan dua buah populasi yang berpasangan, misalnya membandingkan pendapatan sebelum dan sesudah adanya penyuluhan.

- Dua buah populasi yang tidak berpasangan

Hipotesis: $H_0 : \mu_A = \mu_B$

$H_1 : \mu_A \neq \mu_B$

Pengujian Statistik sebagai berikut:

$t_{hitung} = \{(X_A - X_B) - (\mu_A - \mu_B)\}/S_{(XA-XB)}$, untuk data sampel

untuk ragam sama, simpangan baku nya sebagai berikut:

$$S_{(XA-XB)} = \sqrt{S_p\ (1/n_A + 1/n_B)}$$

$$S_p = \sqrt{\{(n_A-1)S^2_A + (n_B-1)S^2_B\}/\{(n_A-1)+(n_B-1)\}}$$

Table yang digunakan nuntuk pembanding yaitu Tabel-t Student, pada taraf nyata α dan derajat bebas (n_A+n_B-2)

Keputusan: Jika $t_{hit.} < t_{tab.}$, maka H_0 diterima

Jika $t_{hit.} \geq t_{tab.}$, maka H_0 ditolak

untuk ragam berbeda, simpangan baku nya sebagai berikut:

$$S_{(XA-XB)} = \sqrt{S^2_A/n_A + S^2_B/n_B}$$

t hitung akan dibandingkan dengan t*

dimana t* dapat dihitung sebagai berikut:

$$t^* = (t_A w_A + t_B w_B) / (w_A + w_B)$$

$w_A = S^2_A/n_A$ dan $t_A = t_{\alpha/2\ (db=nA-1)}$

$w_B = S^2_B/n_B$ dan $t_B = t_{\alpha/2\ (db=nB-1)}$

Keputusan: Jika $t_{hit.} < t^*$, maka H_0 diterima

Jika $t_{hit.} \geq t^*$, maka H_0 ditolak

Keterangan:

X_A = Rata-rata pengamatan atau observasi A

X_B = Rata-rata pengamatan atau observasi B

μ_A = Nilai dugaan A

μ_B = Nilai dugaan B

S_A = Simpangan baku sampel A

S_B = Simpangan baku sampel B

n = Sampel (n_A dan n_B)

Contoh: Membandingkan dua buah populasi yang tidak berpasangan, misalnya membandingkan hasil produksi padi dengan pengolahan lahan dan tanpa pengolahan lahan.

2) ANOVA (*Analysis of Variance*):

a) *One Way* (Rancangan Acak Lengkap)

Analisis *One Way* digunakan untuk membandingkan rata-rata lebih dari dua populasi (perlakuan) yang homogen.

Contoh: Membandingkan pertumbuhan tanaman dengan tiga dosis pemberian pupuk.

Hipotesis: $H_0 : \mu_1 = \mu_2 = \mu_3$

$H_1 : \mu_1 \neq \mu_2 \neq \mu_3$

Cara pengujian:

Tampilan data sebagai berikut:

	Pupuk			
	1	2	3	
	x11 x12	x21 x22	x31 x32	
	T1	T2	T3	

Uji statistik:

Jumlah Kuadrat (JK)

- JK perlakuan = $T_1^2/n_1 + T_2^2/n_2 + T_3^2/n_3 - FK$
- JK total = $x_{11}^2 + x_{12}^2 + \ldots\ldots + x_{31}^2 + x_{32}^2 + \ldots - FK$
- JK Sisaan = JK total - JK perlakuan
- FK = T^2/n dan $n = n_1 + n_2 + n_3$

Kuadrat Tengah (KT)

- KT perlakuan = JK perlakuan / db perlakuan
- KT Sisaan = JK sisaan / db sisaan

$F_{hit.}$ = KT perlakuan / KT sisaan

$F_{tab.} = F_{\alpha(\text{bd perlakuan; db galat})}$

Tabel Analisis of Variance (ANOVA)

Sumber keragaman	db (derajat bebas)	JK	KT	F hitung	F tabel
Perlakuan	t-1				
Sisaan	(n-1) - (t-1)				
Total	n-1				

Keputusan : Jika $F_{hit.} > F_{tab.}$, maka H_0 ditolak

Jika $F_{hit.} \leq F_{tab.}$, maka H_0 diterima

b) *Two Way* (Rancangan Acak Kelompok)

Analisis *Two Way* digunakan untuk membandingkan rata-rata lebih dari dua populasi (perlakuan) yang heterogen.

Contoh: Membandingkan pertumbuhan tanaman dengan tiga jenis pupuk pada tiga daerah yang berbeda.

Hipotesis Perlakuan: H_0: $\mu_1 = \mu_2 = \mu_3$

H_1: $\mu_1 \neq \mu_2 \neq \mu_3$

Hipotesis Blok(Kelompok): H_0 : $\beta_1 = \beta_2 = \beta_3$

H_1 : $\beta_1 \neq \beta_2 \neq \beta_3$

Tampilan data sebagai berikut:

	Pupuk			
	A	B	C	
1	x11	x21	x31	B1
2	x12	x22	x32	B2
3	x13	x23	x33	B3
4	x14	x24	x34	B4

	Pupuk			
	A	B	C	
5	x15	x25	x35	B5
Total	T1	T2	T3	T

Uji statistik:

Jumlah Kuadrat (JK)

- JK perlakuan = $T_A^2/n_A + T_B^2/n_B + T_C^2/n_C - FK$
- JK blok = $B_1^2/n_1 + B_2^2/n_2 + \ldots + B_5^2/n_5 - FK$
- JK total = $x_{11}^2 + x_{12}^2 + \ldots\ldots + x_{31}^2 + \ldots + x_{35}^2 - FK$
- JK sisaan = JK total - JK perlakuan
- FK = T^2/n dan $n = n_A + n_B + n_C = n_1 + n_2 + \ldots + n_5$

Kuadrat Tengah (KT)

- KT perlakuan = JK perlakuan / db perlakuan
- KT blok = JK blok / db blok
- KT sisaan = JK sisaan / db sisaan

F hitung perlakuan = KT perlakuan / KT sisaan

F hitung blok = KT blok / KT sisaan

F tabel = $F_{\alpha(\text{bd perlakuan; db galat})}$

F tabel = $F_{\alpha(\text{bd blok; db galat})}$

Tabel 5. 1 Analisis of Variance (ANOVA)atau Sidik Ragam

Sumber keragaman	db (derajat bebas)	JK	KT	F hitung	F tabel
Perlakuan	t-1				
Blok	r-1				

Sumber keragaman	db (derajat bebas)	JK	KT	F hitung	F tabel
Sisaan	(n-1)-(t-1)-(r-1)				
Total	n-1				

Keputusan : Jika $F_{hit.} > F_{tab.}$, maka H_0 ditolak

Jika $F_{hit.} \leq F_{tab.}$, maka H_0 diterima

3) Uji Korelasi Pearson:

Digunakan untuk mengukur besar hubungan linear antara dua variable, nilai korelasi tidak bisa digunakan untuk menduga, dengan kisaran nilai antara -1 dan +1

Hipotesis: $H_0 : \rho = 0$

$H_1 : \rho \neq 0$

Cara menguji:

Menghitung nilai korelasi (r)

$$r = JHK_{XY} / \sqrt{JK_X * JK_Y}$$

Uji statistik:

$$t \text{ hitung} = r\sqrt{(n-2)/(1-r^2)}$$

Keputusan: Jika $t_{hit.} < t_{tab.}$, maka H_0 diterima

Jika $t_{hit.} \geq t_{tab.}$, maka H_0 ditolak

Contoh: Hubungan antara tinggi badan bapak dengan tinggi badan anak.

4) Regresi Linear:

Digunakan untuk memodelkan hubungan antara variabel tidak bebas dengan variable bebas dan dapat digunakan untuk menduga atau meramal, juga dapat melihat perkembangan variable tidak bebas akibat variable bebas.

Bentuk hubungan:

$Y = a + bX$

$b = JHK_{XY} / JK_X$

Keterangan:

Y = Variabel tidak bebas
X = Variabel bebas
a = konstanta
b = Koefisien regresi

Hipotesis: $H_0 : \beta = 0$

$H_1 : \beta \neq 0$

Uji statistik:

Dengan Uji-t

$t = (b- \beta) / S_b$

$= b / S_b$

$S_b = \sqrt{(JK_Y - b\ JHK_{XY}) / (n-2)\ JK_X}$

Keputusan: Jika $t_{hit.} < t_{tab.}$, maka H_0 diterima

Jika $t_{hit.} \geq t_{tab.}$, maka H_0 ditolak

Dengan Uji-F

Tabel 5. 2 Analisis of Variance (ANOVA) atau Sidik Ragam

Sumber Keragaman	db	JK	KT	F hitung	F tabel
Regresi	1				
Sisaan	n-2				
Total	n-1				

Jumlah Kuadrat:

- JK regresi = b JHK_{XY}
- JK total = JK_Y
- JK sisaan = JK total - JK regresi

Kuadrat Tengah:

- KT regresi = JK regresi / db regresi
- KT sisaan = JK sisaan / db sisaan

F hitung = KT regresi / KT sisaan

F tabel = $F_{\alpha, db=N-2}$

Keputusan: Jika $F_{hit.} < F_{tab.}$, maka H_0 diterima

Jika $F_{hit.} \geq F_{tab.}$, maka H_0 ditolak

Contoh: Memodelkan hubungan dua buah variabel yaitu antara tinggi badan anak dengan tinggi badan bapak.

5.2.4 Kelebihan dan Keterbatasan

Kelebihan uji parametrik yaitu lebih powerful (memiliki daya deteksi yang tinggi) jika asumsi terpenuhi. Kekuatan statistik dan kemudahan interpretasi. Uji ini lebih mampu mendeteksi perbedaan atau hubungan yang nyata dalam data dan hasil uji lebih mudah dipahami, karena didasarkan pada parameter yang dikenal. Uji parametrik lebih efisien dan dapat diandalkan ketika data memenuhi asumsi yang mendasarinya, seperti distribusi normal dan homogenitas ragam.

Keterbatasan uji parametrik yaitu tidak dapat digunakan jika data tidak mengikuti sebaran normal atau homogenitas varians, untuk data yang tidak memenuhi asumsi tersebut atau data pada skala nominal dan ordinal dapat digunakan uji non parametrik.

5.3 Uji Non-Parametrik

5.3.1 Definisi dan Konsep Dasar

Siegel and Castellan (1988), Sudrajat S.W.(1984) dan Steel and Torrie (1981) menyatakan, bahwa uji non-parametrik adalah metode statistik yang tidak memerlukan asumsi distribusi populasi yaitu distribusi normal. Uji non parametrik lebih fleksibel dan dapat digunakan untuk data yang tidak memenuhi asumsi uji parametrik dan untuk data yang berukuran kecil.

5.3.2 Kapan Menggunakan Uji Non-Parametrik?

Uji non parametrik dapat dilakukan jika:

1. Data tidak berdistribusi normal.
2. Skala pengukuran data adalah nominal atau ordinal.
3. Ukuran sampel kecil.
4. Varians antar kelompok tidak homogen.

5.3.3 Jenis-Jenis Uji Non-Parametrik

1) Cuplikan Tunggal

- **Uji Binomium Uji**

Keterangan: Skala ukur nominal

Fungsi: Membandingkan dua katagori, apakah kedua kategori mempunyai proporsi yang sama atau apakah kedua kategori mempunyai perbandingan yang sama.

Cara pengujian :

Hipotesis

$H_0 = A = B$

$H_1 = A \neq B$

Jika besar cuplikan lebih dari 25 dinamakan cuplikan besar dan cuplikan kurang dari 25 dinamakan cuplikan kecil.

Uji statistik cuplikan kecil:

- Hitung masing-masing frekuensi kedua kategori, x ialah kategori dengan frekuensi terkeci dan n adalah jumlah frekuensi dari dua kategori.
- Penggunaan Tabel D (Lihat di buku Siedney Siegel, 1985)
- Hitung berapa peluang yang diperoleh dar Table D biasa disebut dengan probabilitas (p)
- Bandingkan dengan taraf nyata (α)
- Keputusan, tolak H_0 jika $p \leq \alpha$ dan terima H_1 jika $p > \alpha$

Uji statistic cuplikan besar:

- Hitung masing-masing frekuensi dari kedua kategori, berikan notasi x untuk salah satu frekuensi dari dua kategori tersebut.
- Pergunakan metode pendekatan normal, sebagai berikut:
- $Z = \{(x \pm 0.5) - n/2\}/0.5\sqrt{n}$, dengan catatan bahwa $(x+0.5)$ jika $x<0.5$ dan $(x-0.5)$ jika $x>o.5$
- Pergunakan Tabel A (Lihat di buku Siedney Siegel, 1985)
- Hitung berapa peluang yang diperoleh dari Tabel A
- Bandingan dengan taraf nyata (α)
- Keputusan tolak H_0 jika $p \leq \alpha$ dan terima H_1 jika $p > \alpha$

- **Uji Chi square (χ^2)**

Keterangan: Skala ukur nominal

Fungsi: Seperti halnya uji Binomium, akan tetapi untuk kategori lebih dari dua

Cara pengujian sebagai berikut:

Hipotesis

H_0 : A = B = C

H_1 : A ≠ B ≠ C

Pengujian apakah setiap kategori mempunyai frukuensi yang sama.

Analisis sebagai berikut:

	Kategori				Total
	A	B	C	D	
O_i(Observasi)	O_A	O_B	O_C	O_D	
E_i(Harapan)	E_A	E_B	E_C	E_D	
$O_i - E_i$					
$(O_i - E_i)^2/E_i$					

Ei untuk setiap kategori mempunyai nilai frekuensi (haraspan) yang sama besarnya dan dicari dengan cara sebagai berikut:

$E_i = \sum O_i/k$

Keterangan : E_i = nilai harapan

O_i = Nilai Observasi

k = Jumlah kategori

χ^2 hitung $= \sum (O_i - E_i)^2/E_i$, dengan db = k-1

Keputusan: $\chi^2_{\text{hit.}} \geq \chi^2_{\text{tabel}}$, maka Ho ditolak

$\chi^2_{\text{hit.}} < \chi^2_{\text{tabel}}$, maka Ho diterima

- **Uji Kalmogorov-Smirnov**

Keterangan: Skala ukur ordinal

Fungsi:

- Menguji kecenderungan kategori
- Pengganti uji χ^2 jika kategorinya tersusun menurut skala ordinal

Prinsip dasar analisis, yaitu memperbandingkan selisih peluan observasi dengan peluang teoritis dalam bentuk kumulatif terhadap peluang standard pada Tabel E.

Pilih tarafnyata dan N sesuai dengan penelitian yang kita hadapi,

Cara pengujian:

Hipotesis: H_0 : f1 : f2 : f3:……= 1 : 1 : 1 : ….

H_1 : f1 : f2 :f3:……≠ 1 : 1 :1 : ….

Misalkan banyak kategori 5 (Lima) yang diukur menurut skala ukur ordinal

	A	**B**	**C**	**D**	**E**
Frekuensi Observasi (O_i)	a	b	c	d	e
O_i Kumulatif	Oa	Ob	Oc	Od	Oe
Frekuensi Kumulatif O_i (S_{nx})	Oa/x	Ob/x	Oc/x	Od/x	Oe/x
Frekuaensi Kumulatif (F_{ox})	1/5	2/5	3/5	4/5	5/5
$F_{0x} - S_{nx}$	D_A	D_B	D_C	D_D	D_E

Keterangan: x = a + b + c + d + e

Oa = a; Ob = a + b; Oc = a+b+c; dst.

$D_A = (1/5) - (Oa/x)$; $D_B = (2/5) - (Ob/x)$; dst

Cari nilai diantara D_A sampai dengan D_E yang mempunyai nilai mutlak tertinggi

$$D_{max.} = | \text{max. } D_i |$$

Pergunakan Tabel E

Keputusan: $D_{max.} \geq D_{tabel}$, maka tolak H_0

$D_{max.} < D_{table}$, maka terima H_0

- **Uji Deret**

Keterangan: Skala ukur ordinal

Fungsi: Untuk menguji sederatan data yang terdiri atas dua kategori, apakah tersusun random ataukah sistematik.

Cara pengujian:

Hipotesis: H_0: Data tersusun random

H_1: Data tersusun sistematis

Pengujian Hipotesis:

Cuplikan kecil (n1 dan n2 < 20)

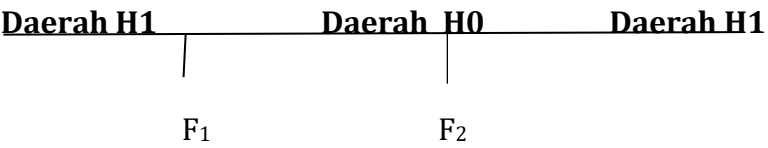
Menghitung nilai r (banyak deret)

Perhatikan contoh: Deret L dan P yang antri beli karcis, sebagai berikut :

r = 9; n1 = 8; n2 = 9^{δ}

n1 = L dan n2 = P

Pergunakan Tabel F1 dan F2



F_1 = Nilai-nilai batas terkecil r untuk menolak $H_{0\delta}$

F_2 = Nilai-nilai batas terbesar r untuk menolak H_0

Keputusan:

Jika r ada diantara F_1 dan F_2, maka Ho diterima, begitu pula sebaliknya

Jika $r < F_1$ dan $r > F_2$, maka H_0 ditolak

Cuplikan besar (n1 dan n2 > 20)

Menghitung nilai r (banyak deret) sama seperti pada sampel kecil

Uji statistika:

$$Z_r = (r - \mu_r)/\sigma_r$$

Dimana $r \sim N(\mu_r, \sigma_r)$

$\mu_r = (2\ n1\ n2)/(n1 + n2) + 1$

$\sigma_r = \sqrt{\{2n1n2(2n1n2 - n1 - n2)\}/(n1+n2)^2\ (n1+n2-1)}$

Pergunakan Tabel A

Keputusan: $Z_r \leq$ Z table, maka tolak H_0

$Z_r >$ Z table, maka terima H_0

2) Cuplikan Ganda

Dalam kasus ini harus ada dua kategori atau perlakuan yang akan diperbandingkan, dan dalam setiap

kategori diulang paling sedikitnya dua (2) kali untuk skala nominal, dan enam (6) kali untuk skala ordinal.

- **Cuplikan Ganda Berpasangan**

a) Uji Mc. Nemar

Keterangan: Skala ukur nominal

Fungsi: Digunakan untuk menguji signifikansi perubahan frekuensi "Sebelum" dan "Sesudah" suatu kegiatan.

Cara pengujian:

Hipotesis: H_0: p(A) = p(D)

H_1: p(A) ≠ p(D)

Data frekuensi disusun dalam bentuk table kontingensi 2x2, sebagaiberikut:

		Sesudah	
		-	+
sebelum	+	A	B
	-	C	D

Selanjutnya yang diperhitungkan dari tabel di atas adalah sel yang mengalami perubahan, yaitu:

A: Berubah dari + ke –

D: Berubah dari – ke +

Uji statistik :

$$\chi^2 = (|A-D| - 1)^2 / (A+D)$$

Pergunakan Tabel C, dengan db = 1

Kaidah keputusan: Jika $\chi^2_{hit.} \geq \chi^2_{tabel}$, maka tolak H_0

$\chi^2_{hit.} < \chi^2_{tabel}$, maka terima H_0

b) Uji Tanda

Keterangan: Skala ukur ordinal

Fungsi: Digunakan untuk menguji signifikansi dua keadaan atau perlakuan apakah ada perbedaan.

Data sifatnya kontinu, dan hanya diperlukan tanda mana dari dua keadaan tersebut yang lebih besar (>) dan lebih kecil (<).

Jika A > B, maka diberi tanda +

Jika A < B, maka diberi tanda –

Jika A = B, maka diberi tanda 0, dan tanda 0 akan dikeluarkan dari analisis.

Cara pengujian sebagai berikut:

Hipotesis: H_0 : p(+) = p(-)

H_1 : p(+) ≠ p(-)

Uji statistik:

Cuplikan kecil (tanda + dan tanda – kurang dan sama dengan dari 25) digunakan Uji Binomium dan pergunakan Tabel D

Cuplikan besar (tanda + dan tanda – lebih dari25) digunakan uji pendekatan normal:

$$Z = \{(x \pm 0{,}5) - 0{,}5n\}/0{,}5\sqrt{n}$$

Pergunakan Tabel A

Keputusan:

p hitung lebih besar dari taraf nyata (α), maka H0 ditolak

P hitung lebih besar dari taraf nyata (α), maka H0 diterima.

c) Uji Wilcoxon

Keterangan: Skala ukur ordinal

Fungsi: Uji ini hampir sama dengan uji tanda, yang membedakan adalah selain untuk signifikansi beda (A dan B), juga ingin diketahui besar beda rankingnya

Cara pengujian:

Hipotesis: H_0: $p(A) = p(B)$

H_1: $p(A) \neq p(B)$

Uji statistik:

Sususn data sebagai berikut, seperti di bawah table ini:

Pasangan	Data A	Data B	Selisih (d_i)	Rank (d_i)	Rank dengan tanda
1	A_1	B_1	$A_1 - B_1$		
2	A_2	B_2	$A_2 - B_2$		
3	A_3	B_3	$A_3 - B_3$		
.	.	.	.		
.	.	.	.		
.	.	.	.		
N	A_n	B_n	$A_n - B_n$		

Catatan rank ditentukan dengan tidak memperhatikan tanda + atau -, selanjutnya rank dengan tanda hanya memasukan tanda yang ada di selisih. Rank dengan tanda dijumlahkan untuk yang + dan –

Total rank tanda + dan – dijumlahkan dan diberi simbol T.

Jika N < 25, maka bandingkan nilai T dengan T_{table} pada Table G.

Jika N ≥ 25, maka digunakan uji dengan pendekatan normal:

$$Z_T = (T - \mu_T)/ \sigma_T$$

$$\mu_T = \{N(N+1)\}/4$$

$$\sigma_T = \sqrt{\{N(N+1)(2N+1)\}}/24$$

Keputusan:

Cuplikan kecil T hitung ≤ T table tolak H_0 dan T hitung > T table terima H_0 (Tabel G)

Cuplikan besar $p_{hit.} \leq \alpha$, maka tolak H_0 dan p hitung > α, maka terima H_0

Alternatif non-parametrik untuk uji-t berpasangan.

Contoh: Membandingkan tekanan darah sebelum dan setelah perawatan.

- **Cuplikan Ganda Tidak Berpasangan**

a) Uji Fisher

Keterangan: Skala ukur nominal

Fungsi: Digunakan untuk menguji adakah perbedaan antara dua keadaan atau perlakuan yang mungkin dari dua populasi yang berbeda.

Hipotesis: H_0 : p(X) = p(Y)

H_1 : p(X) ≠ p(Y)

Uji statistik:

Data disusun dalam bentuk tabel kontingensi 2x2

N tidak begitu besar kurang dari sama dengan 30

A	B	A+B
C	D	C+D
A+C	B+D	N

$p = \{(A+B)!(C+D)!(A+C)!(B+D)!\}/N!A!B!C!D!$

Keputusan: Jika $p \leq \alpha$, maka tolakm H_0

Jika $p > \alpha$, maka terima H_0

b) Uji Chi-Square (χ^2)

Keterangan: Skala ukur nominal

Fungsi: Uji ini hampir sama dengan uji Fisher, data observasi setiap sel lebih besar sama dengan 5, dapat digunakan untuk uji korelasi.

Hipotesis: $H_0 : p(X) = p(Y)$

$H_1 : p(X) \neq p(Y)$

Cara pengujian:

Data disusun dalam bentuk table kontingensi 2x2

Uji statistik:

- Untuk n < 20 kembali menggunakan uji Fisher
- Untuk $20 \leq N \leq 40$ pergunakan uji Chi-Square cuplikan ganda biasa
- Untuk N > 40 pergunakan uji Chi-Square cuplikan ganda dengan koreksi kontinuitas.
- N adalah total sampel

Uji Chi-Square cuplikan ganda biasa, data disusun dalam tabelm kontingensi 2x2, berikut:

O_{11} E_{11}	O_{12} E_{12}	$O_{1.}$
O_{21} E_{21}	O_{22} E_{22}	$O_{2.}$
$O_{.1}$	$O_{.2}$	N

Dimana: O ialah data observasi (frekuensi)

E ialah data frekuensi harapan

E dicari dengan cara sebagaiberikut: misalkan $E_{11} = (O_{1.} * O_{.1})/N$, dst,

$$\chi^2 = \sum\sum (O_{ij} - E_{ij})^2 / E_{ij}$$

Pergunakan Tabel C sebagai pembanding

Keputusan: χ^2 hitung $\geq$ χ^2 tabel, maka tolak H_0

χ^2 hitung $<$ χ^2 tabel, maka terima H_0

Uji Chi-Square untuk N > 40 adalah sebagaiberikut:

A	B	A+B
C	D	C+D
A+C	B+D	N

$$\chi^2 = N(|AD - BC| - N/2)^2 /(A+B)(C+D)(A+C)(B+D)$$

Pergunakan Tabel C sebagai pembanding

Keputusan: χ^2 hitung $\geq$ χ^2 tabel, maka tolak H_0

χ^2 hitung $<$ χ^2 tabel, maka terima H_0

c) Uji Mann-Whitney U

Keterangan: Skala ukur ordinal

Fungsi: Untuk menguji adakah perbedaan dua keadaan atau perlakuan

Hipotesis: H0 : X = Y

H1 : X ≠ Y

Cara pengujian:

Satu data dengan lainnya harus bias ditentukan ranking nya dan data bersifat continue.

Susun data dalam bentuk table berikut:

Data X			Data Y		
Nomor	Data asli	Ranking	Nomor	Data asli	Ranking
1	.	.	1	.	.
2	.	.	2	.	.
3	.	.	3	.	.
.	.	.	4	.	.
.	.	.	5	.	.
.	.	.	.	.	.
n_X	.	.	n_Y	.	.
		R_X			R_Y

Catatan:

- Berikan ranking dari nilai terkecil ke nilai terbesar
- nilai ranking dari 1 sampai dengan (n_X+n_Y)
- Jika ada rank kembar dibuat ranking rata-rata

- R_X ialah total ranking pada Data X dan R_Y ioalah total ranking pada Data Y

Uji statistik:

Hitung nilai μ_X, μ_Y dan μ

$\mu_X = n_X n_Y + n_X(n_X + 1)/2 - R_X$

$\mu_Y = n_X n_Y + n_Y(n_Y + 1)/2 - R_Y$

μ adalah nilai terkecil diantara nilai μ_X dan μ_Y

Kaidah keputusan:

- Jika n_X dan $n_Y \leq 8$, maka gunakan Tabel J. Cari nilai peluang (p) pada Tabel J, sesuaikan nilai μ dan α nya. Jika $p \leq \alpha$ tolak H_0 dan jika $p > \alpha$ terima H_0.
- Jika n_X dan n_Y antara 9 dan 20, maka gunakan Tabel K. Bandingkan μ dengan μ pada Tabel K (sesuai dengan α , n_X dan n_Y). Jika $\mu \leq \mu$ di Tabel K, maka tolak H_0 dan jika $\mu > \mu$ di Tabel K, maka terima H_0.
- Jika n_X dan $n_Y > 20$, maka pergunakan uji statistika untuk pendekatan normal.

$$Z_U = (\mu - \mu_U)/\sigma_U$$

Dimana $\mu_U = (n_X \, x \, n_Y)/2$

$$\sigma_U = \sqrt{(n_X \, x \, n_Y)(n_X + n_Y + 1)/12}$$

Pergunakan Tabel A, untuk menentukan peluang dari nilai Z_U yang telah dihitung. Jika $p \leq \alpha$ tolak H_0 dan jika $p > \alpha$ terima H_0.
Jika Ranking kembar terlalu banyak, maka perlu ada koreksi yaitu dengan mengkoreksi nilai σ_U

$$\sigma_U = \sqrt{\{(n_X n_Y)/N(N-1)\}\{(N^3 - N)/12 - \sum T\}}$$

$T = (t^3 - t) / 12$

Ada beberapa uji cuplikan tidak berpasangan lain nya seperti Uji Median, Uji Kalmogorov-Smirnov, Uji Wald-Wolfowitz dan Uji Moses

Alternatif non-parametrik untuk uji-t tidak berpasangan.

Contoh: Membandingkan skor kepuasan pasien antara dua rumah sakit.

3) Cuplikan Majemuk

- **Cuplikan Majemuk Berpasangan**

a) Uji Cockran Q

Keterangan Skala ukur nomional

Uji ini merupakan perluasan dari uji McNemar untuk kasus lebih dari dua kelompok dan dapat dianggap sebagai analog non parametrik dari ANOVA dengan pengukuran berulang untuk data biner.

Contoh: Seorang peneliti ingin mengetahui apakah ada perbedaan dalam preferensi konsumen terhadap tiga model kemasan produk. Dipilih 20 konsumen untuk memilih model tersebut yang paling mereka sukai, dan datanya adalah biner (0= tidak menyukai dan 1= menyukai). Uji Cockran Q dapat digunakan untuk menentukan apakah ada perbedaan yang signifikan.

Hipotesis: H0: tidak ada perbedaan

H1: adan perbedaan

Tabel 5. 3Tabel Tampilan Data

No	Model 1	Model 2	Model 3	R_i	R_i^2
1 2 . . . 20					
Colom	C1	C2	C3	ΣR_i	ΣR_i^2

Formulasi uji sebagai berikut:

$$Q = \frac{(k-1)\{k \Sigma Ci\text{^}2-(\Sigma Ci)\text{^}2\}}{k \Sigma Ri - \Sigma Ri\text{^}2}$$

Keterangan:

Q= Nilai untuk Cockran test

k= banyaknya kolom

Ci= Jumlah sukses dalam kolom ke j

Ri= jumlah sukses dalam baris ke i

Q dibandingkan dengan Chi-Kuadrat

Jika $Q \leq \lambda^2$, maka H_0 di terima

$Q > \lambda^2$, maka H_0 di tolak

b) Uji Friedman

Keterangan: Skala ukur ordinal

Fungsi: Menguji signifikansi perbedaan beberapa keadaan atau perlakuan yang berpasangan atau berkelompok.

Hipotesis: H_0 : A : B : C : D = 1 : 1 : 1 : 1

H_1 : A : B : C : D ≠ 1 : 1 : 1 : 1

Cara pengujian:

- Setiap keadaan atau perlakuan, diulang sama banyak
- Data harus dalam bentuk ranking
- Ulangan setiap keadaan atau perlakuan minimal dua.
- Susun data dengan keadaan atau perlakuan sebagai kolom, sedangkan ulangan sebagai baris.

Uji statistika:

- Susun data sebagai berikut:

	A		B		C		D	
	data	rank	data	rank	data	rank	data	rank
1 2 3 4 . N								
		R_A		R_B		R_C		R_D

- Buat ranking dari 1 sampai dengan k untuk setiap perulangan atau kelompok (k ialah banyaknya keadaan atau perlakuan)
- Hitunglah total ranking untuk setiap keadaan atau perlakuan, yaitu R_A, R_B, R_C, R_D
- Uji statistik sebagai berikut:

$$\chi^2_R = \{12/Nk(k-1)\}\{\sum R_j^2\} - 3N(k-1)$$

Keterangan;
K = banyak keadaan atau perlakuan
N = banyak ulangan
Rj = total rank setiap keadaan atau perlakuan (j = 1,2,3,….,k)

- Keputusan
  Jika k = 3 dan N = 2 sampai dengan 9
  k = 4 dan N = 2 sampai dengan 3

Gunakan Tabel N, dan tentukan nilai peluang dari nilai χ^2_R yang diperoleh, selanjutnya bandingkan p dengan taraf nyata, jika $p \leq \alpha$ tolak H_0 dan jika $p > \alpha$, maka terima H_0.
Jika k > 4 dan N lebih besar dari ketentuan Tabel N, maka gunakan Tabel C pada db = k-1

$\chi^2_R \geq \chi^2_{tabel}$, maka tolak H_0

$\chi^2_R < \chi^2_{tabel}$, maka terima H_0

Alternatif non-parametrik untuk ANOVA.

- **Cuplikan Majemuk Tidak Berpasangan**

a) Uji χ^2

Keterangan: Skala ukur nominal

Digunakan untuk data kategorik.

Contoh: Hubungan antara jenis kelamin dan preferensi pengobatan.

b) Uji Median dan Uji Kruskal-Wallis

Keterangan: Skiala ukur ordinal

- **Uji Kruskal-Wallis**

Fungsi: Digunakan untuk menguji perbedaan beberapa keadaan atau perlakuan, dari cuplikan yang tidak berpasangan.

Hipotesis: H_0: A : B : C : D = 1 : 1 : 1 : 1

H_1: A : B : C : D ≠ 1 : 1 : 1 : 1

Cara pengujian:

- Data dalam bentuk ranking
- Setiap keadaan atau perlakuan kemungkinan diulang tidak sama banyak nya
- Susun data, dengan keadaan atau perlakuan dalam bentuk kolom dan ulangan dalam bentuk baris
- Buatlah ranking dari 1 sampai dengan N

Dimana $N = n_A + n_B + n_C + n_D$

n_A = jumlah ulangan pada keadaan atau perlakuan A

n_B = jumlah ulangan pada keadaan atau perlakuan B

n_C = jumlah ulangan pada keadaan atau perlakuan C

n_D = jumlah ulangan pada keadaan atau perlakuan D

Uji statistik:

➢ Sususn data sebagai berikut:

A			B			C			D		
no	data	rank	no	data	rank	no	data	rank	no	data	rank
1			1			1			1		
2			2			2			2		
3			3			3			3		

A			B			C			D		
no	data	rank	no	data	rank	no	data	rank	no	data	rank
4 5			4 5 6 7			4			4 5 6		
		R_A			R_B			R_C			R_D

Dimana:

k = 4 keadaan atau perlakuan

$n_A = 5, n_B = 7, n_C = 4$ dan $n_D = 6$

$N = 5 + 7 + 4 + 6 = 22$

- Hitung jumlah rank di setiap keadaan atau perlakuan (R_A , R_B, R_C dan R_D)
- Pergunakan uji statistik nya

$$H = \{12/N(N+1)\}\{\sum R_j^2/nj\} - 3(N+1)$$

➢ Keputusan:

- Jika k = 3 dan $n_A, n_B, n_C, n_D \leq 5$, pergunakan Tabel O, hitung peluang (p) dengan mempertimbangkan niali H, n_A , n_B , n_C dan n_D
  $p \leq \alpha$ tolak H_0 dan jika $p > \alpha$ terima H_0.
- Jika k > 3 dan $n_A, n_B, n_C, n_D > 5$, pergunakan Tabel C dengan db = k -1

Jika $H \geq \chi^2$ tabel, maka tolak H_0 dan jika $H < \chi^2$ tabel, maka terima H_0 .

Alternatif non-parametrik untuk ANOVA.

Contoh: Membandingkan tingkat nyeri pada pasien dengan tiga jenis terapi.

3) Korelasi

- **Uji Koefisien Kontingensi**

Keterangan Skala ukur nominal

Fungsi: Menentukann nilai koefisien korelasi atau hubungan antara varibel X dan Y

Hipotesis: H_0 : Ada korelasi antara X dan Y

H_1 : Tidak ada korelasi antara X dan Y

Cara pengujian:

- Tabel kontingensi (k x r), untuk k dan r lebih besar dari 2, pergunakan nuji Chi-kuadrat cuplikan majemuk
- Table kontingensi (2x2), bisa ditentukan dengan uji Chi-kuadrat cuplikan ganda, uji Chi-kuadrat korelasi kontinuitas, uji Fisher

- **Uji Korelasi Spearman, Uji Kendal Rank**

Keterangan Skala ukur nominal

a) Uji Korelasi Spearman

Fungsi: Digunakan untuk mengukur derajat hubungan dua variable X dan Y dengan skala ordinal. Uji ini mempunyai kemampuan hamper sama dengan korelasi pearson, nilai koefisien ada di antara -1 dan +1 dan dengan skala ukur ordinal.

Hipotesis: H_0 : $\rho = 0$

H_1 : $\rho \neq 0$

Cara pengujian:

Susun data sebagai berikut:

No.	Data		Ranking		d_i	d_i^2
	X	Y	X	Y		
1 2 . N						
						$\sum d_i^2$

Data X di ranking dari 1 sampai dengan N

Data Y di ranking dari 1 sampai dengan N

Keterangan: N = Banyaknya cuplikan
d_i = Selisih ranking X dengan ranking Y

Uji Statistik:

- Tanpa rank kembar atau rank kembar hanya sedikit

$$r_s = 1 - (6 \sum d_i^2)/(N^3 - N)$$

- Rank kembar yang cukup banyak

$$r_s = (\sum x^2 + \sum y^2 - \sum d_i^2)/\sqrt{\sum x2} \cdot \sum y2$$

$$\sum x^2 = \{(N^3-N)/12\} - \sum T_x$$

$$\sum y^2 = \{(N^3-N)/12\} - \sum T_y$$

$$T_x = (t^3 - t)/12$$

$$T_y = (t^3 - t)/12$$

t = banyak kembaran data

Pengujian signifikansi:

Untuk $4 \leq N \leq 30$, Pergunakan Tabel P (eka arah)

Kaidah Keputusan:

Jika $r_s \leq r_s$ table, maka H_0 diterima dan jika rs > rs table, maka H_0 ditolak

Untuk N > 30, pergunakan Tabel B (Tabel t-Student)

Pergunakan Table-t Student (Tabel B)

Hitung nilai t_r

$t_r = r_s\sqrt{(N-2)/(1- r_s)}$

Kaidah Kepuitusan

Jika $t_r \leq t_{table}$, maka H_0 diterima dan jika $t_r > t_{table}$, maka H_0 ditolak

$T_{table} = t_{\alpha(N-2)}$

Alternatif non-parametrik untuk korelasi Pearson.

Contoh: Hubungan antara tingkat stres dan kualitas tidur.

4) Kelebihan dan Keterbatasan

Kelebihan uji non parametrik adalah kemudahan dalam penerapan karena tidak memerlukan asumsi distribusi, cocok untuk data dengan skala pengukuran rendah, seperti skala nominal dan ordinal. Juga lebih Tangguh untuk data dengan *outlier* atau ukuran sampel kecil.

Keterbatasan uji non parametrik adalah kurang *powerful* dibandingkan uji parametrik jika asumsi parametrik terpenuhi. Uji ini kurang kuat dalam mendeteksi perbedaan atau hubungan yang sebenarnya dalam data dibandingkan dengan uji parametrik.

5.4 Perbandingan Uji Parametrik dan Non-Parametrik

Pada Uji parametrik asumsi distribusi yaitu distribusi normal harus terpenuhi, sedangkan untuk uji non parametrik tidak memerlukan asumsi distribusi normal. Skala Pengukuran yang digunakan pada uji parametrik yaitu skala interval dan rasio, sedangkan untuk uji non parametrik yaitu skala nominal dan ordinal. Uji parametrik lebih *poweful* dari pada uji non parametrik. Untuk ukuran sampel uji parametrik lebihn efektif untuk sampel besar, sedangkan uji non parametrik cocok untuk sampel keci.

5.5 Pemilihan Uji yang Tepat

Pemilihan antara uji parametrik dan non-parametrik harus didasarkan pada:

1) Jenis data dan skala pengukuran.
2) Ukuran sampel.
3) Pemenuhan asumsi distribusi.
4) Tujuan penelitian.

5.6 Studi Kasus Uji Statistika

1) Contoh Uji Parametrik

- Studi: Membandingkan kadar kolesterol antara pasien yang diberi diet A dan diet B.

  Metode: Uji-t independen, jika data sampel dan uji Z jika data populasi.

  Hasil: Terdapat perbedaan yang signifikan, jika $p < 0,05$ antara kedua kelompok.

- Studi: Membandingkan hasil produksi padi varietas tertentu pada petani yang menggunakan pupuk organik dan pupuk an organik di Desa A.

  Metode: Uji-t independent, jika data sampel dan uji Z jika data populasi.

  Hasil: Terdapat perbedaan yang signifikan antara kedua kelompok, jika $p < 0,05$.

- Studi: Pengaruh faktor-faktor produksi terhadap hasil produksi jagung di Desa B.

  Metode: Analisis Regresi, uji-F dan uji-t

  Hasil: Terdapat perbedaan signifikan jika $p < 0,05$ antara kedua kelompok.

2) Contoh Uji Non-Parametrik

- Studi: Membandingkan tingkat keparahan gejala penyakit antara tiga kelompok pasien.

  Metode: Uji Kruskal-Wallis.

  Hasil: Tidak terdapat perbedaan yang signifikan antara ketiga kelompok ($p > 0,05$).

- Studi: Seorang peneliti ingin menguji apakah ada hubungan antara jenis kelamin dengan preferensi terhadap produk tertentu. Data kategori jenis kelamin (pria/wanita) dan preferensi produk (suka/tidak suka)

  Metode: Uji chi-square digunakan untuk melihat apakah ada hubungan independensi antara kedua variabel kategori tersebut.

  Hasil: Tidak terdapat perbedaan signifikan jika chi-square hitung lebih kecil dari chi square tabel.

- Studi: Melihat hubungan kinerja penyuluh dengan adopsi teknologi pengelolaan tanaman terpadu pada padi sawah.

Metode: Uji Korelasi Rank Spearman.

Hasil: Tidak terdapat hubungan yang signifikan, jika t_{hitung} lebih kecil dari t_{tabel}.

5.7 Kesimpulan

Uji parametrik dan uji non-parametrik adalah dua pendekatan penting dalam statistika inferensial. Pemilihan metode yang tepat tergantung pada karakteristik data dan tujuan penelitian. Uji parametrik lebih powerful jika asumsi terpenuhi, sedangkan uji non-parametrik lebih fleksibel dan dapat digunakan dalam berbagai kondisi terutama untuk ukuran sampel yang kecil.

DAFTAR PUSTAKA

Gomes, K.A. and A.A. Gomes. 1984. Statistical Procedures for Agricultural Research, 2 nd edition, an International

Rice Research Intitude book, A.Wiley-Intersci. Publ., John Wiley and Sons, New York-Chichester-Brisbane Toronto-Singapore.

Siegel, S., & Castellan, N. J. 1988. Nonparametric Statistics for the Behavioral Sciences. McGraw-Hill.

Steel R.G.D and Torrie J.H. 1981. Principles and Procedures of Statistics. A Biometrical

Sudradjat S.W. M. 1984. Mengenal Ekonometrika Pemula. Penerbit Armico, Bandung.

Walpolle, R, E . 1974. Introduction to Statistics. New York, Macmillan.

BIODATA PENULIS

Tedi Hartoyo, Ir. MSc.

Selain sebagai Dosen tetap di Program Studi Agribisnis Fakultas Pertanian Universitas Siliwangi Tasikmalaya. Penulis mengajar Statistika di Jurusan Kimia Analis BTH Tasikmalaya , Keperawatan Respati Tasikmalaya dan di keperawatan Gigi Tasikmalaya. Penulis lahir di Tasikmalaya tanggal 26 Juli 1962. Penulis adalah Dosen Program Studi Agribisnis Fakultas Pertanian Universitas Siliwangi. Menyelesaikan pendidikan S1 pada Jurusan Statistika IPB Bogor dan melanjutkan S2 pada Agriculture Development, Gent Unversity Belgium. Saat ini penulis mengampu beberapa mata kuliah, diantaranya Matematika, Statistika, Ekonometrika, Metode Penelitian Sosial Ekonomi dan Perancangan Percobaan.

BIODATA PENULIS

Drs. Bambang Suhartawan, M.MT
Dosen Jurusan Teknik Lingkungan
Fakultas Teknik Sipil dan Perencanaan
Universitas Sains dan Teknologi Jayapura

Penulis lahir di Blitar Tahun 1964. Penulis adalah dosen tetap dengan status PNS-DPK pada Jurusan Teknik Lingkungan Fakultas Teknik Sipil dan Perencanaan (FTSP) Universitas Sains dan Teknologi Jayapura. Menyelesaikan pendidikan S1 pada Program Studi Kimia Jurusan Matematika dan Ilmu Pengetahuan Alam (MIPA) Fakultas Keguruan dan Ilmu Pendidikan Universitas Cenderawasih Jayapura pada tahun 1989. Pendidikan Pasca Sarjana diperoleh dari Institut Teknologi Sepuluh Nopember (ITS) Surabaya jurusan Manajemen Teknik Lingkungan pada tahun 2002. Beberapa buku yang telah ditulis terkait bidang lingkungan antara lain : Pengelolaan Sumber Daya Air; Penyehatan Udara; Adaptasi dan Mitigasi Perubahan Iklim di Indonesia; Ekotoksikologi, Pengelolaan Limbah Industri; Sistem Lingkungan Industri; Analisis Kualitas Lingkungan; Kesehatan Lingkungan Bancana; Manajemen Penyakit Tripos Berbasis Lingkungan; Pengelolaan Limbah Padat, Limbah Industri dan B3; Lingkungan : Pemahaman Masalah dan

Solusi; Kesehatan Lingkungan : Memahami Dampak Lingkungan Terhadap Kesehatan; Kimia Ramah Lingkungan; Pencemaran Udara dan Perubahan Iklim; Entomologi dalam Kesehatan Lingkungan; Pengelolaan Limbah Cair; Kimia Lingkungan; Kesehatan Lingkungan Global; Pengenalan Pembangunan Berkelanjutan dan Memahami Konsep dan Teori Ekologi Lingkungan;

BIODATA PENULIS

Deasy Wahyuni, S.Si, M.Si
Dosen Program Studi Teknik Informatika
Fakultas Ilmu Komputer, Universitas Dumai

Penulis lahir di Dumai pada tanggal 16 Desember 1987. Penulis merupakan dosen tetap di Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dumai. Ia menyelesaikan pendidikan sarjana di Departemen Matematika, Universitas Riau dan melanjutkan studi magister di Departemen Matematika, Universitas Andalas. Bidang keahlian penulis adalah Matematika Terapan. Penulis memiliki minat khusus dalam pengembangan model pembelajaran berbasis digital, terutama dalam mata kuliah Statistik dan Analisis Data. Selama karirnya, penulis telah berkontribusi pada berbagai penelitian dan publikasi ilmiah terkait pengembangan metode pengajaran statistik yang didukung aplikasi, seperti R, SAS, dan SPSS, serta pendekatan digital dalam pengajaran.

Selain aktif dalam kegiatan akademik, Deasy Wahyuni juga terlibat dalam kegiatan pelayanan masyarakat dan pelatihan untuk meningkatkan literasi data bagi mahasiswa dan pelaku usaha mikro, kecil, dan menengah (UMKM) lokal. Buku ini ditulis sebagai upaya untuk memberikan panduan praktis dan aplikatif dalam memahami konsep dasar statistik, terutama bagi mahasiswa pemula di bidang teknologi informasi dan komputer.