# MANAJEMEN DAN ANALISIS DATA

KMT Lasmiatun
Solehudin
Maitri Anindita
Rusydi Fauzan
Suwito Pomalingo
Herda Ariyani
Ryryn Suryaman Prana Putra
Putri Rahmadani
Joni Wilson Sitopu
Sediyanto
Nova Suryani
Abdul Wahab
Adhar Arifuddin
Syamsu rijal
A Fahira Nur

ISBN 978-623-198-259-9

# MANAJEMEN DAN ANALISIS DATA

**KMT Lasmiatun**
**Solehudin**
**Maitri Anindita**
**Rusydi Fauzan**
**Suwito Pomalingo**
**Herda Ariyani**
**Ryryn Suryaman Prana Putra**
**Putri Rahmadani**
**Joni Wilson Sitopu**
**Sediyanto**
**Nova Suryani**
**Abdul Wahab**
**Adhar Arifuddin**
**Syamsu rijal**
**A Fahira Nur**

PT GLOBAL EKSEKUTIF TEKNOLOGI

# MANAJEMEN DAN ANALISIS DATA

**Penulis :**
KMT Lasmiatun
Solehudin
Maitri Anindita
Rusydi Fauzan
Suwito Pomalingo
Herda Ariyani
Ryryn Suryaman Prana Putra
Putri Rahmadani
Joni Wilson Sitopu
Sediyanto
Nova Suryani
Abdul Wahab
Adhar Arifuddin
Syamsu rijal
A Fahira Nur

**ISBN : 978-623-198-259-9**

**Editor :** Dr. Helmi Buyung Aulia Safrizal., ST. SE., M.MT
**Penyunting :** Diana Purnama Sar., SE. M.E
**Desain Sampul dan Tata Letak :** Atyka Trianisa, S.Pd

**Penerbit :** PT GLOBAL EKSEKUTIF TEKNOLOGI Anggota IKAPI No. 033/SBA/2022

**Redaksi :**
Jl. Pasir Sebelah No. 30 RT 002 RW 001
Kelurahan Pasie Nan Tigo Kecamatan Koto Tangah
Padang Sumatera Barat
Website : www.globaleksekutifteknologi.co.id
Email : globaleksekutifteknologi@gmail.com

Cetakan pertama, Mei 2023

# KATA PENGANTAR

Segala Puji dan syukur atas kehadirat Allah SWT dalam segala kesempatan. Sholawat beriring salam dan doa kita sampaikan kepada Nabi Muhammad SAW. Alhamdulillah atas Rahmat dan Karunia-Nya penulis telah menyelesaikan Buku Manajemen Dan Analisis Data ini.

Buku ini membahas Pengantar Manajemen dan Analisis Data, Pengenalan SPSS, Struktur Data Penelitian, Distribusi Frekuensi, Tranformasi data dan analisis univariat, Tranformasi data dan analisis univariat, Pengelompokan nilai variable dan pengembangan varaibel baru, analisis deskriptif variable kategorikal dan numerical Penyajian dan interprestasi analisis secara univariat, Analisis Bivariat, analisis data kategorikal dengan kai-kuadrat, penyajian dan interpretasi kai kuadrat, Analisis Bivariat. Analisis data katagorikal dengan ANOVA, Penyajian interpretasi analisis ANOVA, analisis regresi linier sederhana, panyajian interprestasi analisis, Structural Equation Model/SEM Analisis Multivariat, analisis regresi linier ganda, Analisis Mutivariat, analisis regresi logistik penyajian interpretasi analisis, Analisis Multivariat – Analisis Faktor, Analisis Mutivariat – Analisis Diskriminan.

Proses penulisan buku ini berhasil diselesaikan atas kerjasama tim penulis. Demi kualitas yang lebih baik dan kepuasan para pembaca, saran dan masukan yang membangun dari pembaca sangat kami harapkan.

Penulis ucapkan terima kasih kepada semua pihak yang telah mendukung dalam penyelesaian buku ini. Terutama pihak yang telah membantu terbitnya buku ini dan telah mempercayakan mendorong, dan menginisiasi terbitnya buku ini. Semoga buku ini dapat bermanfaat bagi masyarakat Indonesia.

Padang, Mei 2023

Penulis

## DAFTAR ISI

## DAFTAR GAMBAR

## DAFTAR TABEL

# BAB 1
# PENGANTAR MANAJEMEN DAN ANALISIS DATA

*Oleh KMT Lasmiatun*

## 1.1 Pendahuluan

Tingkat Produktivitas tidak hanya membutuhkan teknologi mutakhir dan karyawan yang terampil, tetapi juga kombinasi antara kreativitas dan kesuksesan untuk membuat keseluruhan sistem bekerja dengan baik. Peran Individu maupun kelompok bisa bekerjasama dengan baik adalah dasar dari produktivitas organisasi. Oleh sebab itu, upaya untuk memberikan kenyamanan bagi aktivitas individu dan kelompok merupakan tantangan bagi manajemen.

Upaya perencanaan, pengorganisasian, pengarahan, pengkoordinasian dan pengendalian tersebut dikenal dengan proses manajemen, tujuan dari proses tersebut adalah untuk mengelola sumber daya yang digunakan untuk mencapai tujuan organisasi.

Didalam organisasi data merupakan aset berharga. Data tersebut berisi informasi-informasi yang berguna untuk proses kelancaran operasi kegiatan bisnis yang dilakukan. Dari akses data, organisasi bisa memaksimalkan kegiatan manajemen. Manajemen data pada dasarnya digunakan untuk pengambilan keputusan dengan organisasi dengan cepat dan tepat. Jika analisis data dilakukan dengan tepat, kegiatan bisnis bisa mengalami perkembangan jauh lebih baik dari sebelumnya.

Dalam mengelola aset perusahaan memerlukan proses manajemen dan kemampuan untuk mengambil keputusan,

pemecahan masalah dan *action* untuk menggunakan sumber daya secara efektif dan efisien. Efisiensi adalah ukuran untuk memaksimalkan biaya sumber daya guna mencapai tujuan, yaitu hasil yang direalisasikan terhadap input yang dikonsumsi.

Efektivitas mengukur hasil dari suatu tugas atau pencapaian tujuan. Efektivitas berarti pemenuhan terget produksi baik kualitas ataupun kuantitas. Efisiensi dan efektifitas, memang penting untuk mencapai tujuan, tetapi harus diperhatikan pula penggunaan sumber daya, harus digunakan dengan bijak.

## 1.2 Pengantar Manajemen

### 1.2.1 Konsep Dasar Manajemen

Pada dasarnya manajemen sendiri berarti proses yang dilakukan bertujuan untuk mencapai tujuan organisasi. Proses dalam kegiatan manajemen adalah perencanaan, pengorganisasian, pengarahan, pengoordinasian dan pengawasan, pelaksanaan, pengawasan, pengarahan hingga evaluasi kegiatan.

Proses ini dilakukan agar tujuan yang hendak dicapai organisasi bisa berjalan dengan dengan efektif dan efisien. Organisasi yang dimaksud dalam pembahasan ini angap saja merupakan suatu badan usaha, meliputi pemasaran, produksi, personalia, keuangan dan administrasi akuntansi.

Proses perencanaan terdiri dari, perumusan tujuan, pengamatan lingkungan sekitar usaha, analisis faktor-faktor kelemahan dan kekuatan organisasi, analisis peluang, ancaman dan risiko dalam organisasi dilakukan dengan ilmu manajemen untuk menghasilkan keputusan.

Proses mengorganisasi terdiri dari, fungsi, hubungan dan struktur dalam perusahaan. Fungsi terdiri dari penentuan tugas karyawan dalam melaksanakan fungsi-fungsi dalam organisasi. Hubungan terdiri dari tugas, tanggung jawab dan laporan. Semua itu ditempatkan kedalam struktur organisasi untuk

memaksimalkan semua peran sumber daya yang ada pada organisasi.

Proses pengarahan yaitu pemberian intruksi, motivasi untuk menjalankan kegiatan. Dalam menjalankan kegiatan diperlukan komando demi kelancaran proses untuk mencapai tujuan organisasi.

Proses pengoordinasian bertujuan untuk menciptakan harmoni antara kegiatan dengan tujuan. Untuk mencapai tujuan yang sama diperlukan komunikasi serta kerjasama tim, disinilah fungsi koordinasi.

Proses pengawasan dan evaluasi dilakukan untuk melihat timbal balik antara proses telah dilakukan terhadap hasil yang kemudian dari hasil tersebut akan dipergunakan sebagai dasar mengambil keputusan di waktu mendatang.

Proses-proses manajemen yang telah dijelaskan diatas dilaksanakan sesuai dengan tingkatan-tingkatan dalam organisasi.

Manajemen berdasarkan peringkat dibagi menjadi 3 yaitu, manajemen tingkatan bawah (*supervisor*), menengah (*middle*), dan puncak (*top*). Sedangkan berdasarkan wawasan terdiri dari manajemen fungsional dan manajemen umum. Semuanya bertanggung jawab dalam proses manajemen organisasi.

Manajemen tingkat bawah memiliki tanggung jawab untuk melaksanakan tugas pada unit tertentu. Tugas *supervisor* lebih banyak di proses pengawasan, perencanaan, dan pengorganisasian. Manajemen tingat menengah memiliki tanggung jawab untuk melaksanakan tugas yang berkaitan dengan integrasi juga ikut andil dalam proses perencanaan, pengorganisasian dan pengawasan. Sedangkan manajemen puncak memiliki tanggung jawab yang lebih banyak dalam proses perencanaan dan pengorganisasian, namun memiliki peran yang sedikit dalam proses pengawasan.

### 1.2.2 Proses Manajemen

Menurut (Hariyanto, 2018) proses manajemen secara garis besar antara lain sebagai berikut.

1. **Perencanaan**, Rumusan rinci untuk mencapai beberapa tujuan akhir adalah kegiatan manajemen yang disebut perencanaan. Dengan demikian, perencanaan menyediakan kondisi untuk menetapkan tujuan dan mengidentifikasi cara untuk mencapainya.
2. **Pengendalian,** Pengendalian dan perencanaan hanyalah setengah dari proses manajemen. Setelah rencana dibuat, itu harus dilaksanakan. Manajer dan pekerja kemudian harus memantau implementasinya untuk memastikan rencana tersebut berjalan dengan baik.
3. **Pengambilan Keputusan**, artinya proses pengambilan keputusan diantara beberapa opsi. Fungsi ini merupakan penghubung antara perencanaan dan pengendalian. Manajer harus memiliki visi, keterampilan dan metode untuk mencapai keputusan yang diambil.

### 1.2.3 Fungsi Manajemen

Agar proses manajemen berjalan dengan lancar, diperlukan kemampuan untuk mengambil keputusan, memecahkan masalah dan mengambil tindakan agar dapat menggunakan sumber daya secara efisien dan efektif. Efisiensi adalah ukuran biaya sumber daya yang terkait dengan pencapaian suatu tujuan, yaitu hasil yang dicapai dibandingkan dengan input yang digunakan.

Efektivitas adalah ukuran hasil tugas atau pencapaian tujuan. Efektivitas kinerja berarti memenuhi sasaran produksi unit kerja, baik kuantitatif maupun kualitatif. Namun selain efektifitas juga diperlukan efisiensi, karena kita tidak hanya ingin mencapai tujuan, tetapi juga memperhatikan penggunaan

sumber daya secara tepat atau rasional (efisiensi). Untuk itu perlu diterapkan proses atau fungsi manajemen, yaitu:

1. Perencanaan, yaitu menentukan apa yang akan atau harus dicapai, menetapkan tujuan dan mengidentifikasi langkah-langkah (tindakan) untuk mencapainya.
2. Pengorganisasian, yaitu memaksimalkan SDM dan SDA untuk mencapai tujuan.
3. Pengarahan yaitu membuat pedoman kegiatan untuk mengarahkan pekerja untuk menindaklanjuti rencana yang hendak dilaksanakan.
4. Pengoordinasian, yaitu sinkronisasi tugas kelompok supaya tercipta satu tujuan yang sama.
5. Pengawasan. Pada fungsi ini dilakukan pemantauan kinerja, apakah hasil yang didapat sesuai dengan tujuan yang di capai. Proses pengawasan digunakan untuk meninjau hubungan timbal balik kinerja untuk dasar pengambilan keputusan, perubahan tujuan di masa mendatang.

## 1.3 Manajemen Analisis Data

Setelah mengetahui secara singkat mengenai manajemen lalu apa hubunganya manajemen dengan data? Lalu apa itu manajemen analisis data?

Sebelum membahas hubungan itu kita tilik dahulu apa itu pengertian manajemen analisi data.

Manajemen analisis data adalah rangkaian kegiatan untuk mengelola data hasil penelitian. Didalam organisasi data merupakan aset berharga. Data tersebut berisi informasi-informasi yang berguna untuk proses kelancaran operasi kegiatan bisnis dilakukan. Dari akses data, organisasi bisa memaksimalkan kegiatan manajemen. Manajemen data terdiri dari semua kebijakan, alat, dan prosedur guna menaikan ketersediaan data dalam hal hukum dan peraturan.

Manajemen data pada dasarnya digunakan untuk pengambilan keputusan dengan organisasi dengan cepat dan tepat. Jika analisis data dilakukan dengan tepat, kegiatan bisnis bisa mengalami perkembangan jauh lebih baik dari sebelumnya. Dari manajemen analisis data bisa diketahui customer behavior dan customer experience dapat juga digunakan untuk mengetahui tren yang sedang berkembang dan masih banyak kegunaan lainya. Seiring dengan perkembangan zaman analisis data sudah dimudahkan dengan kehadiran software analisis data, jadi tidak perlu lagi dilakukan secara manual.

Setelah mengetahui pengertian mengenai manajemen analisis data, lalu apa yang dimaksud data itu sendiri? Kenapa data begitu penting? Berikut penjelasanya.

### 1.3.1 Pengertian Data

Data dapat diartikan sebagai kumpulan informasi atau nilai yang diperoleh dari mengamati suatu objek, data dapat berupa angka dan dapat juga berupa simbol atau sifat. Contoh dari Beberapa jenis data antara lain: populasi, sampel, data observasi, data primer, dan data sekunder.

Pada hakekatnya, penggunaan data (setelah diolah dan dianalisis) merupakan landasan objektif dalam proses pengambilan keputusan/pembuatan kebijakan untuk memecahkan masalah yang dihadapi oleh pengambil keputusan. Keputusan yang baik hanya dapat dibuat oleh pembuat keputusan yang objektif dan didasarkan pada data yang baik.

Data yang baik adalah informasi yang handal, tepat waktu yang mencakup wilayah yang luas atau memberikan gambaran masalah secara keseluruhan adalah data yang terkait. Riset menghasilkan informasi. Terdapat tiga tingkatan data yaitu data mentah, hasil survey, data olahan berupa besaran, rata-rata, persentase dan data analitik berupa kesimpulan. Yang terakhir memiliki peringkat tertinggi karena memungkinkan perumusan

proposal atau proposal secara langsung sebagai dasar pengambilan keputusan. (SH Situmorang, 2014)

Data yang baik harus memiliki kriteria valid, reliabel dan obyektif. Valid artinya akurat sesuai dengan pengukuran yang dilakukan. Reliabel artinya tepat atau stabil, saat dilakukan penelitian berulang data yang didapat harus menghasilkan ukuran yang sama. Objektif artinya kesesuaian persepsi, contoh apabila obyek itu warnanya merah maka orang lain yang melihat juga menyatakan hal yang sama yaitu merah. (Sabri, 2010)

Analisis data adalah dari proses penelitian setelah memperoleh sepenuhnya semua informasi yang diperlukan untuk memecahkan masalah yang diteliti. Ketekunan dan ketelitian penggunaan alat analisis sangat menentukan proses penarikan kesimpulan, sehingga kegiatan analisis data menjadi hal tidak boleh diremehkan di dalam proses penelitian. Kecacatan dalam menganalisis dapat berakibat fatal pada kesimpulan, dengan implikasi yang lebih buruk lagi bagi kegunaan dan penerapan hasil penelitian. Oleh sebab itu, ilmu tentang berbagai teknik analisis merupakan sesuatu yang wajib dipahami oleh seorang peneliti, supaya penelitian yang dilakukan bisa memberi kontribusi bermanfaat dalam memecahkan masalah dan hasil tersebut dapat dijelaskan secara ilmiah. (Muhson, 2006)

### 1.3.2 Pembagian Data

Dalam Bukunya berjudul Analisis Data Untuk Riset Manajemen dan Bisnis (SH Situmorang, 2014) Pembagian Data dikelompokan sebagai berikut.

1. Sifat, dari sifatnya data dibagi menjadi dua :
    a) Data Kualitatif, data ini berbentuk deskriptif bukan berupa angka dan tidak bisa dihitung menggunakan rumus matematika, misalnya tingkah laku manusia, kualitas layanan. Data kualitatif di pecah lagi menjadi dua, yakni data nominal dan data ordinal. Data nominal:

data yang berada pada tingkat pengukuran data “paling rendah”. Apabila hasil pengukuran data hanya terdapat satu kategori, maka data itu disebut data nominal (data kategorikal). Data ordinal: Informasi urutan adalah informasi kualitatif, seperti halnya informasi nominal, tetapi pada tingkat yang lebih tinggi dari informasi nominal. Jika dalam data nominal semua data kategori dianggap sama, maka dalam data ordinal terdapat tingkatan data. Data ordinal memiliki urutan data yang lebih tinggi dan lebih rendah. Misalnya, persepsi suatu individu terhadap layanan tertentu bisa “sangat baik” atau “baik” data ordinal tidak bisa disamakan derajatnya “sangat baik” memiliki nilai lebih tinggi daripada “baik”. Persepsi menentukan hasil dari data ini.

b) Data Kuantitatif, data yang berbentuk angka, dan bisa dihitung mengunakan rumus matematika. Contohnya seperti bunga, pajak, pendapatan, dll. Data kuantitatif dipecah menjadi dua yaitu data interval dan data rasio. Data interval memiliki tingkat pengukuran data yang "lebih tinggi" dibandingkan data ordinal, hal tersebut dikarenakan urutan datanya bisa betingkat dan urutan datanya juga dapat dikuantifikasi. Data Rasio merupakan data yang memiliki tingkat pengukuran paling tinggi diantara tipe data lainnya. Data rasio sebenarnya adalah data numerik dalam arti yang sebenarya (tidak masuk kategori seperti data nominal dan ordinal) dan dapat diitung secara matematis. Ini berbeda dengan data interval dimana data rasio memiliki titik nol. Metode statistik nonparametrik umumnya digunakan untuk data nominal dan ordinal, sedangkan metode parametrik digunakan untuk data kuantitatif.

2. Sumber, menurut sumbernya data dibagi menjadi dua:
    a) Data Internal, adalah data yang terdapat dalam suatu suatu organisasi yang mencerminkan keadaan sebenarnya dari organisasi itu. Contohnya dalam suatu perusahaan terdapat jumlah pekerja, jumlah kas dan modal, jumlah produksi.
    b) Data Eksternal, adalah data dari luar organisasi yang memiliki kemungkinan untuk memberi gambaran pada organisasi tersebut. Contohnya: pendapatan mempengaruhi daya beli terhadap pembelian produk perusahaan.
3. Cara memperolehnya, dari cara memperoleh data dibagi menjadi dua:
    a) Data primer (*primary* data) adalah informasi yang diperoleh langsung oleh dikumpulkan oleh individu/organisasi secara tentang obyek penelitian dan untuk kepentingan penelitian yang bersangkutan. Misalnya melalui proses wawancara atau pengamatan obyek.
    b) Data sekunder (*secondary* data) adalah data didapat dari kumpulan penelitian sebelumnya yang disatukan atau dipublikasikan oleh berbagai pihak lain. contoh umumnya berupa dokumenter atau arsip.
4. Waktu, menurut waktu pengumpulannya data dibagi menjadi dua:
    a) Data *cross section* adalah data yang dikumpulkan saat periode waktu tertentu guna menggambarkan keadaan saat itu, contohnya adalah penelitian dengan menggunakan kuesioner
    b) Data *time series* data adalah data yang didapatkan dari waktu ke waktu untuk melihat perkembangan peristiwa selama periode tersebut, contohnya

perkembangan uang beredar, harga macam bahan pokok, pertumbuhan penduduk.

## 1.4 Megapa Manajemen Data Penting?

Manajemen data adalah proses mengumpulkan, menyimpan, melindungi, dan menggunakan data organisasi. Meskipun memiliki banyak sumber data yang berbeda saat ini, organisasi harus menganalisis dan mengintegrasikan data untuk mendapatkan kecerdasan bisnis untuk perencanaan strategis. Manajemen data mencakup semua kebijakan, alat, dan prosedur untuk meningkatkan ketersediaan data dalam konteks hukum dan peraturan.

Data dianggap sebagai sumber daya berharga dalam organisasi modern. Dengan akses ke sejumlah besar data dan berbagai jenis data, organisasi banyak berinvestasi dalam penyimpanan data dan infrastruktur manajemen. Organisasi menggunakan sistem manajemen data untuk melakukan operasi intelijen bisnis dan analitik data secara lebih efektif. Manfaat analisis data adalah:

1. Meningkatkan penjualan dan keuntungan. Analisis data membuka pemahaman yang lebih dalam tentang semua aspek bisnis. Dengan wawasan ini, manajer dapat mengoptimalkan sektor bisnis dan mengurangi biaya. Proses Analisis data juga dapat memprediksi dampak dari keputusan yang diambil di masa depan, membantu pengambilan keputusan dan perencanaan bisnis. Oleh karena itu, dengan meningkatkan teknik analisis datanya, perusahaan mengalami pertumbuhan penjualan dan laba yang signifikan.
2. Pengurangan inkonsistensi data
   *Silo* data merupakan kumpulan data mentah dari organisasi yang hanya bisa digunakan oleh satu departemen atau grup. *Silo* informasi bisa menciptakan

inkonsistensi keandalan pada hasil analisis data. Solusi manajemen data bermanfaat untuk menggabungkan informasi dan membuat tampilan informasi terpusat untuk meningkatkan kolaborasi antar departemen.

3. Memenuhi kepatuhan terhadap peraturan
   Undang-undang yang mengatur mengenai *General Data Protection Regulation* (GDPR) dan *California Consumer Privacy Act* (CCPA) memberi konsumen kendali terhadap perlindungan data mereka. Konsumen bisa mengajukan gugatan apabila organisasi tersebut:
   1) Mengambil data mereka tanpa izin
   2) Tidak mengelola lokasi dan menggunakan data dengan bijak.
   3) Menyimpan data mereka meskipun sudah diminta untuk menghapusnya.

   Dengan demikian, diperlukanya sistem dalam organisasi untuk memanajemen data yang baik, transparan, dan rahasia untuk menjaga akurasi.

# DAFTAR PUSTAKA


Hariyanto, S. 2018. '*Sistem Informasi Manajemen'*, *Sistem Informasi Manajemen*, 9(1), pp. 80–85.

Hastono, S. P. 2001. Analisis data. *Depok: Fakultas Kesehatan Masyarakat Universitas Indonesia*.

Muhson, A. 2006. '*Teknik Analisis Kuantitatif 1 Teknik Analisis Kuantitatif*', *Academia*, pp. 1–7. Available at: http://staffnew.uny.ac.id/upload/132232818/pendidikan/Analisis+Kuantitatif.pdf.

SH Situmorang, I Muda, M Doli, F.F. 2014. *Analisis Data Untuk Riset Manajemen dan Bisnis*. 3rd edn. Medan: USU Press.

Timotius, K. H. 2017. *Pengantar Metodologi Penelitian Pendekatan Manajemen Pengetahuan Untuk Perkembangan Pengetahuan*. Yogyakarta: Penerbit Andi.

_____. 2022. *Apa Itu Manajemen Data*. Dikutip dari https://aws.amazon.com/id (diakses pada 2 Maret 2023)

# BAB 2
# PENGENALAN SPSS

*Oleh Solehudin*

## 2.1 Pendahuluan

Penelitian melibatkan proses yang terorganisir, terencana, dan mengikuti prinsip-prinsip ilmiah untuk memperoleh hasil yang objektif dan rasional. Data digunakan atau diperlukan dalam penelitian dengan mengacu pada indikator tertentu yang telah ditentukan. Analisis data adalah tahap dalam pengolahan data penelitian kuantitatif (Priyatno, 2009). Definisi statistik adalah sebuah disiplin ilmu yang mempelajari teknik-teknik pengumpulan, penyajian, analisis data menggunakan metode-metode tertentu, serta penafsiran hasil analisis data tersebut (Sunjoyo *et al.*, 2012).

Ilmu statistik kerap diidentikkan dengan serangkaian angka-angka dan dikenal sebagai numerical description. Namun, sebenarnya data statistik merujuk pada semua jenis data yang dapat dikumpulkan dan diorganisir dengan metode tertentu. Statistik juga digunakan untuk melakukan berbagai jenis analisis, seperti pembuatan grafik, tabel, peramalan, dan pengujian hipotesis. Secara umum, ilmu statistik dapat dibagi menjadi dua bagian: Pertama, statistik deskriptif yang memberikan gambaran tentang berbagai karakteristik data, seperti rata-rata, variasi data, median, dan sebagainya. Kedua, statistik inferensial yang digunakan untuk membuat inferensi terhadap populasi dengan memanfaatkan data sampel, seperti melakukan estimasi terhadap ukuran populasi, pengujian hipotesis, peramalan, dan lain-lain (Santoso, 2016).

## 2.2 Konsep Dasar SPSS

SPSS merupakan sebuah perangkat lunak yang digunakan untuk memproses data statistik. Menurut Priyatno (2009), program ini termasuk dalam jenis perangkat lunak statistik. SPSS sendiri merupakan singkatan dari "Statistical Product and Service Solution (Sunjoyo *et al.*, 2012).

SPSS merupakan program yang populer dalam melakukan analisis statistik di bidang ilmu sosial. Banyak pengguna dari program ini termasuk peneliti pasar, perusahaan survei, peneliti kesehatan, pemerintah, peneliti pendidikan, organisasi pemasaran dan lainnya. Awalnya, Nie, Bent, dan Hull (1970) melengkapi SPSS dengan manual, yang diakui sebagai salah satu "buku sosiologi yang paling berpengaruh". Selain fitur analisis statistik, SPSS juga menawarkan fitur manajemen data seperti pemilihan kasus, pengorganisasian ulang file, pembuatan data turunan, dan dokumentasi data yang mencakup kamus metadata yang disimpan dalam file data (Abdillah, 2022).

IBM SPSS memiliki beberapa fitur, di antaranya IBM SPSS *Data Collection* untuk mengumpulkan data, IBM SPSS Statistics untuk menganalisis data, IBM SPSS Modeler untuk memprediksi tren, dan IBM Analytical Decision Management untuk membantu dalam pengambilan keputusan. Untuk menginstal versi terbaru program ini di komputer Windows, diperlukan spesifikasi minimum seperti prosesor Intel atau AMD dengan kecepatan 1 GHz, memori (RAM) sebesar 1 GB, resolusi monitor minimal 1024x768 piksel, dan harddisk dengan kapasitas kosong minimal 800 MB. Setelah melakukan analisis, hasilnya akan muncul dalam SPSS Output Navigator. Kebanyakan prosedur dalam Base System menghasilkan pivot table, yang dapat diubah tampilannya sesuai kebutuhan. Hal ini memungkinkan pengguna untuk menyesuaikan hasil output yang dihasilkan oleh SPSS agar sesuai dengan kebutuhan mereka (Budiyanto, 2018).

### 2.2.1 Sejarah SPSS

SPSS dahulu digunakan sebagai perangkat lunak untuk mengolah data statistik dalam ilmu sosial. Namun, seiring berjalannya waktu, SPSS mengalami kemajuan dan semakin kompleks dalam digunakan dalam berbagai disiplin ilmu seperti ekonomi, psikologi, pertanian, teknologi, industri, serta kesehatan dan bidang lainnya (Priyatno, 2009). SPSS merupakan salah satu alat riset dan pelaporan yang dapat digunakan bersama dengan berbagai fungsionalitas lainnya dalam pemrograman statistik (Rafiudin and Saepudin, 2009).

Dahulu, SPSS merupakan singkatan dari "*Statistical Package for the Social Sciences*" dan diciptakan khusus untuk memproses data statistik dalam ilmu sosial. Namun, saat ini kemampuan SPSS telah diperluas dan dapat digunakan oleh berbagai pengguna seperti di pabrik, riset ilmu sains dan sebagainya. Kini, SPSS yang dulunya disebut sebagai singkatan dari "Statistical Package for the Social Sciences" berubah menjadi "Statistical Product and Service Solutions". SPSS dapat membaca berbagai jenis data atau data dapat langsung dimasukkan ke dalam SPSS Data Editor. Apapun struktur dari file data mentahnya, data dalam Data Editor SPSS harus diatur dalam bentuk baris dan kolom. Baris (cases) berisi informasi untuk satu unit analisis, sedangkan kolom (variables) berisi informasi yang dikumpulkan dari masing-masing kasus (Budiyanto, 2018).

SPSS, singkatan dari Paket Statistik untuk Ilmu Sosial, awalnya dirilis pada tahun 1968. Dikembangkan oleh Norman H. Nie dan C. Hadlai Hull, SPSS telah menjadi salah satu software analisis data paling populer di kalangan peneliti dan statistikawan. Norman Nie sendiri merupakan seorang ilmuan politik pasca sarjana di Stanford University yang pada saat itu sedang melakukan Riset Profesor di Departemen Ilmu Politik di Stanford dengan seorang Profesor Emeritus Ilmu Politik di University of Chicago (Abdillah, 2022).

Pada tahun 1998, SPSS diperkenalkan untuk pertama kali. Setelah itu, pada tahun 2009, IBM Corporation membeli SPSS dan menggabungkannya dengan software IBM Analytic. Sejak saat itu, perangkat lunak ini dikenal dengan nama IBM SPSS Statistics. SPSS dikembangkan menggunakan bahasa pemrograman Java dan dapat digunakan pada sistem operasi Microsoft Windows, Linux, dan Mac OS. Selain itu, SPSS juga dapat diintegrasikan dengan bahasa pemrograman R, Microsoft.NET, dan Python untuk digunakan lebih lanjut (Advernesia, 2021).

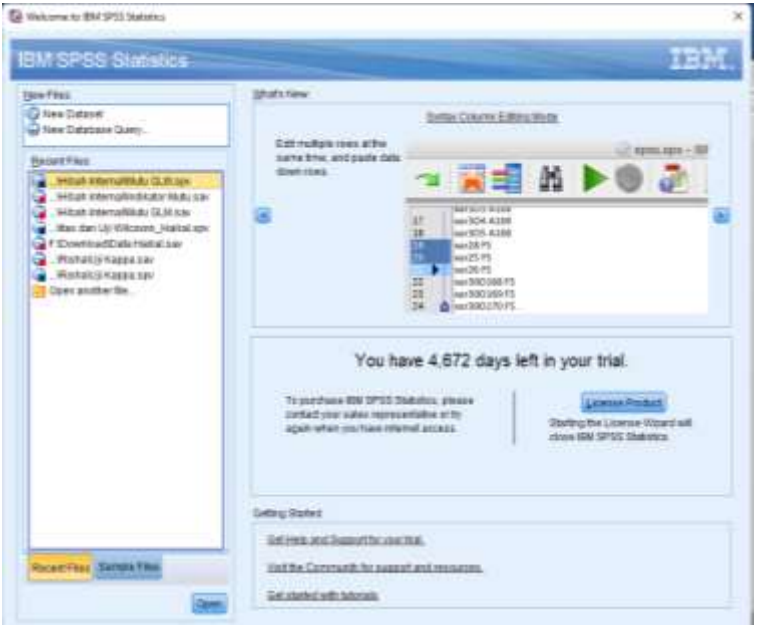
### 2.2.2 Pengertian SPSS

SPSS adalah sebuah aplikasi analisis statistika yang dapat melakukan analisis data menggunakan algoritma machine learning, analisis string, dan analisis big data yang dapat diintegrasikan untuk membangun platform analisis data yang kompleks. SPSS merupakan kependekan dari Statistical Package for the Social Sciences. Aplikasi ini telah dikenal luas di kalangan peneliti dan ahli statistik karena kemampuannya untuk melakukan perhitungan terkait analisis data. SPSS menyediakan perpustakaan untuk perhitungan statistika dengan antarmuka interaktif, sehingga membuatnya menjadi perangkat lunak analisis data yang sangat populer di berbagai universitas, instansi, dan perusahaan yang membutuhkan kemampuan analisis tingkat lanjut (Advernesia, 2021).

SPSS merupakan software aplikasi yang mampu melakukan analisis statistik dengan kemampuan tinggi dan memiliki fitur manajemen data pada lingkungan grafis. SPSS dilengkapi dengan menu-menu deskriptif dan kotak dialog yang sederhana sehingga mudah dipahami dan dioperasikan. Cara penggunaannya sangat mudah, cukup dengan mengklik dan memilih opsi yang diinginkan menggunakan mouse (Abdillah, 2022).

### 2.2.3 Pengenalan Fitur SPSS

1. Tampilan Awal
   a. Dialog IBM SPSS Statistics




**Gambar 2.1.** Dialog IBM SPSS

Ketika pengguna membuka perangkat lunak SPSS, secara default akan muncul tampilan awal. Tampilan ini dilengkapi dengan beberapa kolom shortcut., yaitu:

1) *New Files*: untuk membuat file baru.
2) *Recent Files*: untuk membuka file yang terakhir digunakan.
3) *What's New*: untuk menampilkan fitur baru yang disediakan SPSS.
4) *Modules and Programmability*: agar dapat melihat daftar modul dan program yang terpasang dan

belum terpasang pada SPSS, kita dapat melakukan pengecekan.

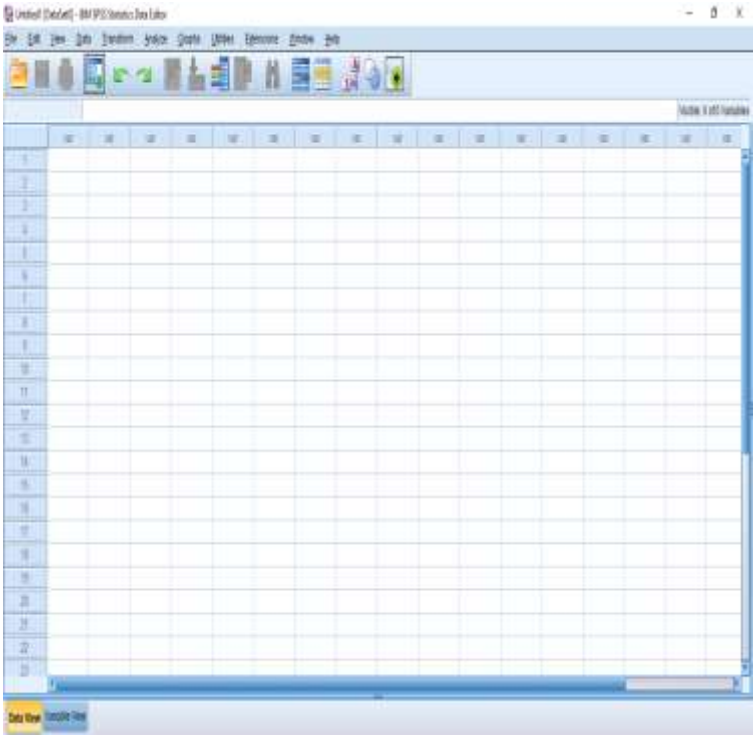
5) *Tutorials*: untuk membuka tutorial yang disediakan perangkat lunak SPSS.
6) Checkbox *Don't show this dialog in the future*: fungsinya adalah untuk menghentikan tampilan dialog IBM SPSS Statistics saat perangkat lunak dibuka SPSS (Advernesia, 2021).

b. Data Editor

Pada perangkat lunak SPSS, jendela utama yang terdiri dari Data View dan Variable View disebut Data Editor. Tampilannya menyerupai spreadsheet pada Microsoft Excel, di mana data dan variabel ditampilkan. Data View menampilkan isi dataset yang sedang dibuka, sedangkan Variable View memberikan informasi terkait karakteristik dan atribut lain dari variabel (Advernesia, 2021):

1) Data View

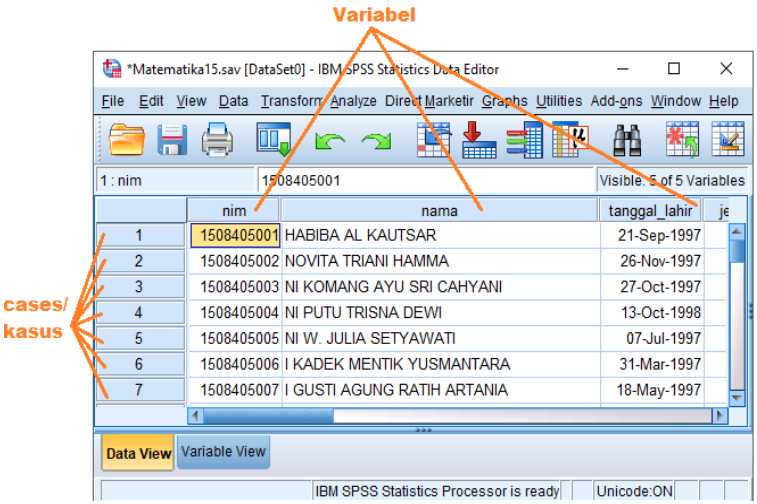
Tampilan lembar kerja SPSS yang menampilkan data dan variabel dalam suatu tabel disebut Data View. Pada tampilan ini, setiap baris mewakili sebuah kasus (*case*), dan setiap kolom mewakili sebuah variabel. Untuk memasukkan data ke dalam program SPSS, kita dapat menggunakan tampilan Data View.



**Gambar 2.2.** Data View

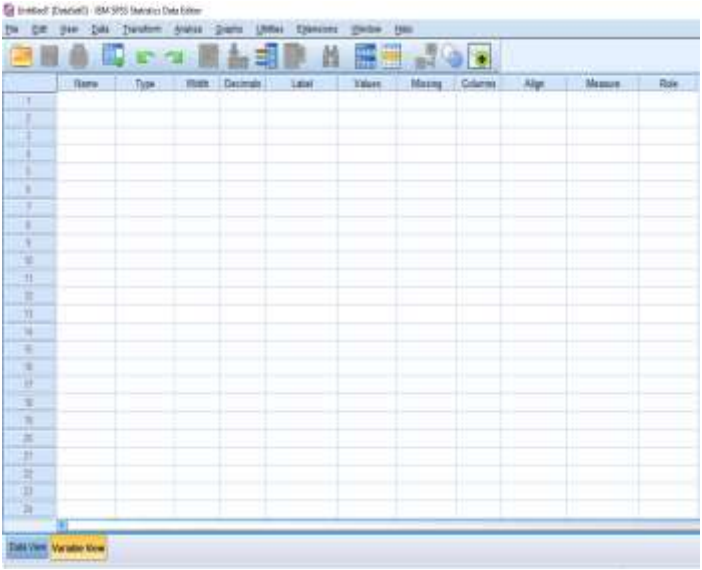
Berikut adalah penjelasan mengenai konsep variabel dan kasus yang terdapat pada tampilan Data View.

a) Kasus (*cases*); dalam SPSS, suatu hasil pengamatan pada sebuah objek dapat diwakili oleh pengamatan yang didapat dari observasi atau eksperimen. Contohnya adalah hasil survei dalam sebuah penelitian, data mahasiswa yang terkumpul, dan sebagainya.
b) Variabel; Merupakan ciri, karakteristik, atau ukuran yang digunakan untuk menjelaskan sebuah kasus (*case*). Beberapa contohnya adalah usia, nama, tingkat pendidikan, dan sebagainya.




**Gambar 2.3.** Tampilan Data View
Sumber: (Advernesia, 2021)

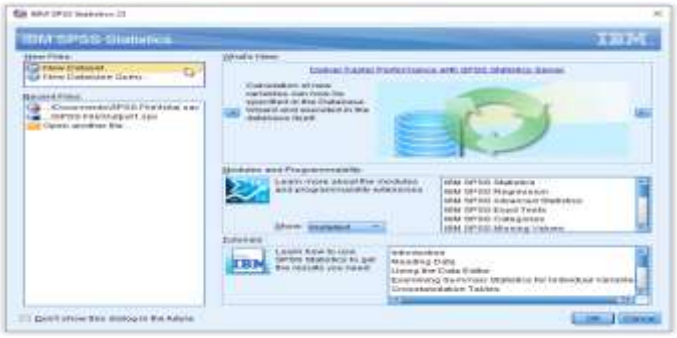
2) Variabel View
Variable View pada SPSS merupakan tampilan spreadsheet yang digunakan untuk mengelola variabel dalam SPSS, termasuk dalam pembuatan dan pengeditan variabel. Semua variabel yang digunakan dalam SPSS dapat dilihat melalui tampilan Variable View. Terdapat beberapa opsi yang tersedia di dalam Variable View, seperti name, type (tipe variabel), width, decimals, label, values, missing, columns, align, measure, dan role. Jendela Variable View pada perangkat lunak SPSS menampilkan informasi terkait karakteristik, atribut, dan lain-lain dari variabel tersebut. Sebelum membuat variabel, disarankan bagi pengguna untuk memahami tipe data (measure) yang ada pada SPSS.



**Gambar 2.4.** Variabel View

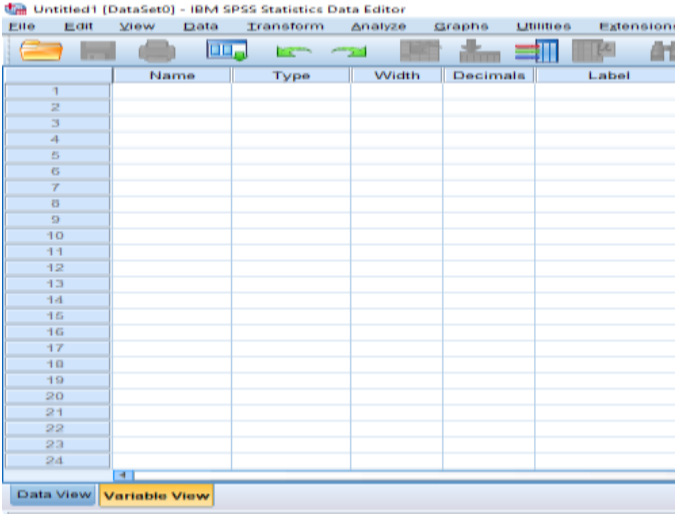
Berikut ini adalah cara membuat variabel pada SPSS:

a) Buka perangkat lunak SPSS



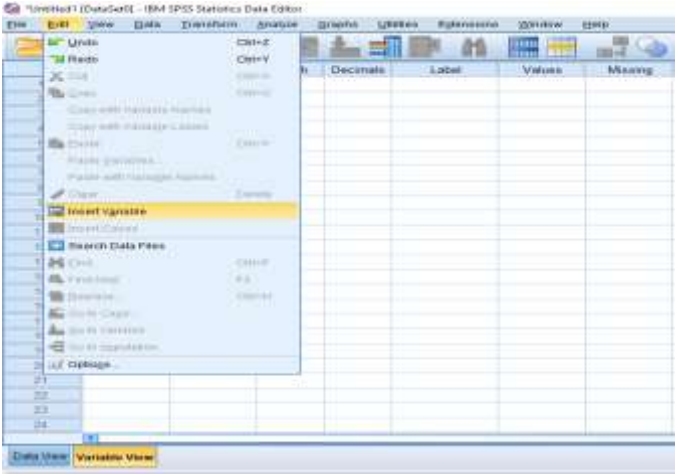
**Gambar 2.5.** Tampilan awal
Sumber: (Advernesia, 2021)

b) Arahkan lembar kerja ke Variable View



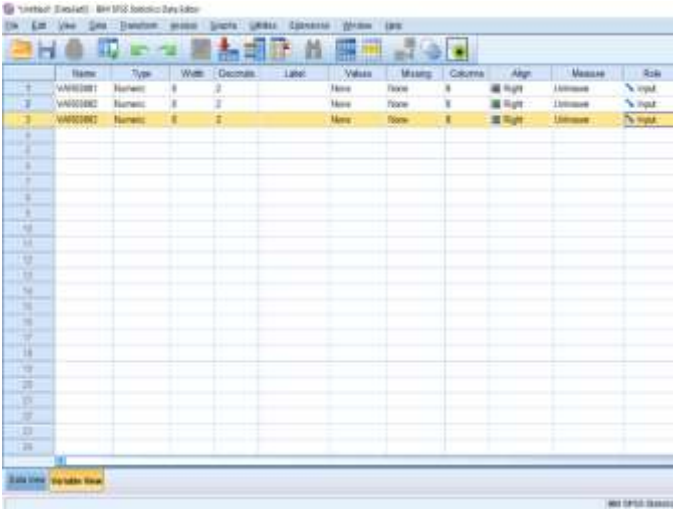

**Gambar 2.6.** Variabel View

c) Menyisipkan Variabel Edit › Insert Variable




**Gambar 2.7.** Insert Vriabel

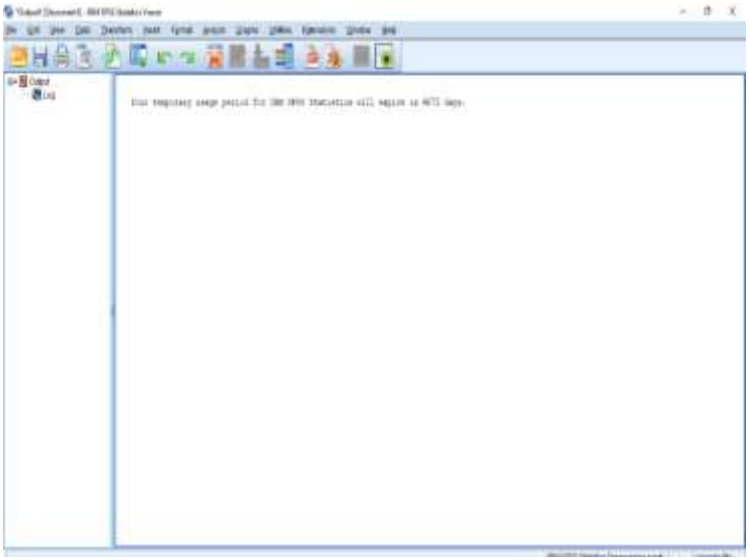
d) Variabel telah didefinisikan



**Gambar 2.8.** Mendefinisikan Variabel

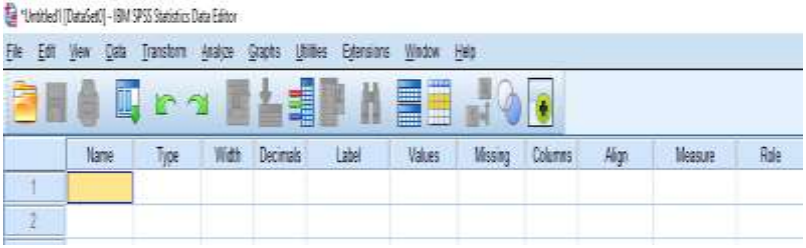
c. Output Viewer

Jendela Output Viewer pada perangkat lunak SPSS menampilkan hasil analisis, deskripsi, dan informasi lainnya secara otomatis. Setiap kali pengguna membuat atau membuka file, Output Viewer akan terbuka untuk data tersebut. Output Viewer terbagi menjadi dua bagian, yaitu bagian kiri dan kanan. Bagian kiri menampilkan garis besar hasil output yang telah dibuat seperti daftar isi, sedangkan bagian kanan menampilkan isi output yang telah dihasilkan. Selain itu, pengguna dapat menyimpan hasil output dalam format file dot spv, serta format lain seperti pdf, doc, dan html melalui opsi Export pada tab File (Advernesia, 2021).



**Gambar 2.9.** Output View

2. Menu SPSS




**Gambar 2.10.** Menu Pada SPSS

Setelah selesai memasukkan data dan menentukan parameter variabel sesuai dengan petunjuk tutorial SPSS yang telah dijelaskan sebelumnya, pengguna dapat menggunakan berbagai menu yang tersedia di SPSS. Menu-menu tersebut memiliki beragam fungsi (Hidayat, 2017), seperti:

a) File; menu ini digunakan untuk melakukan input dan output data. SPSS menyediakan berbagai fitur, seperti membuat dataset baru, membuka dataset yang telah ada, membuka output yang telah dibuat sebelumnya, menjalankan syntax script, mengimpor data dari database atau sumber lain, mengekspor data, menyimpan data, menyimpan data dengan nama yang berbeda, melihat ringkasan file kerja, melihat tampilan printout, membuka file kerja terbaru yang telah dibuat, serta beberapa fitur lainnya.
b) Edit; menu tersebut berfungsi untuk memodifikasi atau memperbarui data, menyisipkan kasus baru, menyisipkan variabel, dan melakukan pencarian pada variabel atau data.
c) View; menu "view" pada SPSS digunakan untuk mengubah tampilan antarmuka, termasuk menampilkan atau menyembunyikan menu bar, toolbar, status bar, dan gridlines.
d) Data; menu ini berfungsi untuk melakukan pemilihan atau penyaringan data, mengurutkan data, mengedit parameter data, merubah struktur data, menyalin dataset dan fungsi lainnya
e) Transform; menu ini berfungsi untuk membuat variabel baru, melakukan operasi matematis, membentuk data time series, serta mempersiapkan data untuk pemodelan data.
f) Analyze; menu ini digunakan untuk melakukan analisis statistik yang beragam, termasuk analisis parametrik dan non-parametrik
g) Graph; digunakan untuk membuat grafik
h) Utilities; menu ini digunakan untuk memberikan catatan pada data, menjalankan skrip macro, menggabungkan model XML, dan sebagainya

i) Add-ons; digunakan untuk menjalankan AddOns atau fitur tambahan dari SPSS
j) Window; menu ini digunakan untuk mengontrol tampilan jendela aplikasi SPSS, seperti mengurangi jendela, membagi layar, dan memilih jendela yang sedang aktif.
k) Help; menu ini digunakan untuk membuka bantuan SPSS, yang dapat berisi referensi tentang cara penggunaan SPSS atau tutorial statistik

# DAFTAR PUSTAKA


Abdillah. 2022. *SPSS*. Available at: https://rumusrumus.com/spss-adalah/.

Advernesia. 2021. *Pengertian SPSS Statistika*. Available at: https://www.advernesia.com/blog/spss/pengertian-spss-statistika/.

Budiyanto. 2018. 'SPSS', *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*. doi: 10.4135/9781506326139.n655.

Hidayat, A. 2017. *Pengantar Tutorial SPSS*, *Statistikian*. Available at: https://www.statistikian.com/2017/05/tutorial-spss-bahasa-indonesia.html.

Priyatno, D. 2009. *Mandiri Belajar SPSS: Untuk Analisis Data dan Uji Statistik*. 3rd edn. Yogyakarta: Mediakom.

Rafiudin, R. and Saepudin, A. 2009. *Praktek Langsung SPSS 17*. Jakarta: Elex Media Komputindo.

Santoso, S. 2016. *Panduan Lengkap SPSS Versi 23*. Jakarta: Elex Media Komputindo.

Sunjoyo, S. *et al.* 2012. *Analisis SPSS Untuk Smart Riset*. Bandung: Alfabeta.

# BAB 3
# STRUKTUR DATA PENELITIAN

*Oleh Maitri Anindita*

## 3.1 Variabel Penelitian dan Pengukurannya

Suatu penelitian pada umumnya memiliki pertanyaan penelitian. Pertanyaan penelitian adalah tujuan dari sebuah penelitian terkait kepentinga dan variabel yang termuat di dalam penelitian. Beberapa contoh pertanyaan penelitian sederhana adalah:

- Berapa distribusi usia mahasiswa di kelas Biostatistik ?
- Apakah ada hubungan antara status kesehatan mahasiswa kelas Biostatistik dengan jenis kelamin mereka ?
- Apakah ada hubungan antara status kesehatan mahasiswa kelas Biostatistik dengan usia mereka ?

Pertanyaan penelitian yang sangat jelas dan tepat akan memandu para peneliti melakukan penelitian dan memastikan bahwa peneliti tidak berakhir dengan tumpukan informasi yang tidak menghasilkan kontribusi pengetahuan. Peneliti membutuhkan pertanyaan penelitian yang jelas sebelum melakukan analisis Biostatistik untuk menghindari situasi dimana sejumlah besar data tidak perlu dikumpulkan dan tidak mengarah pada hasil yang mempunyai makna.

Peneliti menghadapi kebingungan terkait dengan analisis Biostatistik sebenarnya muncul dari ketidaktepatan dalam pertanyaan penelitian yang digunakan untuk memandunya. Sangat sulit untuk memilih jenis analisis yang relevan untuk melakukan pengolahan, mengingat banyak analisis yang dapat diterapkan pada sekumpulan data tertentu. Apabila peneliti tidak dengan jelas

mendefnisikan pertanyaan penelitian maka peneliti tidak akan tahu apa yang harus dilakukan dengan data penelitian setelah dikumpulkannya. Sebaliknya, jika peneliti jelas mendefinisikan tentang pertanyaan penelitian, peneliti akan mudah dalam menangani teknik Biostatistik yang diterapkan. Setiap pertanyaan penelitian di atas mengidentifikasi entitas yang akan di teliti atau sering kita sebut sebagai unit analisis. Sebuah Kasus penelitian adalah entitas yang menampilkan atau memiliki ciri-ciri dari suatu variabel.

Beberapa kasus dapat bersifat individu seperti pada contoh pertanyaan penelitian diatas, tetapi selalu demikian. Misalnya, jika peneliti tertarik pada tingkat retensi sekolah menengah umum di area tertentu, kasusnya adalah sekolah menengah umum. Contoh tersebut adalah sekolah menengah menengah individu yang beri label untuk menunjukkan tingkat retensi masing-masing. Dalam pertanyaan penelitian yang tercantum di atas, semua mahasiswa di kelas Biostatistik merupakan target populasi peneliti (kadang disebut semesta).

Populasi adalah himpunan semua kemungkinan kasus yang menarik. Dalam menentukan populasi yang diinginkan, peneliti biasanya menentukan titik waktu untuk menentukan populasi. Sebagai contoh apakah peneliti tertarik dengan mahasiswa Biostatistik yang terdaftar saat ini, atau mereka yang telah menyelesaikan mata kuliah Biostatistik tahun lalu juga. Peneliti juga menentukan wilayah geografis dimana populasi tersebar. Peneliti memiliki beberapa alasan tertentu untuk tidak mengambil seluruh populasi yang diminati. Peneliti memilih hanya sebagian dari populasi, dan himpunan. Bagian ini serting disebut sebagai sampel.

Sampel adalah sekumpulan kasus yang tidak mencakup setiap anggota populasi. Alasan penggunaan sampel dalam penelitian misalnya adalah mungkin terlalu mahal atau memakan waktu untuk menyertakan setiap mahasiswa dalam suatu

penelitian. Sebagai gantinya, peneliti hanya dapat memilih mahasiswa di kelas Biostatistik yang nama belakangnya dimulai dengan "S", dan hal ini berarti peneliti bekerja dengan menggunakan sampel. Contoh lain misalnya peneliti mengambil sampel mahasiswa dari kelas Biostatistik. Peneliti akan mengamati bahwa para mahasiswa akan berbeda satu sama lain dalam banyak hal. Perbedaan itu mungkin dalam hal jenis kelamin, tinggi badan, usia, agama, sikap terhadap mata kuliah Biostatistik dan lain lain. Topik penelitian mungkin berbeda satu sama lain, dan masing-masing kemungkinan ekspresi perbedaan ini adalah variabel.

`Variabel adalah kondisi atau kualitas yang dapat berbeda dari satu kasus ke kasus lainnya. Lawan kata dari variabel adalah konstanta. Konstanta adalah sebuah kondisi atau kualitas yang tidak bervariasi atau tidak mengalami perubahan antara berbagai kasus. Sebagai contoh Jumlah gram dalam 1 kilogram adalah konstan. Setiap 1 kilogram akan selalu 1000 gram. Sebagian besar penelitian ditujukan untuk memahami variabel tentang apakah dan mengapa suatu variabel mengambil sifat tertentu untuk beberapa kasus dan sifat yang berbeda untuk kasus lain.

## 3.2 Konseptualisasi dan Operasionalisasi Variabel

Suatu penelitian biasanya mempelajari variabel tertentu dan bukan variabel yang lain. Pilihan variabel untuk diselidiki dipengaruhi oleh sejumlah faktor, hal tersebut meliputi .

1. Kerangka teoritis. Teori adalah cara untuk menafsirkan dan memahami kondisi dunia. Peneliti boleh jadi menerima begitu saja bahwa suatu variabel layak untuk diteliti, pada kenyataannya seringkali merupakan proses seleksi yang sangat berat yang mengarahkan perhatian peneliti pada suatu variabel. Peneliti boleh jadi menganggap variabel tertentu sebagai sesuatu yang penting tetapi peneliti lain menganggap variabel tertentu tidak menarik sesuai dengan pandangannya masing-masing. Tanpa teori untuk

mengatur persepsi peneliti tentang dunia, penelitian akan sering menjadi kumpulan pengamatan yang tidak saling terkait dalam cara yang berarti.

2. Agenda penelitian yang telah ditentukan sebelumnya. Terkadang pertanyaan penelitian dan variabel yang akan diteliti tidak ditentukan sendiri oleh peneliti. Misalnya, sebuah lembaga penelitian dapat mengontrak untuk melakukan penelitian yang memiliki kerangka acuan yang ditetapkan sebelumnya oleh badan kontrak. Dalam situasi seperti itu, orang atau kelompok yang benar-benar melakukan penelitian mungkin memiliki sedikit ruang untuk memilih variabel yang akan diselidiki dan bagaimana akan didefinisikan, karena orang atau kelompok tersebut melakukan pekerjaan untuk orang atau organisasi lain.
3. Penelitian yang digerakkan oleh rasa keingintahuan. Kadang-kadang peneliti mungkin tidak memiliki kerangka teori yang jelas untuk beroperasi, atau arahan yang jelas dari orang atau badan lain mengenai konsep kunci yang akan diselidiki. Peneliti ingin menyelidiki variabel murni berdasarkan firasat, perasaan yang dipahami secara luas bahwa sesuatu yang berguna atau penting dapat terungkap jika peneliti mempelajari variabel tertentu. Hal ini dapat menjadi alasan penting untuk melakukan penelitian sebagai keharusan teoretis. Peneliti yang mencoba untuk mengambil bidang penelitian yang sama sekali baru dimana teori yang ada belum banyak mengungkap, firasat sederhana bisa menjadi motivasi yang bermanfaat.

Ketiga faktor diatas ini tidak harus saling terpisah. Jika agenda penelitian ditentukan oleh orang lain, orang tersebut hampir pasti beroperasi dalam kerangka teori tertentu. Apapun motivasinya, penyelidikan pada awalnya akan mengarahkan peneliti ke variabel tertentu untuk diselidiki. Pada tahap awal

peneliti diberikan definisi konseptual variabel. Definisi konseptual variabel menggunakan istilah literal untuk menentukan kualitas variabel.

Definisi konseptual sangat mirip dengan definisi yang ada di kamus. Definisi tersebut memberikan definisi yang berfungsi dari variabel sehingga peneliti memiliki pengertian umum tentang apa artinya variabel tersebut. Sebagai contoh, peneliti mungkin mendefinisikan kesehatan secara konseptual sebagai keadaan kesejahteraan individu. Definisi ini adalah definisi yang masih abstrak dan belum mengarah ke operasional pengamatan data. Sebagai contoh jika lembaga penelitian menginstruksikan para peneliti mengukur keadaan kesejahteraan individu, sekelompok peneliti akan mengalami kebingungan untuk mengamatinya dan memiliki pendapat yang berbeda-beda terkait bagaimana mengamati keadaan kesehjateraan individu.

Definisi konseptual dari suatu variabel hanyalah sebuah permulaan. Peneliti juga membutuhkan seperangkat aturan dan prosedur operasi yang akan memungkinkan peneliti untuk benar-benar mengamati variabel untuk masing-masing kasus, apa yang akan peneliti cari untuk mengidentifikasi status kesehatan individu dan bagaimana para peneliti mencatat kondisi kesejahteraan individu yang bervariasi dari satu orang ke orang lainnya.

Definisi operasional suatu variabel menentukan prosedur dan kriteria untuk mengambil pengukuran variabel itu untuk kasus individu. Sebagai contoh, pernyataan seperti status kesehatan mahasiswa yang diukur dengan seberapa jauh mahasiswa dapat berjalan tanpa bantuan dalam 5 menit memberikan satu definisi operasional dari status kesehatan. Dengan definisi ini peneliti dapat mulai mengukur status kesehatan mahasiswa di kelas Biostatistik dengan mengamati jarak yang ditempuh dalam batas waktu yang ditentukan.

Penentuan definisi operasional untuk variabel adalah yang sumber ketidaksepakatan dalam penelitian. Variabel apa pun

biasanya dapat dioperasionalkan dalam berbagai cara, dan tidak ada satu pun dari definisi operasional yang sempurna. Misalnya, mengoperasionalkan status kesehatan dengan mengamati kemampuan mahasiswa siswa untuk berjalan menghilangkan persepsi subjektif individu tentang seberapa sehat perasaan mereka.

Definisi operasional perlu dibuat agar memenuhi kriteria tertentu. Pepenuhan kriteria ini dalam literatur teknis ini dikenal sebagai masalah validitas konstru. Idealnya, peneliti mencari operasionalisasi yang akan bervariasi ketika variabel yang mendasarinya bervariasi. Sebagai contoh, Termometer air raksa adalah instrumen yang baik untuk mengukur perubahan suhu harian karena ketika variabel yang mendasarinya yaitu suhu berubah, cara mengukurnya (ketinggian batang air raksa) juga berubah. Jika termometer diisi dengan air dan bukan air raksa, variasi suhu harian tidak akan diimbangi dengan perubahan pembacaan termometer. Pada contoh terkait kondisi kesehjateraan individu, apabila peneliti mengandalkan definisi operasional yang hanya mengukur jarak berjalan kaki yang ditempuh dalam waktu tertentu, peneliti mungkin mencatat dua orang sama-sama sehat, padahal sebenarnya mereka berbeda. Misal dua orang yang masing masing berjalan sejauh 700 meter dalam 5 menit, tetapi salah satu dari orang tersebut tidak dapat membungkuk untuk mengikat tali sepatunya karena sakit punggung. Terlihat jelas adanya variasi antara kedua orang tersebut dalam hal kesehatan mereka (kesejahteraan mereka). Namun variasi ini tidak akan terekam jika kita hanya mengandalkan ukuran kemampuan berjalan.

Sejumlah faktor mempengaruhi sejauh mana kita bisa sampai pada definisi operasional variabel yang memiliki validitas konstruk yang tinggi.

1. Kompleksitas konsep. Beberapa variabel tidak terlalu rumit seperti jenis kelamin seseorang dapat mudah ditentukan oleh atribut fisik yang diterima secara umum. Tetapi

sebagian besar variabel jarang begitu mudah. Status kesehatan, misalnya, memiliki sejumlah ukuran. Pada tingkat yang luas kita dapat membedakan antara kesehatan fisik, mental, dan emosional. Dua orang mungkin sehat secara fisik, tetapi yang satu hancur secara emosional sementara yang lain bahagia dan puas. Jika kita mengoperasionalkan status kesehatan dengan melihat hanya pada dimensi fisik, perbedaan penting dalam variabel ini mungkin tidak diamati. Memang masing-masing dimensi yang luas dari status kesehatan – fisik, mental, dan emosional – adalah variabel konseptual dalam diri mereka sendiri, dan menimbulkan masalah operasionalisasi mereka sendiri. Jika peneliti mengambil kesehatan fisik sebagai fokusnya, maka peneliti perlu memikirkan semua bentuk ekspresi khususnya, seperti kemampuan berjalan, mengangkat beban, persentase lemak tubuh dan lain-lain.

2. Ketersediaan data. Peneliti boleh jadi memiliki operasionalisasi yang tampaknya menangkap dengan sempurna variabel minat yang mendasarinya tetapi para peneliti mungkin tidak diizinkan, karena alasan privasi, untuk meninjau catatan kesehatan dari rumah sakit untuk mengumpulkan informasi. Hal ini membuat operasionalisasi yang kurang sempurna harus dilakukan, hanya karena kita tidak bisa mendapatkan data yang ideal.
3. Biaya dan kesulitan mendapatkan data. Peneliti mungkin dapat meninjau catatan kesehatan tetapi biaya dan usaha untuk melakukannya mungkin besar, baik dari segi waktu maupun uang. Contoh lain, peneliti mungkin merasa bahwa ukuran tertentu dari pencemaran air sangat ideal untuk menilai degradasi sungai, tetapi kebutuhan untuk mempekerjakan seorang ahli dengan peralatan pengukur yang canggih mungkin menghalangi ini sebagai pilihan, dan

sebagai gantinya penilaian subjektif dari kekeruhan air mungkin lebih dipilih sebagai langkah cepat dan mudah.

4. Etika. Suatu penelitian apakah dibenarkan dilakukan dengan melihat rincian catatan kesehatan seseorang hanya untuk memenuhi tujuan penelitiannya itu sendiri. Rumah sakit mungkin mengizinkannya dan mungkin ada banyak waktu dan uang yang tersedia, tetapi apakah hal ini membenarkan melihat dokumen yang tidak dimaksudkan sebagai bagian dari proyek penelitian. Masalah etika dalam penelitian membutuhkan pembahasan yang lebih mendalam.

Faktor-faktor diatas membuat validitas penelitian dari operasionalisasi variabel menghasilkan banyak perdebatan. Faktanya, banyak perdebatan seputar penelitian kuantitatif sebenarnya bukan tentang metode analisis atau hasil penelitian, melainkan apakah variabel telah didefinisikan dan diukur dengan benar sejak awal. Jika kriteria operasional yang digunakan untuk mengukur suatu variabel sensitif terhadap cara variabel tersebut dioperasionalkan maka penelitian tersebut akan menghasilkan hasil yang menyesatkan.

## 3.3 Skala pengukuran

Tujuan menurunkan definisi operasional dari definisi konseptual adalah untuk memungkinkan peneliti melakukan pengukuran variabel. Pengukuran (atau pengamatan) adalah proses menentukan dan merekam sifat-sifat yang mungkin dari suatu variabel yang ditunjukkan atau dimiliki oleh kasus penelitian. Variabel jenis kelamin menghasilkan dua sifat yang mungkin yaitu laki-laki dan perempuan. Pengukuran variabel tersebut melibatkan penentuan yang mana dari dua kategori ini yang termasuk dalam kategori seseorang.

Skala pengukuran menentukan rentang yang dapat diberikan pada kasus selama proses pengukuran. Ketika peneliti membangun skala pengukuran, peneliti perlu mengikuti dua aturan tertentu. Aturan tersebut meliputi :

1. Skala pengukuran harus menangkap variasi yang cukup untuk memungkinkan peneliti menjawab pertanyaan penelitian. Peneliti dapat menggunakan jumlah tahun penuh yang telah berlalu sejak lahir sebagai definisi operasional untuk mengukur usia pekerja di suatu perusahaan. Hal ini menghasilkan skala dengan seluruh tahun sebagai poin, dan akan menghasilkan berbagai skor untuk usia karena pekerja lahir pada tahun yang berbeda-beda. Pengukuran usia pelajar sekolah dasar dengan skala ini tidak akan memadai karena sebagian besar atau semua siswa akan lahir pada tahun yang sama. Menggunakan skala pengukuran untuk usia yang hanya mencatat seluruh tahun tidak akan mendapatkan variasi yang cukup untuk membantu peneliti mencapai tujuannya karena setiap siswa di kelas akan terlihat seumuran. Peneliti dapat mempertimbangkan sebagai gantinya dengan menggunakan jumlah keseluruhan bulan berlalu sejak lahir sebagai skala pengukuran. Usia dalam beberapa bulan akan menangkap variasi di antara siswa pada kelas utama usia itu selama bertahun-tahun. Contoh usia untuk sekelompok siswa ini menyoroti masalah intrinsik ketika peneliti mencoba membuat skala untuk mengukur variasi yang cukup untuk suatu variabel kontinyu yang tidak muncul ketika peneliti mencoba mengukur variabel diskrit.
   Variabel diskrit memiliki jumlah nilai yang terbatas. variabel kontinyu dapat bervariasi dalam kuantitas dengan derajat yang sangat kecil. Misalnya, jenis kelamin siswa adalah variabel diskrit dengan hanya dua kemungkinan kategori (laki-laki atau perempuan). Variabel diskrit

seringkali memiliki satuan pengukuran yang tidak dapat dibagi lagi, seperti jumlah anak per rumah tangga. Contoh variabel diskrit lainnya adalah jumlah kecelakaan industri pada tahun tertentu. Usia mahasiswa, di sisi lain, adalah variabel kontinu. Usia dapat dibayangkan berubah dalam bertahap cara dari orang ke orang atau untuk orang yang sama dari waktu ke waktu. Karena itu, variabel kontinu diukur dengan satuan yang dapat dibagi lagi secara tak terhingga. Usia, misalnya, tidak memiliki satuan dasar untuk mengukurnya. Peneliti mungkin mulai dengan mengukur usia dalam tahun. Tapi satu tahun bisa dibagi menjadi bulan, dan bulan menjadi minggu, minggu menjadi hari, dan seterusnya. Satu-satunya batasan adalah seberapa tepat yang diinginkan oleh peneliti. Tahun menangkap lebih sedikit variasi daripada bulan, dan bulan kurang dari minggu. Secara teoritis, dengan variabel kontinu peneliti dapat berpindah secara bertahap dan lancar dari satu nilai variabel ke nilai berikutnya tanpa harus melompat. Meskipun demikian pada prakteknya peneliti harus selalu membulatkan pengukuran dan memperlakukan variabel kontinu seolah olah variabel itu diskrit, dan ini menyebabkan skala pengukuran melompat dari satu titik ke titik berikutnya. Skala tersebut adalah dengan diskrit, meskipun variabel yang mendasarinya kontinu.

Penggunaan skala pengukuran diskrit untuk mengukur usia, apakah peneliti melakukannya dalam tahun atau bulan, menyebabkan peneliti mengelompokkan kasus menjadi beberapa kelompok. Titik-titik pada skala bertindak seperti pusat gravitasi yang menarik semua variasi kecil di sekitarnya. Peneliti mungkin dapat mengatakan bahwa dua orang masing-masing berusia 20 tahun, tetapi sebenarnya mereka akan berbeda dalam hal

usia, kecuali jika mereka dilahirkan tepat pada saat yang sama. Tapi sedikit perbedaan antara seseorang yang berumur 20 tahun, 3 bulan, 4 hari, 5 jam, 27 detik... dan seseorang yang berumur 20 tahun, 4 bulan, 12 hari, 3 jam, 15 detik menjadi tidak relevan untuk masalah penelitian yang sedang peneliti selidiki dan peneliti memperlakukan mereka sama dalam hal variabel, meskipun mereka benar benar berbeda. Tidak ada skala pengukuran yang bisa berharap untuk menangkap variasi penuh yang diungkapkan oleh variabel kontinu. Masalah praktis yang peneliti hadapi adalah apakah skala menangkap cukup variasi untuk membantu peneliti menjawab pertanyaan penelitian.

2. Skala pengukuran harus memungkinkan peneliti dapat menetapkan setiap kasus menjadi satu dan hanya satu titik pada skala. Pernyataan ini sebenarnya mewujudkan dua prinsip pengukuran yang terpisah. Pertama adalah prinsip eksklusivitas yang menyatakan bahwa tidak boleh ada kasus yang memiliki lebih dari satu nilai untuk variabel yang sama. Misalnya, seseorang tidak boleh berusia 19 tahun dan 22 tahun. Kedua, pengukuran juga harus mengikuti prinsip kelengkapan yang menyatakan bahwa setiap kasus dapat diklasifikasikan ke dalam suatu kategori. Skala status kesehatan yang hanya memiliki poin sehat dan sangat sehat tidak mencukupi untuk siapa pun yang kurang sehat karena tidak dapat diukur pada skala ini.

## 3.4 Tingkat pengukuran

Skala pengukuran memungkinkan peneliti untuk mengumpulkan data yang memberikan informasi tentang variabel yang ingin diukur. Data adalah hasil pengukuran yang diambil untuk variabel tertentu untuk setiap kasus dalam sebuah

penelitian. Skala pengukuran tidak memberikan jumlah informasi yang sama tentang variabel yang coba diukur. Faktanya, peneliti umumnya berbicara tentang skala pengukuran yang memiliki satu dari empat skala berbeda pada tingkat pengukurannya: nominal, ordinal, interval, dan rasio.

Peneliti berbicara tentang tingkat pengukuran karena semakin tinggi tingkat pengukuran, semakin banyak informasi yang peneliti miliki tentang suatu variabel. Tingkat pengukuran ini adalah perbedaan mendasar dalam Biostatistik, karena menentukan banyak hal yang dapat peneliti lakukan dengan data yang dikumpulkannya. Lebih lanjut ketika peneliti mempertimbangkan teknik Biostatistik mana yang dapat digunakan untuk menganalisis data, biasanya pertanyaan pertama yang diajukan adalah tingkat pengukuran suatu variabel. Terdapat hal-hal yang dapat peneliti lakukan dengan data yang dikumpulkan pada tingkat pengukuran interval yang tidak dapat peneliti lakukan dengan data yang dikumpulkan pada tingkat nominal.

## 3.5 Skala nominal

Tingkat pengukuran terendah adalah skala nominal. Skala nominal pengukuran mengklasifikasikan kasus ke dalam kategori yang tidak memiliki urutan kuantitatif. Sebagai contoh ketika peneliti tertarik dengan agama seseorang. Secara operasional peneliti mendefinisikan agama seseorang memberikan rentang kategori berikut: Islam, Kristen, Katolik, Hindu, Budha dan lainnya. Perhatikan bahwa untuk memastikan skalanya lengkap, seperti sebagian besar skala nominal, memiliki kategori umum 'Lainnya'. Kategori seperti itu, terkadang diberi label bermacam-macam di akhir skala dan hal ini memberikan cara cepat untuk mengidentifikasi skala pengukuran nominal. Cara lain yang mudah untuk mendeteksi skala nominal adalah mengatur ulang urutan kategori yang terdaftar dan melihat apakah skala tersebut masih

'masuk akal'. Sebagai contoh terdapat urutan agama pada ilustrasi dibawah.

1. Kristen Islam Katolik Hindu Budha Lainnya
2. Islam Katolik Hindu Budha Kristen Lainnya

Terlihat jelas bahwa urutan kemunculan kategori tidak masalah, asalkan aturan eksklusivitas dan kelengkapan diikuti. Hal ini karena tidak ada pengertian peringkat agama atau urutan besarnya agama. Seseorang tidak dapat mengatakan bahwa seseorang dalam kategori 'Kristen' memiliki agama lebih banyak atau lebih sedikit daripada seseorang dalam kategori 'Hindu'. Dengan kata lain,variabel yang diukur pada tingkat nominal bervariasi secara kualitatif tetapi tidak secara kuantitatif. Seseorang dalam kategori Kristen secara kualitatif berbeda dengan seseorang dalam kategori Hindu terkait dengan variabel 'Agama', tetapi mereka tidak memiliki agama yang lebih atau kurang daripada agama yang lainnya.

Peneliti dapat menetapkan nomor untuk setiap kategori untuk alasan kenyamanan dan kemudahan yang akan sangat berguna ketika nanti harus memasukkan data ke dalam aplikasi statistik seperti SPSS. Peneliti dapat menetapkan nomor untuk kategori agama dengan cara berikut:

1. Islam
2. Kristen
3. Katolik
4. Hindu
5. Budha
6. Lainnya

Namun, angka-angka ini hanyalah label kategori yang tidak memiliki arti kuantitatif. Angka-angka tersebut hanya mengidentifikasi kategori yang berbeda, tetapi tidak mengungkapkan hubungan matematis antara kategori tersebut.

Angka (kode) tersebut digunakan untuk kenyamanan untuk memasukkan dan menganalisis data. Peneliti dapat dengan mudah menggunakan skema pengkodean berikut untuk menetapkan nilai numerik ke setiap kategori.

1 Muslim
4 Kristen
5 Katolik
6 Hindu
8 Budha
9 Lainnya

Sebuah urutan skala pengukuran juga mengkategorikan kasus. Jadi skala nominal dan ordinal terkadang disebut secara kolektif skala kategoris. Namun, skala ordinal memberikan informasi tambahan. Sebuah urut skala pengukuran, selain fungsi klasifikasi, memungkinkan kasus diurutkan berdasarkan derajat menurut pengukuran variabel. Timbangan ordinal, yaitu, memungkinkan kita untuk pangkat kasus. Pemeringkatan melibatkan pengurutan kasus dalam arti kuantitatif, seperti dari 'terendah' ke 'tertinggi', dari 'kurang' ke 'lebih', atau dari 'terlemah' ke 'terkuat', dan sangat umum digunakan saat mengukur sikap atau kepuasan dalam suatu survei opini.

Misalnya, sebuah asumsi bahwa peneliti menetapkan dalam mencoba mengukur usia dengan skala sebagai berikut:

- Kurang dari 20 tahun
- 20 sampai 60 tahun
- Lebih dari 60 tahun.

Skala diatas adalah skala nominal, yaitu menetapkan kasus ke dalam kategori. Selain itu juga memungkinkan peneliti untuk menggolongkan bahwa seseorang yang termasuk dalam kategori '20 hingga 60 tahun' lebih tua daripada seseorang dalam kategori 'kurang dari 20 tahun'. Tua adalah peringkat di atas seseorang yang

'kurang dari 20 tahun'. Tidak seperti data nominal, kasus dalam satu kategori tidak hanya berbeda dengan kasus di kategori lain, tetapi juga 'lebih tinggi', atau 'lebih kuat', atau 'lebih besar', atau lebih 'intens'. Peneliti tidak dapat mengatur ulang kategori skala ordinal tidak seperti pada skala nominal. Jika peneliti membuat skala dengan cara berikut, pertambahan usia bergerak melintasi halaman akan hilang keteraturannya :

- 20 sampai 60 tahun
- Lebih dari 60 tahun.
- Kurang dari 20 tahun

Seperti halnya data nominal, nilai numerik dapat diberikan ke titik-titik pada skala sebagai bentuk singkatan, tetapi dengan skala ordinal angka-angka ini juga perlu menjaga arti peringkat. Jadi salah satu dari rangkaian angka berikut dapat digunakan:

1 = kurang dari 20 tahun
2 = 20 sampai 60 tahun
3 = Lebih dari 60 tahun

Atau :

20 = kurang dari 20 tahun
30 = 20 sampai 60 tahun
60 = Lebih dari 60 tahun

Salah satu sistem pengkodean memungkinkan kategori untuk diidentifikasi dan diurutkan satu sama lain, tetapi angka itu sendiri tidak memiliki signifikansi kuantitatif di luar fungsi peringkat ini. Salah satu kesalahan umum dalam analisis Biostatistik adalah memperlakukan skala yang memungkinkan jawaban 'Tidak' atau 'Ya' hanya sebagai nominal, padahal hampir selalu ordinal. Pertimbangkan pertanyaan yang meminta peserta dalam penelitian 'Apakah Anda merasa sehat ? '. Peneliti dapat mengatakan bahwa seseorang yang menjawab 'Ya' tidak hanya berbeda dalam tingkat (persepsi) kesehatannya, tetapi juga

memiliki tingkat kesehatan yang lebih tinggi daripada seseorang yang menjawab 'Tidak'. Hampir semua pertanyaan yang menawarkan opsi jawaban Ya/Tidak dapat ditafsirkan dengan cara ini sebagai skala ordinal.

## 3.6 Skala interval/rasio

Skala ordinal memungkinkan peneliti untuk mengurutkan kasus dalam kaitannya dengan variabel. Sebagai contoh peneliti dapat mengatakan bahwa satu kasus 'lebih baik' atau 'lebih kuat' dari yang lain. Tetapi skala ordinal tidak memungkinkan peneliti untuk mengatakannya seberapa banyak sebuah kasus lebih baik atau lebih kuat jika dibandingkan dengan yang lain. Jika peneliti menggunakan skala usia di atas maka peneliti tidak bisa mengatakan seberapa jauh lebih tua atau lebih muda seseorang dalam satu kategori dibandingkan dengan seseorang dalam kategori lain. Hal ini akan menyesatkan bagi peneliti untuk menggunakan skema pengkodean kedua di atas dan mengatakan bahwa seseorang di kelompok tertua memiliki 40 unit usia lebih banyak daripada seseorang di kelompok termuda (yaitu 60 – 20 = 40). Jarak interval antara kategori tidak diketahui.

Sebagai pertimbangan jika peneliti mengukur usia dengan cara alternatif, dengan menanyakan setiap orang berapa tahun telah berlalu antara kelahiran dan ulang tahun terakhir mereka, peneliti dapat menugaskan orang ke dalam kelompok berdasarkan jumlah tahun. Peneliti juga dapat melakukan tugas mengurutkan kasus berdasarkan pengukuran ini dengan menunjukkan siapa yang lebih tua atau lebih muda. Namun, tidak seperti pada skala nominal dan ordinal, peneliti dapat juga mengukur perbedaan jumlah usia antara setiap kasus. Dalam skala pengukuran ini, angka-angka jumlah tahun yang peneliti dapatkan benar-benar menandakan nilai kuantitatif. Kemampuan untuk mengukur jarak antar titik pada skala inilah yang menjadikan metode pengamatan usia ini sebagai skala interval/rasio.

Sebuah skala interval memiliki unit pengukuran interval jarak yang sama antara nilai-nilai pada skala. skala rasio memiliki nilai nol yang menunjukkan kasus di mana tidak ada kuantitas variabel adalah ada. Dengan kata lain, kita tidak hanya dapat mengatakan bahwa satu kasus memiliki lebih banyak (atau lebih sedikit) variabel yang dipertanyakan daripada yang lain, tetapi kita juga dapat mengatakan berapa lebih banyak (atau lebih sedikit). Dengan demikian seseorang yang berumur 30 tahun memiliki umur 8 tahun lebih dari seseorang yang berumur 22 tahun. Peneliti dapat mengukur interval diantara nilai umur tersebut. Lebih-lebih lagi,interval antara titik-titik pada skala memiliki nilai yang sama di seluruh rentangnya, sehingga selisih umur antara 22 dan 30 tahun sama dengan selisih umur antara 64 dan 72 tahun. Angka-angka pada skala interval memang jelas memiliki signifikansi kuantitatif. Oleh karena itu angka-angka ini disebut nilai-nilai untuk variabel. Peneliti juga akan merujuk pada angka yang digunakan untuk mewakili kategori data nominal dan ordinal sebagai 'nilai' atau 'skor', sehingga istilah 'nilai', 'skor', dan 'kategori' digunakan secara bergantian. Namun untuk data nominal dan ordinal, nilai semacam itu hanyalah label kategori tanpa signifikansi kuantitatif nyata.

Peneliti perlu memperhatikan bahwa pengamatan 0 tahun merepresentasikan kasus dimana tidak memiliki kuantitas variabel usia. Kondisi yang demikian disebut dengan ‘titik nol sebenarnya’ dan merupakan karakteristik yang menentukan skala rasio berlawanan dengan skala interval. Misalnya, suhu yang diukur dalam derajat Celcius tidak memiliki angka yang benar-benar nol. Terdapat titik nol, tetapi 0°C tidak menunjukkan kasus di mana tidak ada panas terjadi. Dingin tetapi tidak sedingin itu. Sebaliknya, 0°C menunjukkan sesuatu yang lain yaitu titik di mana air membeku. Namun, perbedaan halus antara skala pengukuran interval dan rasio ini tidak penting untuk kasus ini.

Kita umumnya dapat melakukan analisis Biostatistik yang sama pada data yang dikumpulkan pada skala interval yang dapat kita lakukan pada data yang dikumpulkan pada skala rasio, dan dengan demikian kita berbicara tentang satu tingkat pengukuran interval/rasio.

Pentingnya pembedaan antara skala nominal, ordinal, dan interval/rasio adalah banyaknya informasi tentang suatu variabel yang disediakan oleh setiap tingkatan (Tabel 3.1).

**Tabel 3.1.** Skala Pengukuran

| Tingkat pengukuran | Contoh | Prosedur Pengukuran | Operasi yang diijinkan |
|---|---|---|---|
| Nominal (tingkat terendah) | Jenis kelamin<br>Ras<br>Agama<br>Status Pernikahan | Klasifikasi ke dalam kategori | Menghitung jumlah kasus di setiap kategori; membandingkan jumlah kasus di setiap kategori |
| Ordinal | Kelas sosial<br>Skala sikap dan opini | Klasifikasi ditambah peringkat dari setiap kategori yang mempertimbangkan | Semua di atas ditambah penilaian 'lebih besar dari' atau 'kurang dari' |
| Interva/rasio | Umur dalam tahun<br>Detak jantung | Klasifikasi ditambah peringkat ditambah deskripsi jarak antar skor dalam satuan yang sama | Semua di atas ditambah operasi matematika lainnya seperti penjumlahan, pengurangan, perkalian, dan lain-lain. |

Sumber: Healey (1993)

Tabel 3.1 diatas meringkas jumlah informasi yang diberikan oleh setiap tingkat pengukuran dan tugas yang peneliti dapat lakukan dengan data yang dikumpulkan pada setiap tingkat. Data nominal memiliki informasi paling sedikit, data ordinal memberikan lebih banyak informasi karena kita dapat mengurutkan kasus, dan data interval/rasio menangkap informasi paling banyak karena memungkinkan peneliti untuk mengukur perbedaan.

Sebelum menyimpulkan pembahasan tentang tingkat pengukuran ini, terdapat dua hal penting yang perlu menjadi perhatian. Pertama adalah pada setiap variabel yang diberikan dapat diukur pada tingkat yang berbeda-beda, tergantung pada definisi operasionalnya. Sebagai contoh, peneliti dapat mengukur usia dalam satu tahun penuh (interval/rasio), tetapi peneliti juga dapat mengukur usia dalam pengelompokan luas (ordinal). Sebaliknya, skala tertentu dapat memberikan tingkat pengukuran yang berbeda tergantung pada variabel tertentu yang diyakininya sedang diukur sampai taraf tertentu terkait dengan interpretasi. Sebagai contoh peneliti mungkin memiliki skala jenis pekerjaan yang dipecah menjadi administrasi, pengawasan, dan manajemen. Jika peneliti mengartikan skala ini hanya sebagai menandakan pekerjaan yang berbeda, maka hal tesebut mengukur klasifikasi pekerjaan yang merupakan skala nominal. Jika peneliti melihat skala ini sebagai mengukur status pekerjaan maka peneliti dapat mengurutkan secara hierarkis kategori-kategori ini ke dalam skala ordinal.

# DAFTAR PUSTAKA


Argyrous, G. 2011. *Statistics for research: With a guide to SPSS*. Sage Publications..

Black, K. 2019. *Business statistics: for contemporary decision making*. John Wiley & Sons.

Daniel, W. W., & Cross, C. L. 2018. *Biostatistics: a foundation for analysis in the health sciences*. Wiley.

Healey, J. F. 1993. Statistics: A Tool for Social Research; Wadsworth Pub. *Co.: Belmont, CA, USA*.

Rosner, B. 2015. *Fundamentals of biostatistics*. Cengage learning.

# BAB 4
# DISTRIBUSI FREKUENSI

*Oleh Rusydi Fauzan*

## 4.1 Pendahuluan

Distribusi frekuensi adalah metode statistik yang digunakan untuk meringkas dan menginterpretasikan sejumlah besar data. Ini adalah alat yang berguna untuk banyak manfaat, seperti mengidentifikasi rentang nilai, menjelaskan tendensi sentral, menganalisis bentuk data, membandingkan kumpulan data, menyederhanakan data, dan memperkirakan probabilitas. Dengan menggunakan distribusi frekuensi, ahli statistik dan analis data dapat memperoleh wawasan berharga ke dalam data dan membuat keputusan yang tepat berdasarkan hasil.

Distribusi frekuensi adalah alat fundamental dalam statistik yang memiliki banyak aplikasi di berbagai bidang studi. Dalam penelitian ilmiah, distribusi frekuensi dapat digunakan untuk menganalisis dan menginterpretasikan data dari eksperimen, survei, dan sumber lainnya. Berikut adalah beberapa kemungkinan aplikasi distribusi frekuensi di masa depan dalam penelitian ilmiah yaitu penelitian biomedis, ilmu lingkungan, ilmu kemasyarakatan, pembelajaran mesin, dan kontrol kualitas.

Distribusi frekuensi adalah alat serbaguna yang dapat diterapkan di berbagai bidang penelitian ilmiah. Aplikasinya berkisar dari menganalisis biomarker dalam penelitian biomedis hingga memantau atribut produk dalam proses manufaktur (Hogg and Tanis, 2019). Karena pengumpulan dan analisis data terus memainkan peran penting dalam penelitian ilmiah, penggunaan distribusi frekuensi kemungkinan akan tetap menjadi alat penting untuk menganalisis dan menginterpretasikan data.

## 4.2 Definisi Distribusi Frekuensi

Distribusi frekuensi merupakan konsep statistik yang mengacu pada pola frekuensi kemunculan nilai atau kelompok nilai dalam suatu kumpulan data. Ini digunakan untuk mengatur data ke dalam kelas atau interval, lalu menghitung berapa banyak titik data yang masuk ke dalam setiap interval. Distribusi frekuensi dapat disajikan dalam bentuk tabel, grafik, atau histogram.

Distribusi frekuensi menjadi alat penting dalam statistik karena menyediakan cara untuk meringkas dan menginterpretasikan data dalam jumlah besar. Dengan menganalisis distribusi frekuensi, ahli statistik dapat memperoleh wawasan tentang kecenderungan sentral, variabilitas, dan bentuk data.

Distribusi frekuensi merupakan konsep dasar dalam statistika yang berperan penting dalam meringkas dan menginterpretasikan data. Ini digunakan untuk mengatur data ke dalam kategori atau interval yang bermakna, yang memfasilitasi analisis dan interpretasi. Memahami distribusi frekuensi sangat penting bagi siapa saja yang bekerja dengan data dan statistik.

## 4.3 Tujuan Distribusi Frekuensi

Terdapat beberapa tujuan kenapa ahli statistik atau pihak yang berkepentingan menggunaan distribusi frekuensi menurut (Devore, 2015) dan (Field, 2013) yaitu:

1. *Describing the data*
   Distribusi frekuensi digunakan untuk meringkas data dengan menyajikan frekuensi setiap nilai atau rentang nilai. Hal ini dapat membantu dalam memahami distribusi data dan mengidentifikasi pencilan.
2. *Identifying the mode*
   Distribusi frekuensi dapat digunakan untuk mengidentifikasi modus, yaitu nilai yang paling sering muncul dalam data. Hal ini dapat berguna dalam

menentukan nilai tipikal atau nilai yang paling umum dalam kumpulan data.

3. *Calculating percentiles*
   Distribusi frekuensi dapat digunakan untuk menghitung persentil, yang merupakan ukuran posisi relatif dalam kumpulan data. Hal ini dapat membantu dalam membandingkan individu atau kelompok pada variabel tertentu.
4. *Testing for normality*
   Distribusi frekuensi dapat digunakan untuk menguji apakah data mengikuti distribusi normal, yang merupakan asumsi utama untuk banyak analisis statistik. Hal ini dapat membantu menentukan uji statistik mana yang sesuai untuk data tersebut.
5. *Comparing groups*
   Distribusi frekuensi dapat digunakan untuk membandingkan distribusi variabel antara kelompok yang berbeda, seperti laki-laki vs perempuan atau perlakuan vs kontrol. Hal ini dapat membantu mengidentifikasi perbedaan atau kesamaan di antara kelompok-kelompok tersebut.
6. *Identifying trends over time*
   Distribusi frekuensi dapat digunakan untuk menganalisis data dari waktu ke waktu dan mengidentifikasi tren, seperti peningkatan atau penurunan nilai. Hal ini dapat membantu untuk memahami pola dan perubahan dalam data.
7. *Estimating probabilities*
   Distribusi frekuensi dapat digunakan untuk memperkirakan probabilitas terjadinya peristiwa tertentu, berdasarkan distribusi data. Hal ini dapat berguna dalam memprediksi hasil atau membuat keputusan.
8. *Simplifying data*

Distribusi frekuensi dapat digunakan untuk menyederhanakan data dengan mengelompokkan nilai ke dalam kategori atau interval. Hal ini dapat membuat data lebih mudah dipahami dan dianalisis.

9. *Evaluating data quality*
   Distribusi frekuensi dapat digunakan untuk mengevaluasi kualitas data dengan mengidentifikasi nilai yang hilang, pencilan, atau ketidakkonsistenan. Hal ini dapat membantu mengidentifikasi dan memperbaiki kesalahan dalam data.
10. *Visualizing data*
    Distribusi frekuensi dapat digunakan untuk membuat representasi visual dari data, seperti histogram atau poligon frekuensi. Hal ini dapat membantu mengkomunikasikan distribusi data dengan cara yang jelas dan intuitif.

Distribusi frekuensi merupakan alat fundamental dalam statistik dan analisis data dengan berbagai tujuan. Alat ini digunakan untuk meringkas data, mengidentifikasi nilai tipikal, menghitung persentil, menguji normalitas, membandingkan kelompok, mengidentifikasi tren, memperkirakan probabilitas, menyederhanakan data, mengevaluasi kualitas data, dan memvisualisasikan data.

Dengan memahami distribusi frekuensi dan cara menggunakannya, setiap pihak dapat memperoleh wawasan yang berharga tentang data mereka dan membuat keputusan yang tepat berdasarkan analisis. Kesimpulannya, distribusi frekuensi menyediakan metode yang kuat untuk mengatur dan memahami data, yang sangat penting untuk banyak aplikasi di berbagai bidang seperti bisnis, sains, dan ilmu sosial. Distribusi frekuensi juga menjadi alat yang ampuh untuk meringkas, menganalisis, dan menginterpretasikan data dengan cara yang bermakna.

## 4.4 Tipe Distribusi Frekuensi

Distribusi frekuensi memiliki beberapa jenis yang bentuknya saling berbeda satu sama lain menurut (Montgomery, Peck and Vining, 2012) dan (Larson and Farber, 2019), yaitu:

1. *Ordinary frequency distribution*
   Ini adalah jenis distribusi frekuensi yang paling dasar, yang mencantumkan setiap nilai dalam kumpulan data dan frekuensi kemunculannya. Jenis distribusi frekuensi ini sering digunakan untuk menggambarkan variabel diskrit, seperti jumlah anak dalam sebuah keluarga atau jumlah hewan peliharaan yang dimiliki oleh individu.
2. *Cumulative frequency distribution*
   Jenis distribusi frekuensi ini menunjukkan jumlah total nilai yang berada di bawah atau sama dengan nilai tertentu. Hal ini dapat membantu dalam memahami distribusi data dan mengidentifikasi persentil. Distribusi frekuensi kumulatif dapat ditampilkan dengan menggunakan tabel frekuensi kumulatif atau poligon frekuensi kumulatif.
3. *Relative frequency distribution*
   Jenis distribusi frekuensi ini menunjukkan proporsi atau persentase nilai yang masuk ke dalam setiap kategori. Hal ini dapat membantu dalam membandingkan kelompok atau mengidentifikasi pola dalam data. Distribusi frekuensi relatif dapat ditampilkan dengan menggunakan tabel frekuensi relatif atau histogram frekuensi relatif.
4. *Grouped frequency distribution*
   Jenis distribusi frekuensi ini mengelompokkan nilai ke dalam interval atau kelas dan menunjukkan frekuensi nilai yang masuk ke dalam setiap interval. Distribusi frekuensi yang dikelompokkan sering digunakan ketika berurusan dengan variabel kontinu, seperti usia atau pendapatan.

Distribusi ini dapat ditampilkan dengan menggunakan tabel frekuensi yang dikelompokkan atau histogram.

5. *Skewed frequency distribution*
   Jenis distribusi frekuensi ini tidak simetris, dengan satu ekor lebih panjang dari yang lain. Distribusi frekuensi miring dapat berupa miring positif, di mana ekor berada di sisi kanan, atau miring negatif, di mana ekor berada di sisi kiri. Distribusi frekuensi yang miring dapat disebabkan oleh pencilan atau distribusi data yang tidak normal.
6. *Bimodal frequency distribution*
   Jenis distribusi frekuensi ini memiliki dua puncak, yang mengindikasikan bahwa ada dua mode atau nilai khas yang berbeda dalam kumpulan data. Distribusi frekuensi bimodal dapat disebabkan oleh campuran dua populasi yang berbeda atau kombinasi dari dua proses yang berbeda.

Secara keseluruhan, jenis distribusi frekuensi yang digunakan tergantung pada sifat data dan pertanyaan penelitian yang diajukan. Dengan memahami berbagai jenis distribusi frekuensi, seseorang dapat memilih metode yang paling tepat untuk menganalisis dan menginterpretasikan data mereka.

## 4.5 Bentuk Distribusi Frekuensi

Berikut beberapa bentuk dari distribusi frekuensi menurut (Montgomery, Peck and Vining, 2012), (McClave, Benson and Sincich, 2019) dan (Larson and Farber, 2019) yaitu:

1. *Normal distribution*
   Merupakan distribusi frekuensi simetris yang mengikuti kurva berbentuk lonceng. Rata-rata, median, dan modus dari distribusi normal semuanya sama, dan deviasi standar menentukan penyebaran data. Distribusi normal biasanya

digunakan untuk memodelkan banyak fenomena dunia nyata, seperti tinggi badan, berat badan, dan nilai ujian.

2. *Skewed distribution*
   Distribusi miring adalah distribusi frekuensi yang tidak simetris. Ada dua jenis distribusi miring: miring positif dan miring negatif. Pada distribusi miring positif, ekornya lebih panjang di sisi kanan, dan rata-rata lebih besar daripada median. Pada distribusi miring negatif, ekornya lebih panjang di sisi kiri, dan rata-rata lebih kecil dari median. Distribusi miring sering kali disebabkan oleh pencilan atau distribusi data yang tidak normal.
3. *Bimodal/multimodal distribution*
   Distribusi bimodal/multimodal adalah distribusi frekuensi dengan dua atau lebih modus atau puncak. Hal ini dapat terjadi ketika ada dua atau lebih kelompok atau populasi yang berbeda dalam kumpulan data. Distribusi bimodal/multimodal biasanya digunakan untuk memodelkan nilai ujian, di mana mungkin ada dua kelompok siswa dengan tingkat kinerja yang berbeda atau dalam distribusi pendapatan, di mana mungkin ada beberapa kelompok pendapatan.
4. *Uniform distribution*
   Distribusi seragam adalah distribusi frekuensi di mana setiap nilai dalam kumpulan data memiliki frekuensi yang sama. Dalam distribusi seragam, tidak ada modus, dan rata-rata serta mediannya sama. Distribusi seragam biasanya digunakan untuk memodelkan lemparan dadu, pelemparan koin, dan peristiwa acak lainnya.
5. *Logarithmic/Pareto distribution*
   Distribusi logaritmik/Pareto adalah distribusi frekuensi di mana frekuensi suatu kejadian berbanding terbalik dengan peringkat atau besarnya. Dalam distribusi logaritmik/Pareto, terdapat banyak kejadian kecil dan

sedikit kejadian besar. Distribusi logaritmik/Pareto biasanya digunakan untuk memodelkan distribusi pendapatan, ukuran kota, dan fenomena lain di mana beberapa kejadian besar mendominasi.

6. *PERT/triangular (triangular) distribution*
   Distribusi PERT/triangular (segitiga) adalah distribusi frekuensi di mana nilai-nilai dalam kumpulan data kemungkinan besar mengelompok di sekitar nilai pusat. Dalam distribusi PERT/triangular (segitiga), terdapat nilai minimum, nilai maksimum, dan modus atau nilai yang paling mungkin. Distribusi PERT/triangular (segitiga) biasanya digunakan dalam manajemen proyek dan analisis risiko untuk memperkirakan kemungkinan berbagai hasil.

Bentuk distribusi frekuensi yang diamati tergantung pada sifat data dan proses yang mendasari data tersebut. Dengan memahami berbagai bentuk distribusi frekuensi, seseorang dapat memilih metode yang tepat untuk menganalisis dan menginterpretasikan data mereka.

## 4.6 Penyajian Distribusi Frekuensi

Berikut beberapa bentuk penyajian dari distribusi frekuensi (Larson and Farber, 2019) dan (McClave, Benson and Sincich, 2019) yaitu:

1. *Histogram*
   Histogram adalah representasi grafis dari distribusi frekuensi. Histogram terdiri dari serangkaian persegi panjang atau batang, di mana area setiap batang mewakili frekuensi nilai atau rentang nilai tertentu. Histogram biasanya digunakan untuk memvisualisasikan data kontinu, seperti nilai tes, tinggi badan, dan berat badan.
2. *Polygon*

Poligon mirip dengan histogram, tetapi alih-alih batang, poligon menggunakan serangkaian segmen garis yang terhubung untuk merepresentasikan distribusi frekuensi. Poligon berguna untuk membandingkan beberapa distribusi frekuensi pada grafik yang sama.

3. *Pie chart*
   Diagram lingkaran adalah grafik melingkar yang merepresentasikan distribusi frekuensi sebagai serangkaian irisan, di mana area setiap irisan merepresentasikan frekuensi relatif dari nilai atau kategori tertentu. Diagram lingkaran biasanya digunakan untuk memvisualisasikan data kategorikal, seperti proporsi siswa di kelas yang lebih menyukai jenis musik yang berbeda.
4. *Ogive*
   Ogive adalah grafik yang merepresentasikan distribusi frekuensi kumulatif, yang menunjukkan jumlah total pengamatan yang kurang dari atau sama dengan nilai tertentu. Ogive berguna untuk menganalisis tren dalam data dan mengidentifikasi pencilan.
5. *Stem-and-leaf plot*
   Plot batang-dan-daun adalah representasi grafis dari distribusi frekuensi yang menampilkan setiap pengamatan sebagai batang (angka atau angka pertama) dan daun (angka terakhir). Plot batang-dan-daun berguna untuk memvisualisasikan kumpulan data berukuran kecil hingga sedang.
6. *Dot plot*
   Dot plot adalah representasi grafis dari distribusi frekuensi yang menggunakan titik-titik untuk mewakili setiap pengamatan. Plot titik berguna untuk memvisualisasikan kumpulan data berukuran kecil hingga sedang dan mengidentifikasi pencilan.

Pilihan metode penyajian tergantung pada sifat data dan tujuan analisis. Dengan memahami berbagai metode untuk menyajikan distribusi frekuensi, seseorang dapat memilih metode yang tepat untuk mengkomunikasikan hasil analisisnya kepada orang lain.

## 4.7 Komponen Tabel Distribusi Frekuensi

Berikut adalah beberapa komponen dari tabel distirbusi frekuensi menurut (Larson and Farber, 2019) dan (McClave, Benson and Sincich, 2019) yaitu:

1. *Classes*
   Kelas adalah kategori atau kelompok yang digunakan untuk membagi data. Kategori ini dapat didasarkan pada interval numerik atau kriteria kualitatif, seperti rentang usia atau jenis produk. Jumlah kelas yang digunakan dapat bervariasi, tergantung pada kumpulan data dan tujuan analisis.
2. *Class boundaries*
   Batas kelas adalah titik-titik yang menentukan batas-batas kelas. Batas-batas ini biasanya ditempatkan di tengah-tengah antara batas atas satu kelas dan batas bawah kelas berikutnya. Batas kelas berguna untuk memastikan bahwa setiap titik data hanya ditugaskan ke satu kelas.
3. *Class edges or class real boundaries*
   Tepi kelas, juga dikenal sebagai batas nyata kelas, adalah batas sebenarnya dari kelas. Batas-batas ini biasanya dinyatakan sebagai nilai minimum dan maksimum dari data di setiap kelas.
4. *Class midpoints or class marks*
   Titik tengah kelas, juga dikenal sebagai tanda kelas, adalah titik-titik di tengah setiap kelas. Titik ini dihitung dengan mengambil rata-rata dari batas kelas atas dan bawah.
5. *Class intervals*

Interval kelas adalah rentang nilai yang menentukan setiap kelas. Interval kelas biasanya dipilih dengan ukuran yang sama, meskipun tidak selalu demikian.

6. *Class interval length or class area*
   Panjang interval kelas, juga dikenal sebagai luas kelas, adalah perbedaan antara batas atas dan bawah setiap kelas. Nilai ini penting untuk menghitung area batang histogram atau segmen garis poligon frekuensi.
7. *Class frequency*
   Frekuensi kelas adalah jumlah pengamatan yang masuk ke dalam setiap kelas. Nilai ini dapat dinyatakan sebagai frekuensi absolut, yang merupakan jumlah observasi aktual di setiap kelas, atau sebagai frekuensi relatif, yang merupakan proporsi observasi di setiap kelas relatif terhadap jumlah total observasi.

Distribusi frekuensi dapat dipecah menjadi beberapa komponen utama, termasuk kelas, batas kelas, tepi kelas atau batas nyata kelas, titik tengah kelas atau tanda kelas, interval kelas, panjang interval kelas atau luas kelas, dan frekuensi kelas. Masing-masing komponen ini memainkan peran penting dalam mengatur, meringkas, dan menyajikan data dengan cara yang bermakna.
Memahami komponen-komponen ini sangat penting untuk membuat representasi grafis yang akurat dan informatif dari data, seperti histogram, poligon, diagram lingkaran, dan ogive.

Dengan menganalisis dan menginterpretasikan representasi grafis ini, para peneliti dan analis dapat memperoleh wawasan tentang berbagai fenomena dan membuat keputusan yang tepat berdasarkan temuan mereka.

## 4.8 Cara Menyusun Tabel Distribusi Frekuensi

Berikut adalah beberapa langkah dalam menyusun tabel distribusi frekuensi :

1. *Determine the range of the data*
   Langkah pertama adalah menentukan kisaran data dengan mengurangi nilai terkecil dari nilai terbesar. Hal ini akan membantu dalam menentukan ukuran interval yang akan digunakan.
2. *Decide on the number of intervals*
   Langkah selanjutnya adalah menentukan jumlah interval yang akan digunakan dalam tabel distribusi frekuensi. Hal ini dapat dilakukan dengan menggunakan berbagai metode seperti aturan Sturges, aturan akar kuadrat, atau aturan Freedman-Diaconis. Jumlah interval harus dipilih sedemikian rupa sehingga setiap interval mengandung jumlah observasi yang cukup.
3. *Determine the interval size*
   Setelah jumlah interval ditentukan, ukuran interval dapat dihitung dengan membagi rentang dengan jumlah interval. Ukuran interval harus dibulatkan ke jumlah tempat desimal yang sesuai.
4. *Determine the interval limits*
   Batas interval adalah batas atas dan bawah dari setiap interval. Ini dapat dihitung dengan menambahkan ukuran interval ke batas bawah interval sebelumnya untuk mendapatkan batas atas interval saat ini. Batas bawah dari interval pertama biasanya dipilih sebagai nilai terkecil dalam kumpulan data.
5. *Tally the data*
   Langkah selanjutnya adalah menghitung data untuk setiap interval. Hal ini melibatkan penghitungan jumlah pengamatan yang termasuk dalam setiap interval dan mencatatnya sebagai penghitungan.

6. *Calculate the frequency*
   Setelah data dihitung, frekuensi untuk setiap interval dapat dihitung dengan menjumlahkan hasil penghitungan. Frekuensi adalah jumlah pengamatan dalam setiap interval.
7. *Calculate the cumulative frequency*
   Frekuensi kumulatif adalah jumlah frekuensi untuk saat ini dan semua interval sebelumnya. Hal ini berguna untuk menganalisis distribusi data di seluruh rentang.
8. *Calculate the relative frequency*
   Frekuensi relatif adalah proporsi pengamatan dalam setiap interval relatif terhadap jumlah total pengamatan. Hal ini dapat dihitung dengan membagi frekuensi setiap interval dengan jumlah total pengamatan dan mengalikannya dengan 100%.
9. *Create the frequency distribution table*
   Langkah terakhir adalah membuat tabel distribusi frekuensi, yang harus menyertakan kolom-kolom untuk batas interval, penghitungan, frekuensi, frekuensi kumulatif, dan frekuensi relatif.

Perlu dicatat bahwa langkah-langkah spesifik untuk menyusun tabel distribusi frekuensi dapat bervariasi, tergantung pada sifat data dan tujuan penggunaan tabel. Selain itu, ada berbagai perangkat lunak yang tersedia yang dapat mengotomatiskan proses pembuatan distribusi frekuensi.

*Contoh:*

Tabel distribusi frekuensi digunakan dalam berbagai industri untuk meringkas dan menganalisis data. Berikut ini adalah contoh bagaimana tabel distribusi frekuensi dapat digunakan dalam industri ritel:

Misalkan sebuah toko ritel ingin memahami pola pembelian pelanggannya untuk kategori produk tertentu, misalnya pakaian.

Mereka mungkin mengumpulkan data tentang jumlah yang dibelanjakan pelanggan untuk pembelian pakaian pada bulan tertentu. Toko tersebut kemudian dapat menggunakan tabel distribusi frekuensi untuk meringkas data dan mendapatkan wawasan tentang perilaku pelanggan.

Tabel tersebut dapat mencakup kolom untuk interval pembelanjaan (misalnya $0-50, $50-100, $100-150, dll.), jumlah pelanggan yang membelanjakan uangnya dalam setiap interval ("Frekuensi"), persentase total pelanggan yang membelanjakan uangnya dalam setiap interval ("Frekuensi Relatif"), dan persentase kumulatif pelanggan yang membelanjakan uangnya dalam setiap interval ("Frekuensi Relatif Kumulatif").

Dengan menggunakan tabel ini, toko dapat mengidentifikasi kebiasaan belanja pelanggannya. Sebagai contoh, mereka mungkin menemukan bahwa mayoritas pelanggan menghabiskan antara $50-100 untuk pembelian pakaian setiap bulannya, atau sebagian kecil pelanggan adalah pembelanja besar, menghabiskan lebih dari $500 per bulan untuk pembelian pakaian.

Toko dapat menggunakan wawasan ini untuk menginformasikan strategi pemasaran, penetapan harga, dan manajemen inventaris. Sebagai contoh, mereka dapat membuat promosi yang ditargetkan untuk pelanggan dengan pengeluaran tinggi, atau menyesuaikan harga atau inventaris mereka untuk lebih memenuhi kebutuhan pelanggan dengan kisaran pengeluaran $50-100.

Secara keseluruhan, tabel distribusi frekuensi menyediakan alat yang ampuh untuk meringkas dan menganalisis data, dan dapat digunakan di berbagai industri untuk mendapatkan wawasan dan membuat keputusan yang tepat.

Berikut tabel dari contoh diatas:

**Tabel 4.1.** Tabel Distribusi Frekuensi

| *Spending Intervals ($)* | *Frequency* | *Relative Frequency (%)* | *Cumulative Relative Frequency (%)* |
|---|---|---|---|
| 0-50 | 200 | 40 | 40 |
| 50-100 | 300 | 60 | 100 |
| 100-150 | 50 | 10 | 110 |
| 150-200 | 20 | 4 | 114 |
| 200-250 | 10 | 2 | 116 |
| >250 | 20 | 4 | 120 |

Sumber: Diolah penulis (2023)

Dalam contoh ini, data telah dikelompokkan ke dalam enam interval pembelanjaan, dengan frekuensi yang menunjukkan jumlah pelanggan yang melakukan pembelian dalam interval tersebut. Kolom frekuensi relatif menunjukkan proporsi pelanggan yang berbelanja dalam setiap interval sebagai persentase dari jumlah total pelanggan, dan kolom frekuensi relatif kumulatif menunjukkan proporsi kumulatif pelanggan yang berbelanja dalam setiap interval.

Dengan menggunakan tabel ini, toko ritel dapat dengan mudah melihat bahwa mayoritas pelanggan menghabiskan antara $50-100 untuk pembelian pakaian setiap bulannya, dan sebagian besar pelanggan menghabiskan lebih dari $150. Mereka juga dapat melihat bahwa 60% pelanggan berbelanja dalam interval $50-100, dan 40% pelanggan berbelanja dalam interval $0-50.

## 4.9 Tantangan dan Peluang Penggunaan Distribusi Frekuensi di Masa Depan

Terdapat beberapa tantangan dan peluang mengenai penggunaan distribusi frekuensi dalam prakteknya di masa depan yaitu:

1. *Big Data*
   Karena jumlah data yang dikumpulkan dan dianalisis terus bertambah, menganalisis dan memvisualisasikan distribusi frekuensi akan menjadi semakin menantang. Teknik dan alat baru untuk memproses dan menampilkan kumpulan data yang besar akan dibutuhkan. Penggunaan data harus mampu membantu perusahaan dalam mengambil keputusan (Nugraha *et al.*, 2023), (Seto *et al.*, 2023), dan (Murdana *et al.*, 2023).
2. *Interpretation of Complex Data*
   Dengan munculnya sumber data yang kompleks seperti media sosial, distribusi frekuensi dapat menjadi lebih sulit untuk ditafsirkan karena keragaman data. Teknik seperti penggalian data dan pembelajaran mesin akan diperlukan untuk mengekstrak wawasan yang berarti dari jenis data ini. Dengan itu di masa depan semakin dibutuhkan metode yang lebih baik dalam pengolahan data (Fauzan, Hafidah, *et al.*, 2023), (Santoso *et al.*, 2023), dan (Rachmat *et al.*, 2023).
3. *Data Quality*
   Keakuratan dan keandalan data yang digunakan untuk menyusun distribusi frekuensi dapat memiliki dampak yang signifikan terhadap validitas hasil. Perhatian yang cermat harus diberikan pada pembersihan dan verifikasi data untuk memastikan hasil yang akurat dan bermakna. Data yang akurat dan detail akan sangat berpengaruh dalam berbagai kegiatan yang ada di perusahaan terutama untuk pengambilan keputusan (Fauzan, Hafidah, *et al.*,

2023), (Lestari *et al.*, 2022), dan (Fauzan, A'yun, *et al.*, 2023).

4. *Visualization Techniques*
   Pemilihan teknik visualisasi dapat memiliki dampak yang signifikan terhadap interpretasi distribusi frekuensi. Teknik-teknik baru, seperti visualisasi interaktif dan dinamis, akan menjadi lebih penting untuk memungkinkan pengguna mengeksplorasi dan berinteraksi dengan data dengan cara-cara baru. Perusahaan harus terus menemukan berbagai teknik yang membawa banyak manfaat untuk perusahaan (Fauzan, Nurhayati and Novia, 2020), (Fauzan, Hakim, *et al.*, 2023), dan (Wijaya *et al.*, 2023).
5. *Data Privacy and Security*
   Seiring dengan bertambahnya data, kekhawatiran seputar privasi dan keamanan data juga akan meningkat. Langkah-langkah yang tepat harus diambil untuk melindungi data sensitif dan memastikan bahwa data digunakan dengan cara yang etis dan bertanggung jawab. Keamanan data sangat penting bagi keunggulan perusahaan di masa depan (Sarjana *et al.*, 2022), (Widiana *et al.*, 2023), dan (Fauzan, Putri, *et al.*, 2023).
6. *Cross-Disciplinary Applications*
   Penggunaan distribusi frekuensi tidak terbatas pada bidang tertentu, dan ada peningkatan minat untuk menggunakan teknik ini dalam pengaturan interdisipliner. Kolaborasi antara para ahli di berbagai bidang akan diperlukan untuk sepenuhnya menyadari potensi distribusi frekuensi. Dengan melibatkan banyak pihak maka hasil yang akan didapatkan akan menjadi semakin maksimal (Fauzan, Supryanita and Rahmatika, 2021), (Fauzan, Rizki, *et al.*, 2023), dan (Ernayani *et al.*, 2022).

7. *Real-Time Data*
   Penggunaan distribusi frekuensi dalam analisis data waktu nyata akan menjadi semakin penting dalam bidang-bidang seperti keuangan, kesehatan, dan transportasi. Distribusi frekuensi waktu nyata akan memungkinkan para pengambil keputusan untuk merespons dengan cepat terhadap perubahan dalam data. Sehingga mampu mengambil berbagai peluang dan memenangkan persaingan (Fauzan and Jayanti, 2020), (Kunda *et al.*, 2023), dan (Fauzan and Sari, 2016).
8. *Data Integration*
   Integrasi berbagai sumber data dapat menimbulkan tantangan saat menyusun distribusi frekuensi. Teknik untuk mengintegrasikan data dari berbagai sumber akan menjadi semakin penting untuk memungkinkan analisis yang akurat dan komprehensif. Hasil analisa akan sangat berdampak terhadap kinerja dan kemajuan perusahaan di masa depan (Fauzan and Rahmadani, 2018) dan (Purnamasari *et al.*, 2023).
9. *Dynamic Data*
   Penggunaan distribusi frekuensi dalam pengaturan data yang dinamis, seperti yang ditemukan di lingkungan online, menghadirkan tantangan yang unik. Teknik untuk memvisualisasikan dan menganalisis distribusi frekuensi secara real-time akan menjadi semakin penting dalam pengaturan ini. Hal ini akan sangat membantu banyak pihak yang ingin menggunakan berbagai informasi terkait dengan perusahaan (Fauzan and Sari, 2018) dan (Mustanir *et al.*, 2023).
10. *Data Ethics and Bias*
    Konstruksi dan interpretasi distribusi frekuensi dapat dipengaruhi oleh bias data dan masalah etika. Pertimbangan yang cermat terhadap masalah-masalah ini

akan diperlukan untuk memastikan bahwa analisis yang dilakukan adil, akurat, dan bermakna. Data yang diberikan harus benar dan jujur sehingga membantu setiap pemangku kepentingan dalam mengambil keputusannya (Fauzan, 2014) dan (Fauzan, Setiawan, *et al.*, 2023).

## 4.10 Penutup

Distribusi frekuensi adalah alat statistik yang kuat yang memiliki berbagai aplikasi di berbagai bidang. Karena jumlah data yang dikumpulkan terus meningkat, penggunaan distribusi frekuensi akan menjadi semakin penting untuk analisis data dan proses pengambilan keputusan.

Namun, ada juga tantangan yang perlu diatasi, seperti memastikan keakuratan dan integritas data yang dikumpulkan, menangani outlier dan data yang hilang, serta memilih metode presentasi dan analisis yang sesuai.

Untuk memanfaatkan sepenuhnya manfaat distribusi frekuensi, penting bagi para peneliti, analis, dan pengambil keputusan untuk selalu mengikuti perkembangan teknik dan kemajuan terbaru di bidang ini.

Dengan perkembangan teknologi dan alat baru, seperti algoritme pembelajaran mesin dan analitik data besar, potensi aplikasi distribusi frekuensi akan terus berkembang di masa depan.

Oleh karena itu, sangat penting bagi individu dan organisasi untuk berinvestasi dalam pelatihan dan pendidikan untuk memastikan bahwa mereka memiliki keterampilan dan pengetahuan yang diperlukan untuk secara efektif menggunakan distribusi frekuensi untuk membuat keputusan yang tepat.

# DAFTAR PUSTAKA


Devore, J.L. 2015. *Probability and statistics for engineering and the sciences.* Boston, MA: Cengage Learning.

Ernayani, R. *et al.* 2022. 'The Influence of Sales and Operational Costs on Net Income in Cirebon Printing Companies', *Al-Kharaj: Journal of Islamic Economic and Business*, 4(2020), pp. 81–86.

Fauzan, R. 2014. 'Penilaian Kinerja Pada Lembaga Pendidikan Tinggi Dengan Menggunakan Metode Balanced Scorecard. Studi Kasus: Sekolah Tinggi Ilmu Ekonomi Haji Agus Salim Bukittinggi', *jurnal ekonomi*, 16(2), pp. 50–60.

Fauzan, R., Hakim, M.Z., *et al.* 2023. *Auditing*. Global Eksekutif Teknologi.

Fauzan, R., Hafidah, A., *et al.* 2023. *Ekonomi Manajerial*. Global Eksekutif Teknologi.

Fauzan, R., A'yun, K., *et al.* 2023. *Islamic Marketing*. Global Eksekutif Teknologi.

Fauzan, R., Rizki, M., *et al.* 2023. *Koperasi*. Global Eksekutif Teknologi.

Fauzan, R., Setiawan, R., *et al.* 2023. *Manajemen Perubahan*. Global Eksekutif Teknologi.

Fauzan, R., Putri, R.D., *et al.* 2023. *Wawasan Bisnis*. Global Eksekutif Teknologi.

Fauzan, R. and Jayanti, A. 2020. 'Strategi Pemasaran Untuk Meningkatkan Volume Penjualan Dengan Menggunakan Blue Ocean Strategy Model Pada Usaha Sanjai Nitta Bukittinggi', *Jurnal BONANZA: Manajemen dan Bisnis*, 1(1), pp. 1–12.

Fauzan, R., Nurhayati, N. and Novia, I. 2020. 'Pengambilan Keputusan Strategis dalam Penentuan Harga Jual Produk dengan Menggunakan Pendekatan Activity Based Costing. Studi Kasus UMKM Tia Konveksi', *Jurnal PROFITA: Akuntansi dan Bisnis*, 1(1), pp. 35–46.

Fauzan, R. and Rahmadani, S. 2018. 'Strategi Pengembangan Agrowisata dengan Menggunakan Blue Ocean Strategy Model. Studi Kasus Perkebunan Kopi Green Sago Kabupaten 50 Kota', *jurnal ekonomi*, 21(1), pp. 21–33.

Fauzan, R. and Sari, A.M. 2016. 'Analisis Strategi Pemasaran Untuk Meningkatkan Volume Penjualan (Studi Kasus di Cafe Texas Juice Cabang Tangah Jua Kota Bukittinggi)', *jurnal ekonomi*, 20(2), pp. 147–156.

Fauzan, R. and Sari, R.P. 2018. 'Strategi Pengembangan Taman Marga Satwa dengan Menggunakan SWOT dan QSPM Model. Studi Kasus Taman Marga Satwa dan Budaya Kinantan Kota Bukittinggi', *jurnal ekonomi*, 21(2), pp. 120–131.

Fauzan, R., Supryanita, R. and Rahmatika, R. 2021. 'Analisa Strategi Pemasaran untuk Peningkatan Daya Saing pada Bisnis Kafe di Kota Bukittinggi (Studi Kasus Kafe Teras Kota)', *MABIS: Jurnal Manajemen Bisnis Syariah*, 1(1).

Field, A. 2013. *Discovering statistics using IBM SPSS statistics.* 4th edn. Thousand Oaks, CA: Sage Publications.

Hogg, R. V. and Tanis, E.A. 2019. *Probability and statistical inference.* 10th edn. Boston, MA: Pearson.

Kunda, A. *et al.* 2023. *Pengantar Bisnis: Manajemen, Pembiayaan, Pemasaran Dan Operasional*. Global Eksekutif Teknologi.

Larson, R. and Farber, B. 2019. *Elementary statistics: Picturing the world.* 7th edn. Boston, MA: Pearson.

Lestari, S.P. *et al.* 2022. *Manajemen Operasional*. Global Eksekutif Teknologi.

McClave, J.T., Benson, P.G. and Sincich, T. 2019. *Statistics for business and economics.* 13th edn. Boston, MA: Pearson.
Montgomery, D.C., Peck, E.A. and Vining, G.G. (2012) *Introduction to linear regression analysis.* 5th edn. Hoboken, NJ: Wiley.
Murdana, I.M. *et al.* 2023. *Ekonomi Pariwisata*. Global Eksekutif Teknologi.
Mustanir, A. *et al.* 2023. *Pemberdayaan Masyarakat*. Global Eksekutif Teknologi.
Nugraha, D.B. *et al.* 2023. *Sistem Informasi Akuntansi*. Global Eksekutif Teknologi.
Purnamasari, S. *et al.* 2023. *Manajemen Keuangan Islam*. Global Eksekutif Teknologi.
Rachmat, Z. *et al.* 2023. *Manajemen Syariah*. Global Eksekutif Teknologi.
Santoso, L.W. *et al.* 2023. *Manajemen Sains*. Global Eksekutif Teknologi.
Sarjana, S. *et al.* 2022. *Manajemen Pemasaran*. Global Eksekutif Teknologi.
Seto, A.A. *et al.* 2023. *Analisis Laporan Keuangan*. Global Eksekutif Teknologi.
Widiana, I.N.W. *et al.* 2023. *Perilaku Dan Budaya Organisasi*. Global Eksekutif Teknologi.
Wijaya, K. *et al.* 2023. *Manajemen Pemasaran Lanjutan*. Global Eksekutif Teknologi.

# BAB 5
# TRANSFORMASI DATA DAN ANALISIS UNIVARIAT

*Oleh Suwito Pomalingo*

## 5.1 Pendahuluan

Transformasi data dan analisis univariat sangat penting dalam proses analisis data. Kedua proses ini sangat membantu dalam meningkatkan kualitas data dan memberikan informasi tentang karakteristik data secara mendasar. Transformasi data dan analisis univariat mempunyai hubungan yang erat. Transformasi data sering ditetapkan sebagai tahap awal sebelum melakukan analisis univariat, karena terkadang sebuah data perlu dilakukan normalisasi, penskalaan ulang, ataupun perubahan logaritmik, sehingga data tersebut dapat memenuhi asumsi analisis. Olehnya itu diperlukan proses transformasi data untuk memperbaiki kondisi data yang tidak sesuai untuk dianalisa. Jadi tujuan utama dilakukan proses transformasi data adalah untuk mengubah bentuk data ke dalam bentuk lain agar lebih sesuai untuk dianalisis (Witten, Frank and Hall, 2011) atau memenuhi asumsi analisis (Inmon, 2005).

Adapun analisis univariat adalah salah satu jenis analisis dalam pengolahan data statistik yang digunakan dalam memahami distribusi data, membuat keputusan yang informatif dengan bantuan visualisasi data, dan lainnya. Tujuannya tentunya untuk memahami karakteristik dasar dari suatu variable, seperti *central tendency*, *disperse*, bentuk distribusi, dan *outlier*. Analisis univariat ini merupakan tahap awal yang penting dalam eksplorasi data karena dapat memberikan informasi yang berguna tentang kualitas

data, selain itu analisis univariat berguna memastikan bahwa proses analisis data valid dan akurat.

Pada bagian ini kita akan membahas tentang konsep Transformasi Data dan Analisis Univariat, serta akan disajikan juga beberapa contoh penerapannya.

## 5.2 Transformasi Data

### 5.2.1 Pengertian Transformasi Data

Transformasi Data adalah proses merubah atau mengonversi data ke dalam skala baru untuk memenuhi sebaran data menjadi normal (Inmon, 2005). Data yang perlu ditransformasi merupakan data yang tidak memenuhi syarat untuk dilakukan analisis misalnya mengubah data yang memiliki skala interval ke skala ordinal jika syarat uji statistik mensyaratkan data kategorik. Dalam transformasi data dapat melibatkan perubahan struktur, format, atau bahkan jenis data. Pada banyak kasus, data mentah yang diperoleh dari sumber aslinya mungkin tidak sesuai atau tidak memenuhi kebutuhan analisis. Dengan melakukan proses transformasi data, dapat meningkatkan kualitas data dengan menghilangkan data yang tidak relevan, mengatasi *missing value* ataupun jika ada *outliers*, melakukan normalisasi, standarisasi, dan merubah format data menjadi lebih mudah dipahami (Inmon, 2005).

Transformasi data juga berguna untuk mengubah data ke dalam format yang dapat digunakan oleh algoritma analisis tertentu (Ng, 2018). Misalnya, beberapa algoritma *machine learning* mengharuskan data berada pada rentang nilai tertentu atau mengharuskan data kategorik dibuat menjadi variable numerik, maka transformasi data diperlukan untuk memenuhi kebutuhan tersebut.

Berikut beberapa alasan kenapa perlu dilakukan transformasi data:

1. Untuk memastikan data yang digunakan dalam analisis atau pemodelan memiliki kualitas dan relevansi yang

memadai. Seperti membuat data menjadi lebih simetris, lebih linier, dan data memiliki varian yang konsisten.
2. Untuk meningkatkan akurasi dan performa model analisis yang digunakan.
3. Mengurangi biaya dan waktu yang diperlukan dalam pengolahan data karena data yang telah ditransformasi menjadi lebih mudah dipahami dan digunakan.

Namun yang perlu diingat bahwa transformasi data juga memiliki beberapa resiko, seperti kehilangan informasi atau berkurangnya keakuratan data. Oleh karena itu, proses transformasi data harus dilakukan dengan hati-hati dengan memperhitungkan keuntungan dan resiko yang terkait dengan proses tersebut. Selain itu, data hasil transformasi hanya membantu dalam melakukan proses analisis lanjut, sedangkan untuk ditampilkan pada laporan tetap data aslinya.

### 5.2.2 Jenis-jenis Transformasi Data

Ada berbagai jenis transformasi data yang dapat digunakan dalam analisis data, dan setiap jenis transformasi memiliki tujuan yang berbeda. Berikut jenis-jenis transformasi data yang umum digunakan:

**1. Transformasi Logaritmik**

Transformasi logaritmik merupakan jenis transformasi data yang digunakan untuk menormalkan data yang memiliki distribusi yang sangat condong (*skewed*) atau ada ketimpangan (*asymmetry*). Dengan menggunakan fungsi logaritma, data akan diubah atau ditransformasi menjadi data logaritmik.

Penggunaan transformasi logaritmik biasanya pada data yang memiliki rentang nilai yang sangat luas atau data yang memiliki variasi nilai yang sangat besar pada bagian ekstrim distribusinya. Pada beberapa kasus, transformasi

logaritmik ini dapat digunakan untuk mengatasi hubungan yang tidak linier antara variabel dalam analisis statistik. Setelah dilakukan transformasi, data akan memiliki distribusi yang lebih simetris dan memenuhi asumsi-asumsi yang mendasari analisis ragam (Rosadi, 2017).

Transformasi logaritmik ini dapat dilakukan menggunakan logaritma natural atau logaritma basis 10, tergantung pada kebutuhan analisis yang dilakukan. Fungsi logaritma natural menghasilkan nilai yang lebih rendah untuk nilai-nilai besar dan nilai yang lebih tinggi untuk nilai-nilai kecil. Dengan kata lain, ketika data yang sangat condong atau ada ketimpangan ditransformasikan menjadi data logaritmik, nilai-nilai yang sangat besar akan dikecilkan sedangkan nilai-nilai yang sangat kecil akan diperbesar, sehingga memungkinkan kita untuk melihat variasi data yang lebih baik. Contohnya, kita memiliki data pendapatan dengan rentang nilai dari Rp 10.000 hingga Rp 1.000.000. Data ini memiliki distribusi yang sangat condong dengan sebagian besar data berada pada rentang nilai yang rendah. Dengan transformasi logaritmik, data pendapatan akan diubah menjadi data logaritmik dengan menggunakan fungsi logaritma natural (*ln*). Dalam hal ini, data pendapatan yang asli akan diubah menjadi data logaritmik dengan rumus $ln(x)$, dimana $x$ adalah nilai pendapatan asli. Setelah transformasi logaritmik, rentang nilai data akan lebih terdistribusi secara merata, sehingga memungkinkan kita untuk melihat variasi data yang lebih baik.

Contoh penggunaan logaritma natural misalnya kita memiliki data dengan distribusi skewness positif:
[10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500].

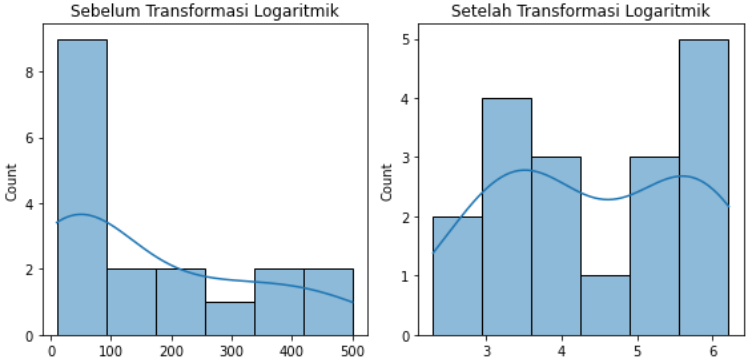
Data terlihat sebagian besar berada pada nilai yang kecil, sedangkan sedangkan beberapa data dengan nilai yang

sangat besar. Setelah transformasi dilakukan, data akan menjadi sebagai berikut:
[2.30, 2.71, 2.99, 3.22, 3.40, 3.56, 3.69, 3.81, 3.91, 4.61, 5.01, 5.30, 5.52, 5.70, 5.86, 6.00, 6.11, 6.21].

Kita bisa lihat bahwa data baru setelah dilakukan transformasi, memiliki distribusi yang lebih simetris dan tidak lagi condong ke arah nilai yang kecil. Perlu diingat bahwa transformasi logaritmik hanya efektif pada data yang memiliki distribusi yang sangat condong atau terdistorsi dan tidak dianjurkan untuk digunakan pada data yang sudah terdistribusi secara merata. Untuk lebih jelasnya, kita lakukan contoh diatas menggunakan bahasa pemrograman Python.

```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
# Membuat data contoh yang berdistribusi skewness positif
data = np.array([10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500])
# Membuat histogram data sebelum dilakukan transformasi logaritmik
plt.figure(figsize=(8, 4))
plt.subplot(1, 2, 1)
sns.histplot(data, kde=True)
plt.title("Sebelum Transformasi Logaritmik")
# Melakukan transformasi logaritmik pada data
data_log = np.log(data)
# Membuat histogram data setelah dilakukan transformasi logaritmik
plt.subplot(1, 2, 2)
sns.histplot(data_log, kde=True)
plt.title("Setelah Transformasi Logaritmik") plt.tight_layout()
plt.show()
```

Setelah dijalankan, maka hasilnya seperti pada gambar 5.1




**Gambar 5.1.** Transformasi Data dengan Transformasi Logaritmik
(Sumber : Dokumentasi Pribadi)

## 2. Transformasi Arc Sin

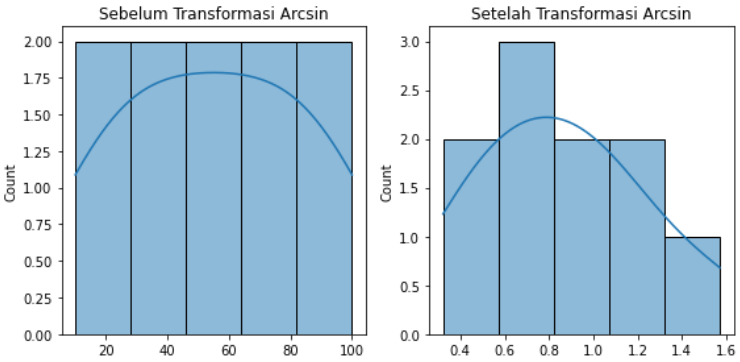
Transformasi arcsin atau transformasi angular merupakan salah satu jenis transformasi data yang digunakan untuk menormalkan distribusi data proporsi atau persentase yang terbatas pada rentang tertentu. Transformasi ini digunakan untuk mengubah data persentase menjadi nilai yang terdistribusi secara normal (Duchesne, 2003), atau menjadi nilai yang berkisar antara -1 hingga 1. Transformasi ini menggunakan fungsi inverse sine atau arcsine. Dalam transformasi ini, nilai persentase *x* akan diubah menjadi arcsine dari akar kuadrat dari *x*.

Contoh penggunaan transformasi arcsin dalam transformasi data, dapat kita lihat pada potongan kode berikut:

```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Membuat data contoh persentase
```

```
data_percent = np.array([10, 20, 30, 40, 50, 60, 70, 80, 90, 100])

# Membuat histogram data sebelum dilakukan transformasi arcsin
plt.figure(figsize=(8, 4))
plt.subplot(1, 2, 1)
sns.histplot(data_percent, kde=True)
plt.title("Sebelum Transformasi Arcsin")

# Melakukan transformasi arcsin pada data persentase
data_arcsin = np.arcsin(np.sqrt(data_percent/100))

# Membuat histogram data setelah dilakukan transformasi arcsin
plt.subplot(1, 2, 2)
sns.histplot(data_arcsin, kde=True)
plt.title("Setelah Transformasi Arcsin")
plt.tight_layout()
plt.show()
```

Kita dapat melihat bahwa data ini memiliki rentang nilai antara 0 hingga 100. Dengan melakukan transformasi arcsin pada data ini dengan rumus *arcsin(sqrt(x/100))*, dimana *x* adalah nilai persentase. Setelah dijalankan potongan kode tersebut, maka akan menghasilkan visualisasi seperti pada gambar 5.2.




**Gambar 5.2.** Transformasi Data dengan Transformasi Arc Sin (Sumber : Dokumentasi Pribadi)

Pada grafik sebelah kiri, kita dapat melihat bahwa distribusi data persentase sangat condong ke arah nilai yang tinggi. Namun, setelah dilakukan transformasi arcsin pada data dan membuat histogram pada grafik sebelah kanan, kita dapat melihat bahwa distribusi data menjadi lebih simetris dan terdistribusi secara normal.

Transformasi arcsin berguna dalam analisis data, terutama pada data yang berkaitan dengan persentase atau proporsi (Kristjansson *et al.*, 2005). Transformasi ini dapat membantu menormalkan data dan memudahkan analisis serta visualisasi data. Namun, perlu diingat bahwa transformasi ini hanya efektif pada data yang memiliki rentang nilai persentase yang terbatas.

### 3. Transformasi Akar Kuadrat

Transformasi akar kuadrat atau *square root transformation*, merupakan jenis transformasi data yang digunakan untuk memperbaiki distribusi data yang tidak sesuai dengan asumsi-asumsi analisis ragam. Transformasi ini

dilakukan dengan mengambil akar kuadrat dari setiap nilai data, sehingga data akan terdistribusi lebih simetris (Schumacker and Lomax, 2004). Biasanya, transformasi ini sering digunakan pada data yang memiliki varian yang tidak konstan atau heteroskedastisitas, di mana variansinya meningkat seiring dengan meningkatnya nilai rata-rata. Selain itu, transformasi ini juga berguna pada data yang memiliki kecondongan atau ketimpangan pada distribusinya. Dengan kata lain, transformasi akar kuadrat ini dapat digunakan pada data yang memiliki nilai yang besar atau memiliki distribusi *skewness* positif. Transformasi ini akan mengurangi nilai yang sangat besar dan mengangkat nilai yang kecil sehingga distribusi data akan terlihat lebih simetris dan memenuhi asumsi-asumsi analisis ragam.

Rumus untuk transformasi akar kuadrat diformulasikan sebagai berikut:

$y = \text{sqrt}(x)$ ……(1)

Di mana:

*y* adalah nilai hasil transformasi

*x* adalah nilai asli data

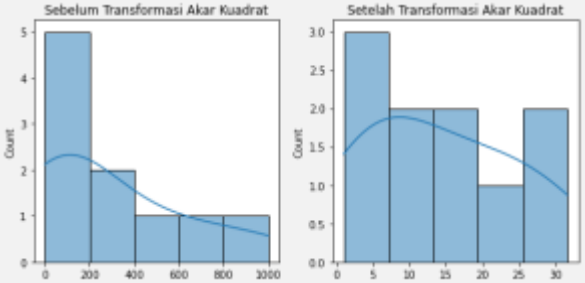
Contoh penerapan transformasi akar kuadrat dapat dilihat pada potongan kode program berikut ini:

```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Membuat data contoh dengan distribusi yang skewness positif
data = np.array([1, 2, 3, 4, 5, 6, 7, 8, 9, 10])**3

# Mencari nilai mean dan variance dari data
mean = np.mean(data)
variance = np.var(data)
```

```
print("Nilai mean awal:", mean)
print("Nilai variance awal:", variance)

# Membuat histogram data sebelum dilakukan transformasi
plt.figure(figsize=(8, 4))
plt.subplot(1, 2, 1)
sns.histplot(data, kde=True)
plt.title("Sebelum Transformasi Akar Kuadrat")

# Melakukan transformasi akar kuadrat pada data
transformed_data = np.sqrt(data)

# Mencari nilai mean dan variance dari data setelah transformasi
mean_transformed = np.mean(transformed_data)
variance_transformed = np.var(transformed_data)

print("Nilai mean setelah transformasi:", mean_transformed)
print("Nilai variance setelah transformasi:", variance_transformed)

# Membuat histogram data setelah dilakukan transformasi
plt.subplot(1, 2, 2)
sns.histplot(transformed_data, kde=True)
plt.title("Setelah Transformasi Akar Kuadrat")
plt.tight_layout()
plt.show()
```

Output dari potongan kode program diatas dapat kita lihat sebagai berikut:

```
Nilai mean awal: 302.5
Nilai variance awal: 106334.25 N
ilai mean setelah transformasi: 14.26723106687563
Nilai variance setelah transformasi: 98.94611768437893
```




Pada contoh ini, kita memiliki data dengan distribusi yang *skewness* positif. Pada data asli kita dapat melihat bahwa nilai *mean* dan *variance* sangat tinggi. Setelah dilakukan transformasi akar kuadrat pada data, nilai *mean* dan *variance* data menjadi lebih rendah dan lebih stabil. Hal ini dapat kita lihat pada histogram grafik sebelah kanan, dimana distribusi data menjadi lebih simetris dan terdistribusi secara normal.

Perlu diperhatikan bahwa transformasi akar kuadrat hanya berguna untuk data dengan skewness positif. Pada data dengan distribusi normal atau skewness negatif, transformasi ini mungkin tidak diperlukan.

### 4. Transformasi z-score

Transformasi Z-score atau *standardization* merupakan teknik transformasi data yang digunakan untuk mengubah skala pengukuran data asli menjadi skala standar, sehingga nilai rata-ratanya adalah 0 dan standar deviasinya adalah 1.

Transformasi ini dilakukan dengan menghitung selisih antara nilai data dengan rata-rata, kemudian hasilnya dibagi dengan standar deviasi. Nilai Z-score menggambarkan seberapa jauh setiap nilai data dari nilai rata-rata, diukur dalam satuan standar deviasi (Hair *et al.*, 2018). Tujuan dari transformasi Z-score adalah untuk membandingkan nilai data yang berasal dari populasi atau sampel yang berbeda dalam satuan yang sama. Transformasi ini sering digunakan dalam analisis statistik, seperti analisis regresi dan analisis faktor.

Rumus untuk menghitung nilai Z-score adalah sebagai berikut:

$$z = (x - mean) / std \quad \text{.....(2)}$$

Di mana:

*z* adalah nilai Z-score
*x* adalah nilai data asli
*mean* adalah nilai rata-rata data
*std* adalah standar deviasi data

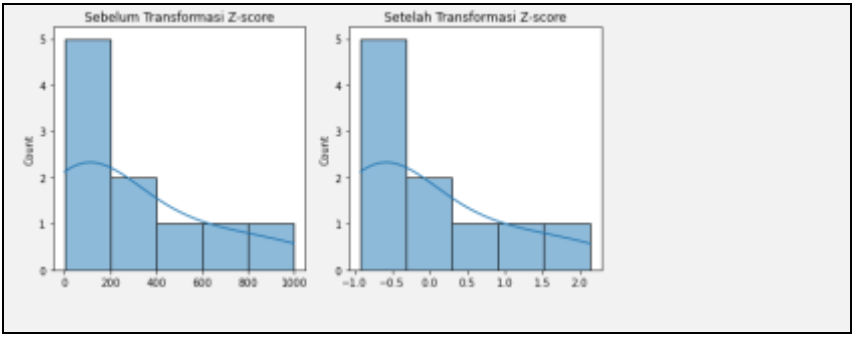
Contoh penerapan transformasi Z-score dapat dilihat pada potongan kode program berikut ini:

```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Membuat data contoh dengan distribusi yang skewness positif
data = np.array([1, 2, 3, 4, 5, 6, 7, 8, 9, 10])**3

# Menghitung nilai mean dan standar deviasi dari data
mean = np.mean(data)
std = np.std(data)

# Membuat histogram data sebelum dilakukan transformasi
plt.figure(figsize=(8, 4))
```

```
plt.subplot(1, 2, 1)
sns.histplot(data, kde=True)
plt.title("Sebelum Transformasi Z-score")

# Melakukan transformasi Z-score pada data
z_score_data = (data - np.mean(data)) / np.std(data)

print("Data asli:", data)
print("Nilai mean:", mean)
print("Nilai standar deviasi:", std)
print("Data setelah transformasi Z-score:", z_score_data)

# Membuat histogram data setelah dilakukan transformasi
plt.subplot(1, 2, 2)
sns.histplot(z_score_data, kde=True)
plt.title("Setelah Transformasi Z-score")
plt.tight_layout()
plt.show()
```

Output dari potongan kode program diatas dapat kita lihat sebagai berikut:

```
Data asli: [  1   8  27  64 125 216 343 512 729 1000]
Nilai mean: 302.5
Nilai standar deviasi: 326.0893282522444
Data setelah transformasi Z-score: [-0.92459328 -0.90312676 -
0.84486052 -0.73139468 -0.54432937 -0.26526474
0.1241991  0.64246199  1.30792382  2.13898444]
```




Pada contoh potongan kode program, setelah dilakukan transformasi, nilai data diubah menjadi nilai Z-score yang terdistribusi secara normal, dengan nilai rata-rata sebesar 0 dan standar deviasi sebesar 1. Perlu diperhatikan bahwa transformasi Z-score dapat digunakan pada data numerik dan kategorikal yang dikonversi menjadi data numerik.

Penting untuk melakukan normalisasi data sebelum melakukan analisis yang memerlukan perbandingan antara data, seperti analisis regresi dan analisis faktor.

## 5.3 Analisis Univariat

### 5.3.1 Pengertian Analisis Univariat

Analisis univariat adalah metode analisis statistik yang digunakan untuk menganalisis satu variabel atau data tunggal pada suatu waktu (Tabachnick and Fidell, 2018). Metode ini bertujuan untuk mengidentifikasi, menyajikan, memahami dan menggambarkan karakteristik dari variabel atau data tunggal dalam bentuk statistik deskriptif seperti *mean*, *median*, *modus*, *range*, *standar deviasi*, dan *persentil* (Hair *et al.*, 2018). Berbagai metode statistik dapat digunakan dalam analisis univariat, seperti tabel frekuensi, histogram, diagram batang, dan lain-lain. Analisis univariat juga dapat memberikan gambaran tentang distribusi data, termasuk simetri, skewness, dan kurtosis. Analisis univariat

sangat penting dalam statistika, karena dapat memberikan informasi dasar tentang data yang diolah (Salkind and Frey, 2019). Selain sebagai alat deskriptif, analisis univariat juga dapat digunakan sebagai alat inferensial. Dalam hal ini, teknik-teknik statistik inferensial, seperti uji hipotesis, dapat digunakan untuk menarik kesimpulan tentang populasi berdasarkan sampel yang digunakan dalam analisis univariat.

### 5.3.2 Teknik Dasar Analisis Univariat

Berikut beberapa teknik dasar yang dilakukan dalam analisis univariat:

**1) Ukuran Pemusatan Data**

Ukuran pemusatan data dalam statistik deskriptif adalah teknik perhitungan untuk menunjukkan nilai tengah atau pusat dari suatu variabel atau dari suatu distribusi data. Teknik ini melibatkan penghitungan statistik deskriptif seperti *mean* (rata-rata), *median*, dan *modus*.

*Mean* (rata-rata) adalah nilai rata-rata dari suatu distribusi data. *Mean* dihitung dengan menjumlahkan semua nilai data dalam dataset dan kemudian membaginya dengan jumlah data.

*Median* adalah nilai tengah dari data setelah diurutkan dari nilai yang terkecil hingga yang terbesar. *Median* sering digunakan ketika terdapat data outlier atau data yang sangat berbeda dengan nilai data lainnya. *Median* dapat memberikan gambaran yang lebih akurat tentang nilai tengah data dalam kasus seperti ini.

*Modus* adalah nilai yang paling sering muncul dalam dataset. *Modus* sering digunakan ketika data memiliki bentuk distribusi yang tidak normal atau simetris. Modus digunakan untuk data yang diukur dalam skala nominal, ordilan, interval, dan rasio.

Ketiga ukuran pemusatan data ini memberikan informasi yang berbeda tentang nilai tengah atau pusat dari suatu distribusi data.

**2) Ukuran Sebaran Data**

Ukuran sebaran data merupakan suatu ukuran yang menyatakan seberapa jauh data menyebar dari rata-ratanya, atau seberapa jauh nilai-nilai data bervariasi dari nilai pusatnya. Beberapa ukuran yang dapat digunakan termasuk jangkauan (*range*), simpangan kuartil (*quartile deviation*), simpangan baku (*standard deviation*), dan ukuran lainnya.

**3) Distribusi Frekuensi**

Distribusi frekuensi adalah suatu cara untuk menyajikan data dalam bentuk tabel yang menunjukkan jumlah frekuensi atau jumlah kejadian dari setiap nilai atau kelompok nilai dalam kumpulan data.

Terdapat dua jenis distribusi frekuensi, yakni distribusi frekuensi absolut dan distribusi frekuensi relatif. Distribusi frekuensi absolut menunjukkan jumlah pengamatan di setiap interval atau kategori, sementara distribusi frekuensi relatif mengindikasikan proporsi pengamatan di setiap interval atau kategori dalam kaitannya dengan jumlah total pengamatan. Tabel distribusi frekuensi umumnya terdiri dari kolom-kolom seperti batas bawah, batas atas, lebar interval, dan frekuensi.

Misalkan kita memiliki data berikut mengenai nilai ujian Bahasa Indonesia siswa-siswa di sebuah sekolah:
75, 80, 85, 90, 70, 75, 85, 80, 90, 95, 70, 75, 80, 85, 80, 85, 75, 90, 80, 85

Dengan mengelompokkan nilai-nilai tersebut, kita dapat membuat distribusi frekuensi untuk data ini. Hasilnya bisa ditampilkan dalam bentuk tabel seperti ini:

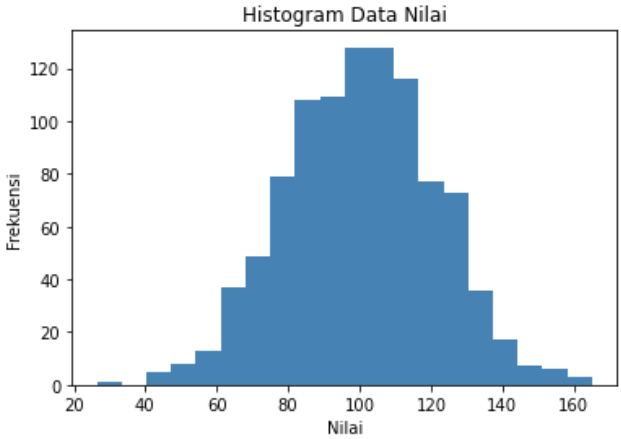
**Tabel 5.1.** Data Distribusi Frekuensi

| Interval | Frekuensi |
|---|---|
| 70-74 | 2 |
| 75-79 | 4 |
| 80-84 | 7 |
| 85-89 | 5 |
| 85-89 | 2 |

**4) Histogram**

Histogram adalah sebuah teknik analisis univariat yang digunakan untuk mempelajari frekuensi nilai yang dimiliki oleh suatu data. Histogram menggambarkan bagaimana data terdistribusi dalam rentang nilai tertentu dengan mengelompokkan data ke dalam kelas atau interval, lalu menunjukkan frekuensi atau jumlah pengamatan dalam setiap interval dengan menggunakan bar. Rentang nilai variabel numerik diposisikan pada sumbu horizontal (*x-axis*) dan frekuensi diposisikan pada sumbu vertikal (*y-axis*).

Histogram berguna untuk memvisualisasikan bentuk distribusi data dan mengidentifikasi pola atau informasi lain yang mungkin sulit dilihat dengan hanya menggunakan statistik deskriptif seperti mean dan standar deviasi.




**Gambar 5.3.** Histogram
(Sumber : Dokumentasi Pribadi)

## 5.4 Kesimpulan

Dalam analisis univariat, transformasi data adalah proses mengubah data dari bentuk awalnya menjadi bentuk lain yang lebih sesuai untuk analisis statistik tertentu. Transformasi data dapat membantu meningkatkan kualitas analisis dan interpretasi data, serta membantu memenuhi asumsi statistik tertentu.

Beberapa jenis transformasi data yang umum dilakukan dalam analisis univariat meliputi transformasi logaritmik, transformasi akar, dan transformasi arc sin, transformasi z-score, dan masih banyak lagi lainnya. Transformasi data juga dapat membantu mengatasi masalah seperti skewness atau asimetri data dan heteroskedastisitas atau ketidakhomogenan variansi data.

Setelah dilakukan transformasi data, analisis univariat dapat dilakukan dengan lebih baik. Beberapa teknik analisis univariat yang dapat digunakan setelah dilakukan transformasi data meliputi distribusi frekuensi, histogram, diagram batang, dan scatter plot. Teknik-teknik ini membantu memvisualisasikan data,

mengidentifikasi pola atau tren dalam data, serta memperoleh informasi statistik yang berguna untuk analisis lebih lanjut.

Dengan demikian, transformasi data dan analisis univariat adalah dua konsep yang erat kaitannya dalam statistik. Transformasi data dapat membantu meningkatkan kualitas analisis univariat, sementara analisis univariat dapat membantu memahami data setelah dilakukan transformasi.

# DAFTAR PUSTAKA


Duchesne, P. 2003. 'Estimation of a Proportion with Survey Data', *Journal of Statistics Education*, 11(3). doi: 10.1080/10691898.2003.11910725.

Hair, J. F. *et al.* 2018. *Multivariate Data Analysis*. Cengage Learning.

Inmon, W. H. 2005. *Building the Data Warehouse 4th Edition*. 4th edn. Wiley.

Kristjansson, E. *et al.* 2005. 'A Comparison of Four Methods for Detecting Differential Item Functioning in Ordered Response Items', *Educational and Psychological Measurement*, 65(6), pp. 935–953. doi: 10.1177/0013164405275668.

Ng, A. 2018. *Machine Learning Yearning*, *deeplearning.ai*.

Rosadi, D. 2017. *Analisis Statistika dengan R*. Yogyakarta: Gadjah Mada University Press.

Salkind, N. J. and Frey, B. B. 2019. *Statistics for People Who (Think They) Hate Statistics*. Sage.

Schumacker, R. E. and Lomax, R. G. 2004. *A Beginner's Guide to Structural Equation Modeling*. Psychology Press. doi: 10.4324/9781410610904.

Tabachnick, B. G. and Fidell, L. S. 2018. *Using Multivariate Statistics*. 7th editio. Pearson.

Witten, I. H., Frank, E. and Hall, M. A. 2011. *Data Mining: Practical Machine Learning Tools and Techniques*. Elsevier. doi: 10.1016/C2009-0-19715-5.

# BAB 6
# ANALISIS DESKRIPTIF VARIABEL KATEGORIKAL DAN NUMERIKAL, PENYAJIAN DAN INTERPRETASI ANALISIS SECARA UNIVARIAT

*Oleh Herda Ariyani*

## 6.1 Dasar-Dasar Statistik Farmasi

Dalam pembicaraan yang lalu kita telah mempresentasikan data dalam bentuk tabel dan grafik yang bertujuan meringkas dan menggambarkan data kuantitatif, untuk mendapatkan gambaran yang lebih jelas tentang sekumpulan data. (Sugiyarto, 2015)

Analisis deskripsi merupakan analisis statistik yang paling mendasar untuk menggambarkan keadaan data secara umum. Pada dasarnya ruang lingkup statistik deskriptif meliputi beberapa hal, yaitu:

1. Penyusunan tabel frekuensi atau distribusi frekuensi.
2. Mengukur nilai-nilai tendensi sentral seperti rata-rata, median, dan modus.
3. Mengukur nilai-nilai fractile seperti Quartil, Quantil, desil, dan persentil.
4. Mengukur nilai-nilai disperse, seperti range, semi-interquartil range, semi 10-90 persentil range, mean deviation, standar deviasi, variasi, dan koefisien variasi.
5. Mengukur koefisien kemencengan (*skewness*).
6. Mengukur koefisien keruncingan (kurtosis)
7. Membuat gambaran visual dalam bentuk grafik seperti histogram, polygon, dan ogive.

Untuk mendapatkan gambaran data yang lebih jelas, selain disajikan dalam bentuk tabel atau diagram, diperlukan juga ukuran-ukuran yang blok (RT) yang terpilih dan dilakukan penelitian.(Sugiyarto,2015).

## 6.2 Data Dan Variabel Penelitian

### 6.2.1 Data Statistik

**Data** menurut bentuknya ada dua macam yaitu kualitatif dan kuantitatif. Statistik hanya bekerja dengan data kuantitafif atau data kualitatif yang sudah dikuantitatifkan. **Data kuantitatif** adalah fakta yang dinyatakan dengan angka. Misalnya tinggi badan mahasiswa, dan sebagainya. **Data kualitatif** adalah data yang dinyatakan dalam bentuk bukan angka. Misalnya profesi, agama, dan sebagainya. Data kualitatif dapat dikuantitatifkan antara lain dengan cara memberi skor, ranking, dan sebagainya.

Data yang dikumpulkan bersifat diukur secara langsung dan tidak dapat diukur secara langsung. Untuk sifat yang kedua perlu dibuat sedemikian hingga secara operasional dapat diukur, yaitu diusahakan untuk memecah atau menguraikan data itu dalam sejumlah dimensi yang dapat diukur. Misalnya status sosial ekonomi dapat dioperasionalkan menjadi dimensi pendapatan dan dimensi status pekerjaan, intelegensia dapat dioperasionalkan menjadi tes. Intelegnsi yang terdiri dari beberapa soal, yang setiap soalnya merupakan satu dimensi, dan sebagainya. (Sugiyarto,2015)

Dalam propses operasionalisasi harus diperhatikan bahwa sifat hakikat pengertian tidak berubah dan validitas (keabsahan) pengukuran cukup baik. Beberapa skala pengukuran yang digunakan adalah:

**a. Skala Nominal (Klasifikasi)**

Skala nominal adalah skala pengukuran berupa bilangan atau lambing-lambang lain untuk mengelompokkan obyek atau orang, atau benda-benda lain. Misalnya, status

pendidikan dikelompokkan menjadi sekolah atau tidak sekolah, jenis kelamin dikelompokkan menjadi laki-laki atau wanita, dan sebagainya. Kelas atau kategori adalah titik skala nominal. Dalam hal Janis kelamin, titik skala adalah laki-laki dan wanita. (Sugiyarto,2015)

**b. Skala Ordinal (Ranking)**

Skala ordinal adalah skala pengukuran yang mengelompokkan objak-objek ke dalam kelas-kelas yang mempunyai hubungan urutan satu dengan yang lain. Hubungan antar kelas-kelas adalah lebih baik, lebih disukai, lebih tinggi dan sebagainya. Misalnya, seorang anggota ABRI dapat dikelompokkan menurut pangkatnya: himpunan Mayor, himpunan Kapten, himpunan Letnan dan sebagainya. Hubungan antar kelas-kelas terdapat urutan tertentu, pangkat Mayor lebih tinggi dari pangkat Letnan (Mayor>Kapten>Letnan). (Sugiyarto,2015)

**c. Skala Interval**

Skala interval adalah skala pengukuran yang mengelompokkan objek-objek ke dalam kelas-kelas yang mempunyai hubungan urutan dan perbedaan dalam jarak (*Interval*) satu dengan yang lain.

Ciri-ciri skala interval:

1) Unit pengukuran sama dengan konstan.
2) Perbandingan antara dua interval sembarang adalah independen dengan unit pengukuran dan titik nolnya.
3) Titik nol dan unit pengukuran sembarang *(Arbitrary)*
   Contoh skala interval adalah tahun-tahun almanac, skala temperatur.(Sugiyarto,2015)

**d. Skala Rasio**

Skala rasio adalah skala pengukuran yang mengelompokkan objek-objek ke dalam kelas-kelas yang mempunyai hubungan, urutan dan berbeda dalam jarak antara objek yang satu dengan yang lain serta titik nolnya

berarti (tidak sembarang). Contoh skala rasio adalah skala untuk mengukur panjang, luas, isi, berat, tinggi, dan sebagainya. (Sugiyarto,2015).

### 6.2.2 Deskriptif Kategori: Nominal dan Ordinal

1. Pengertian dan klasifikasi

   Variabel nominal dan ordinal disebut sebagai variabel kategorik karena mempunyai kategori. Sebagai contoh jenis kelamin adalah variabel, sedangkan laki-laki dan perempuan adalah kategori variabel; klasifikasi kadar kolesterol adalah variabel, sedangkan baik, sedang baik, sedang, dan buruk adalah kategorinya. (Dahlan, 2014).

   Berdasarkan kategori inilah Anda dapat membedakan nominal dari ordinal. Nominal mempunyai kategori yang sederajat (contoh: variabel jenis kelamin dengan kategori laki-laki dan perempuan) sedangkan ordinal mempunyai kategori yang bertingkat (contoh: variabel kolesterol dengan kategori kadar kolesterol baik, kadar kolesterol sedang, dan kadar kolesterol buruk). (Dahlan, 2014).

   Berdasarkan jumlah kategori, variabel kategorikal diklasifikasikan menjadi dikotom dan polimokotom. Yang pertama mempunyai dua kategori contoh nya jenis kelamin (laki-laki dan perempuan). Yang kedua mempunyai lebih dari dua kategori, misal nya tingkat pendidikan (rendah, menengah, dan tinggi). (Dahlan, 2014).

   Sensitivitas, spesifikasi, dan *area under the curve* (AUC) termasuk ke dalam skala kategorik karena ketiganya adalah proposi. (Dahlan, 2014).

2. Cara penyajian

   Variabel kategorik dideskripsikan dalam jumlah (n) dan persentase (%)(Dahlan, 2014). Tabel 1 merupakan contoh penyajian variabel kategorik dalam bentuk tabel.

### 6.2.3 Deskriptif Numerik: Rasio dan Interval

1. Pengertian dan klasifikasi

   Varibel rasio dan interval disebut sebagai variabel numerik karena tidak mempunyai kategori variabel. Rasio dan interval bergantung pada nilai nolnya. Jika variabel mempunyai nilai nol alami (seperti tinggi badan, berat badan, dan jarak), maka ia adalah rasio. Jika variabel tidak mempunyai nilai nol alami (seperti suhu), maka ia adalah interval. Perhatikan bahwa nol derajat pada skala Celcius berbeda dengan nol derajat pada skala *Fahrenheit*. Variabel numerik juga diklasifikasikan menjadi variabel diskrit dan kontinyu. Contoh masing-masing adalah jumlah anak dan berat badan disebut kontinyu karena merupakan hasil mengukur. (Dahlan, 2014).

2. Perubahan skala pengukuran

   Suatu variabel numerik dapat berubah menjadi kategorik. Kadar gula darah adalah numerik. Jika diklasifikasikan menjadi diabetes dan tidak diabetes, maka skalanya berubah menjadi kategorik. Kadar hemoglobin adalah numerik. Jika diklasifikasikan menjadi anemia dan tidak anemia, maka skalanya menjadi kategorik. Suatu variabel numerik bisa dianggap kategorik jika variasinya tidak terlalu banyak. Contohnya adalah jumlah anak. Jika variasi jumlah anak hanya satu atau dua maka kita bisa mengganggapnya sebagai kategorik sehingga disajikan sebagaimana variabel kategorik (Dahlan, 2014).

3. Cara penyajian.

   Variabel numerik dideskripsikan dalam ukuran pemusatan dan ukuran penyebaran. Ukuran pemusatan adalah *mean* dan median. Ukuran penyebaran adalah simpang baku dan persentil. Variabel numerik juga dapat dideskripsikan dalam bentuk grafik. Grafik yang sering digunakan adalah *error bar* dan *box plot.* Jika distribusi (sebaran) data normal

maka kita memilih *mean* untuk ukuran pemusatan, simpang baku (s.b) untuk ukuran penyebaran, dan *error bar* untuk gambar. Jika distribusi tidak normal maka kita memilih median untuk ukuran pemusatan dan persentil sebagai ukuran penyebaran. Nilai persentil yang sering digunakan adalah minimum-maksimum, persentil 25-persentil 75, persentil 2,5-persentil 97,5, persentil 5-persentil 95, dan persentil 10-persentil 90. Salah satu cara untuk mengetahui sebaran data adalah dengan histogram. Data dikatan normal apabila histogram simetris membentuk kurva terbalik (kurva Gauss).

Cara lain untuk mengetahui sebaran data dijelaskan pada bagian selanjutnya. (Dahlan, 2014). Persentil merupakan urutan peringkat data. Data bias kita urutkan mulai dari yang terkecil sampai dengan terbesar. Beberapa peringkat yang cukup populer adalah minumum (persentil satu), kuartil satu (persentil 25), median (persentil 50), kuartil dua (persentil 50), kuartil tiga (persentil 75), dan maksimum (persentil 100). Bila jumlah subjek relatif sedikit, minimal -maksimal dapat digunakan sebagai ukuran penyebaran. Bila sejumlah subjek relatif banyak, kita dapat memilih persentil lainnya.

Contoh sistem persentil yang cukup populer yang cukup populer adalah kurva pertumbuhan. Garis paling bawah pada kurva adalah persentil tiga. Garis yang berada di tengah adalah median. Sementara, garis paling atas adalah persentil 97. Jika berat badan seorang anak berada dibawah persentil tiga, maka anak tersebut beratnya kurang. Jika diatas persentil 97, maka beratnya lebih. Usia dideskripsikan dalam rerata (pemusatan) dan simpang baku (penyebaran) karena sebaran datanya normal. Selanjutnya, lama perawatan dideskripsikan dalam median (pemusatan) dan minimum-maksimum (penyebaran) karena tidak normal. Minumum-maksimum digunakan karena jumlah subjek sedikit (n=30). (Dahlan,2014). Cara penulisan rerata dan simpang baku

ada dua alternatif, yaitu rerata s.b. atau rerata (s.b). Sebagai kesepakatan, buku ini menggunakan alternatif kedua. Untuk publikasi, sebaiknya ikuti aturan yang ditetapkan oleh institusi atau jurnal tempat publikasi. Penulisan rerata dan simpang baku tidak harus selalu presisi. Jumlah desimal cukup satu desimal lebih banyak dibandingkan dengan *raw data.* Untuk variabel umur, jumlah desimal ada dua karena jumlah desimal pada *raw data* satu.

Data disajikan dalam rerata (simpang baku) untuk data berdistribusi normal; dalam median (minimal-maksimal) untuk data berdistribusi tidak normal. Selanjutnya, variabel numerik dapat juga disajikan dalam bentuk gambar. Usia disajikan dalam *error bar* karena sebaran datanya normal. Sementara, lama perawatan disajikan dalam *box plot* karena tidak normal.(Dahlan, 2014).

### 6.2.4 Rate

Selain skala kategorik dan numerik, terdapat satu skala pengukuran yang perlu diketahui, yaitu skala rate. Rate adalah jumlah kejadian dibagi dengan *person time* (waktu orang). (Dahlan, 2014).

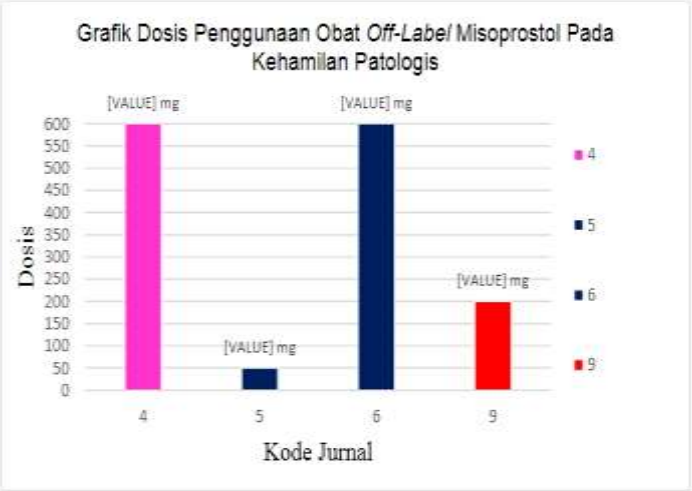
## 6.3 Penggambaran Distribusi Frekuensi

Untuk lebih mempermudah dalam memahami dan menganalisi dan disamping disajikan dalam bentuk tabel distribusi frekuensi, cara yang lain adalah dengan menggambarkan distribusi tersebut dalam bentuk grafik.(Sugiyarto,2015). Cara lain untuk menyajikan data adalah dengan menggunakan grafik (plot). Penyajian secara grafis adalah cara yang baik dan format ini mengindikasikan adanya tren baik pada pelaku percobaan maupun pembaca suatu laporan imiah (Jones, 2010). Tipe grafik yang dipilih pada dasarnya adalah pilihan penulis yang melakukan studi; namun, diluar dari pilihan yang dibuat, grafik harus menyajikan data secara akurat. Penjelasan yang cukup harus terdapat pada tiap

grafik untuk menjamin interprestasinya oleh ilmuan-ilmuan yang lain (Jones, 2010).

Selain itu, bab ini memperkenalkan konsep distribusi frekuensi dan plot-plot terkait serta pembentukannya. Pengunaan dan interprestasi distribusi frekuensi merupakan aspek statistik terpadu dan akan dilanjutkan dalam bab-bab berikutnya. Beberapa grafik yang dibahas disini adalah histogram, polygon, ogive, dan pie chart.

a. Histogram
   Untuk menggambarkan grafik ini interval kelas ditentukan pada sumbu X dan sumbu Y. untuk menggmbar grafik distribusi frekuensi realtif, cara adalah: interval kelas diletakkan pada sumbu Y, dengan tinggi persegi panjang = Frekuensi relative interval kelas
   Lebar interval kelas.

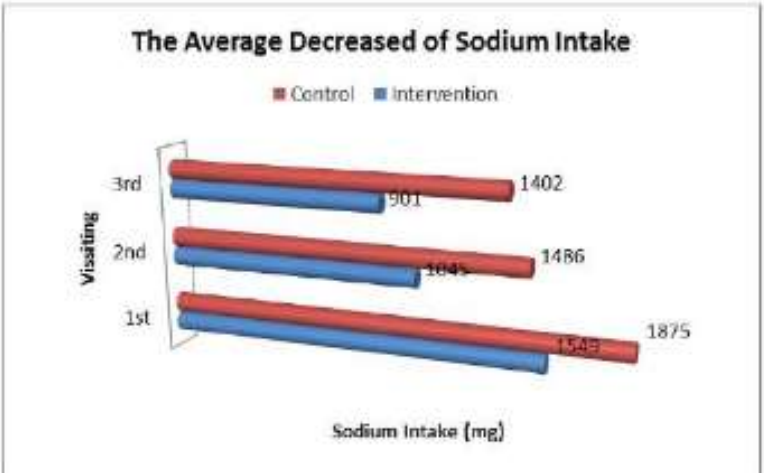



**Gambar 6.1.** Histogram Grafik Dosis Penggunaan Obat Off-LabelMisoprostol Pada Kehamilan Patologis (Ningrum dkk, 2020)

b. Poligon

Cara menggambar poligon

- Tentukan titik tengah interval kelas diletakkan pada sumbu X
- Frekuensi setiap kelas pada sumbu Y
- Tarik gari slurus dari titik tengah interval kelas, tingginya sesuai dengan frekuensi
- Hubungkan antar puncak garis-garis tersebut dengan garis lurus.




**Gambar 6.2.** Rata-rata Nilai Penurunan Asupan Natrium pada Kedua Kelompok di Setiap Kunjungan(Ariyani dkk, 2017; Indriani dkk, 2021)

c. Ogive

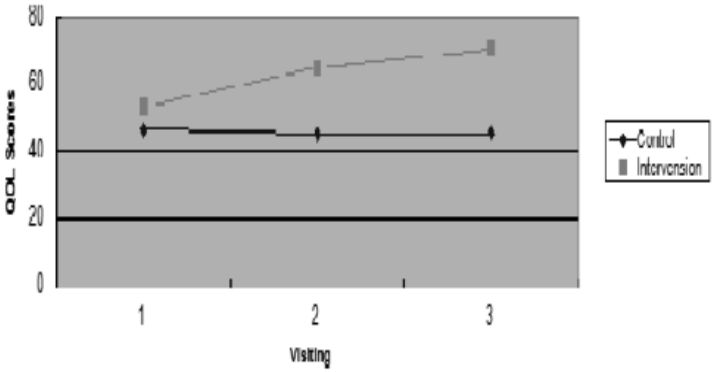
Grafik ini digunakan untuk menggambar distribusi frekuensi kumulatif.

Cara menggambar:

- Harga batas interval kelas diletakkan pada sumbu Y
- Frekuensi diletakkan pada sumbu Y
- Tentukan Koordinat titik-titik dengan

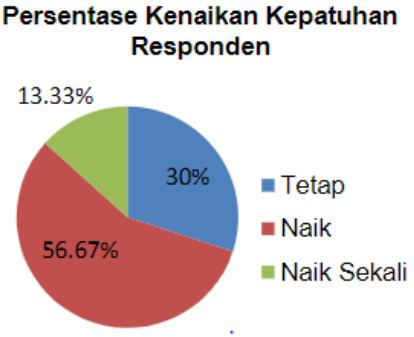
Absis : Batas Interval Kelas
Ordinat : Frekuensi

- Hubungkan antar titik-titik tersebut. (Sugiyarto,2015).




**Gambar 6.3.** Rata-rata Skor Kualitas Hidup Pasien Hipertensi Kelompok Kontrol dan Intervensi pada Setiap Kunjungan
(Ariyani dkk, 2015)

*d.* Grafik Lingkaran *(Pie Chart)*




**Gambar 6.4.** Persenatse Kenaikan Kepatuhan Responden
(Ariyani dkk, 2018)


**Gambar 6.5.** Gambaran Pola Pengobatan Warga (Ariyani & Rahayu, 2017)

Untuk menggambarkan grafik lingkaran, kita perlu menggambarkan sebuah lingkaran kemudian dibagi-bagi menjadi beberapa bagian. Misalkan kita ingin membuat deskripsi perbedaan kenaikan kepatuhan responden dalam penelitian, maka dapat digambarkan dalam pie chart seperti gambar 4 maupun gambar 5 di atas.

e. Tabel

Bentuk penyajian ini merangkum data ke dalam bentuk baris atau kolom. Adapun baris dan kolom tersebut dapat berupa kategori-kategori dan angka frekuensi (Otok & Ratnaningsih, 2022). Dapat Anda lihat contoh pada tabel 1 di bawah ini.

**Tabel 6.1.** Gambaran perubahan perilaku masyarakat dalam mengonsumsi vitamin pada saat pendemik covid-19 di Apotek X

| No | Kategori | Jumlah | |
|---|---|---|---|
| | | N | % |
| 1 | *Precontemplation* | 2 | 6,66 |
| 2 | *Contemplation* | 8 | 26,66 |
| 3 | *Preparation* | 2 | 6,66 |
| 4 | *Action* | 11 | 36,66 |
| 5 | *Maintenance* | 5 | 16,66 |

(Ariyani dkk, 2022)

Penyajian data berbentuk tabel merupakan salah satu metode penyajian yang paling umum, dan biasanya merupakan hasil dari pengumpulan dan penyaringan hasil-hasil percobaan (Jones, 2010).

## 6.4 Distribusi Frekuensi Kumulatif

Metode lain untuk penyajian grafis data dapat berupa suatu distibusi frekuensi kumulatif. Dalam metode ini, data khususnya disajikan dalam bentuk frekuensi total dari semua pengamatan yang kurang dari batas atas kelas dari suatu interval kelas. Hal ini dapat dengan mudah diistilahkan sebagai distribusi frekuensi kumulatif' *kurang dari*". Sebagai alternative yang lain, data dapat disajikan dalam bentuk frekuensi total dari pengamatan-pengamatan yang lebih besar dari atau sama dengan batas bawah kelas dari suatu interval kelas tertentu, yaitu distribusi frekuensi kumulatif 'lebih dari'. Sebuah contoh dari tiap tipe-tipe distribusi frekuensi kumulatif ini ditampilkan dibawah ini, menggunakan data yang disajikan dalam Tabel 2 (Jones, 2010)

**Tabel 6.2.** Data distribusi frekuensi kumulatif yang menggambarkan berat 210 tablet yang dipilih dari suatu bets produktif.

| Berat tablet (mg) | Frekuensi kumulatif (kurang dari) | Berat tablet (mg) | Frekuensi kumulatif (lebih dari) |
|---|---|---|---|
| <290,05 | 0 | <290,05 | 210 |
| <291,05 | 2 | <291,05 | 208 |
| <292,05 | 4 | <292,05 | 206 |
| <293,05 | 8 | <293,05 | 202 |
| <294,05 | 14 | <294,05 | 196 |
| <295,05 | 22 | <295,05 | 188 |
| <296,05 | 31 | <296,05 | 179 |
| <297,05 | 52 | <297,05 | 158 |
| <298,05 | 82 | <298,05 | 128 |
| <299,05 | 114 | <299,05 | 96 |
| <300,05 | 150 | <300,05 | 60 |
| <301,05 | 174 | <301,05 | 36 |
| <302,05 | 183 | <302,05 | 27 |
| <303,05 | 190 | <303,05 | 20 |
| <304,05 | 198 | <304,05 | 12 |
| <305,05 | 201 | <305,05 | 9 |
| <306,05 | 205 | <306,05 | 5 |
| <307,05 | 207 | <307,05 | 3 |
| <308,05 | 209 | <308,05 | 1 |
| <309,05 | 210 | <309,05 | 0 |
| <310,05 | 210 | <310,05 | 0 |

( Jones, 2010).

## 6.5 Interprestasi Beberapa Uji Hipotesis Berdasarkan Nilai *P*

Pada bagian sebelumnya, kita telah membahas prinsip uji hipotesis dengan menghitung nilai p. Dalam bahasa praktis sehari-hari, kita sering menyederhanakannya menjadi *bermakna* dan *tidak bermakna.* Tabel 6.3 menyajikan interpretasi untuk hasil penelitian yang bermakna, yaitu ketika nilai p lebih kecil dari 0,05 (hipotesis nol ditolak).

**Tabel 6.3.** Interpretasi beberapa ujian bila nilai p<0,05 (hipotesis nol ditolak).

| No. | Uji hipotesis | Interpretasi bila hipotesis nol di tolak (p<0'05) |
|---|---|---|
| 1 | Uji t tidak berpasangan | terdapat perbedaan rerata antara dua kelompok. |
| 2 | Uji t berpasangan | Terdapat perbedaan rerata antaradua kelompok. |
| 3 | One-way Anova | Paling tidak terdapat perbedaan rerata antara dua kelompok. Untuk mengetahui detilnya, lakukan analisis post hoc. |
| 4 | Rapeated Anova | Paling tidak terdapat perbedaan rerata antaradua kelompok. Untuk mengetahui detilnya, lakukan analisis post hoc. |
| 5 | Mann-Whitney | Terdapat perbedaan rerata anatara dua kelompok. |
| 6 | Wilcoxon | Terdapat perbedaan rerata antardua kelompok. |
| 7 | Kruskal-Wallis | Paling tidak terdapat perbedaan rerata antardua kelompok. Untuk mengetahui detilnya, lakukan analisis post hoc. |

| No. | Uji hipotesis | Interpretasi bila hipotesis nol di tolak (p<0'05) |
|---|---|---|
| 8 | Friedman | Paling tidak terdapat perbedaan antara rerata antardua kelompok. Untuk mengetahui detilnya, lakukan analisis post hoc. |
| 9 | Chi-square | Terdapat perbedaan proporsi antarkelompok. |
| 10 | Fisher | Terdapat perbedaan proporsi antarkelompok. |
| 11 | McNemar | Terdapat perbedaan proporsi antarkelompok. |
| | | 12. Marginal homogeneity, Terdapat perbedaan proporsi antarkelompok. |
| 12 | Cochran | Paling tidak terdapat perbedaan proporsi antarkelompok. Untuk mengetahui detilnya, lakukan analisis post hoc. |
| 13 | Perbandingan AUC | Terdapat perbedaan AUC antarkelompok. |
| 14 | Log rank | Terdapat perbedaan rate antarkelompok. |
| 15 | Uji Bland-Altman | Terdapat perbedaan pengukuran antarkelompok. |
| 16 | Uji Kappa | Nilai kappa bermakna. |
| 17 | Pearson | Terdapat korelasi antarvariabel. |
| 18 | Spearman | Terdapat korelasi antarvariabel. |
| 19 | Uji normalitas | Sebaran data tidak normal. |
| 20 | Uji varian | Varian tidak sama. |

(Dahlan, 2014).

## 6.6 Beberapa Catatan Mengenai Cara Melaporkan Hasil Analisis

Setiap komunitas ilmiah mempunyai tradisi atau kesepakatan bagaimana melaporkan hasil analisi. Demikian juga dengan komunitas ilmiah kedokteran dan kesehatan. Berikut ini merupakan beberapa catatan yang pada umumnya menjadi kesepakatan cara melaporkan hasil penelitian. Perlu menjadi catatan bahwa beberapa institusi dan jurnal mungkin mempunyai kesepakatan yang berbeda.

1. Format tabel yang dipilih adalah tabel tanpa garis vertikal. Garis horizontal hanya untuk baris judul tabel, baris total (jika ada), serta akhir tabel.
2. Deskripsi variabel kategorik disajikan dalam jumlah (n) dan persentase (%).
3. Presentase disajikan satu angka di belakang koma.
4. Deskripsi variabel numerik berdistribusi normal disajikan dalam rerata simpang baku atau rerata (simpang baku). Pada buku ini, penyajian data disajikan dalam rerata (simpang baku).
5. Deskripsi variabel numerik berdistribusi tidak normal disajikan median dan persentil.
6. Persentil yang digunakan bisa minimum-maksimum, persentil 2,5-97,5, presentil 5-90, atau kuartil satu-tiga. Pemilihannya bergatung pada karakter data.
7. Hasil uji hipotesis disajikan dalam nilai p dan interval kepercayaan.
8. Bila harus menyajikan salah satunya, nilai interval kepercayaan lebih diprioritaskan dari pada nilai p.
9. Simpang baku selisih ditujukan pada tabel analitik numerik berpasangan sebaiknya disertakan dalam laporan.

10. Perhatian khusus perlu ditujukan pada tabel analitik pasangan. Kesalahan yang sering terjadi adalah penyajian berpasangan disajikan secara tidak berpasangan.
11. Nilai rerata dan simpang baku disajikan dengan penambahan satu angka di belakanh koma dibandingjan *raw data.* Jika *raw data* satu desimal, maka rerata dan simpang baku disajikan dalam dua desimal.
12. Nilai p disajikan dengan tiga angka di belakang koma.
13. Bila pada perangkat lunak didapatkan nilai p=0,000, penulisan dalam laporan adalah <0,001.
14. Nilai odds rasio dan risiko relatif disajikan dua angka di belakang koma.
15. Dalam tabel harus ada informasi jumlah subjek.
16. Tabel harus menjelaskan dirinya (*self explanation).* Bila perlu, berikan keterangan dibawah tabel atau pada judul tabel. (Dahlan, 2014).

# DAFTAR PUSTAKA


Jones, David S. 2010. Statistik Farmasi. Alih Bahasa Hesty Utami Ramadaniati & H. Harrizul Rivai; editor Nurul Aini. EGC : Jakarta.

Sugiyarto. 2015. Dasar-dasar Statistik Farmasi Cetakan Pertama. Binafsi Publisher : Jakarta.

Dahlan, M.S. 2014. Statistik untuk Kedokteran dan Kesehatan Deskriptif, Bivariat, dan Multivariat dilengkapi Aplikasi Menggunakan SPSS Edisi 6. Epidemiologi Indonesia : jakarta

Ningrum, H., Ariyani, H., & Muthaharah, M. 2020. Studi Literatur Pola Penggunaan Obat Off-Label Pada Pasien Obstetri Dan Ginekologi. JCPS (Journal Of Current Pharmaceutical Sciences), 4(1), 273-281. Retrieved from https://journal. umbjm.ac.id/index.php/jcps/article/view/599

Ariyani, H., Akrom. R.Alfian. 2015. Impact of Pharmacist Mediated Brief oral counseling and remainder motivation via text messaging (SMS) on quality of life In ambulatory hypertensive Patients at Dr. H. Moch Ansari Saleh Banjar masin Hospital, South Kalimantan, Indonesia. International Conference of Medical and Health Sciences 2015
http://eprints.uad.ac.id/10503/1/DONE%20akrom_8_ICMHS_2015_BRIEFPHARMASICOUNSEL_ANSARI.pdf

Ariyani, H., Akrom, R. Alfian. 2017. Building care of hypertensive patients in reducing sodium intake in Banjarmasin. Book :Unity in Diversity and the Standardisation of Clinical Pharmacy Services 1st Edition Page 8. CRC Press/Balkemia Taylor & Francis. Proceedings Of The 17th Asian Conference On Clinical Pharmacy (ACCP 2017), 28–30 JULY 2017, YOGYAKARTA, INDONESIA. eBook ISBN9781315112756 https://www.taylorfrancis.com/chapters/edit/10.1201/9781315112756-6/building-care-hypertensive-patients-reducing-sodium-intake-banjarmasin-ariyani-akrom-alfian

Indriani, N., Ariyani, H., & Ulfah, M. 2021. Literature Study Of The Effectiveness Of Councelling On The Adherence Of HypertensivePatients In Various Health Facilities. JCPS (Journal Of Current Pharmaceutical Sciences), 4(2), 379-394. Retrieved from https://journal.umbjm.ac.id/index.php/jcps/article/view/725

Ariyani, H., Hartanto, D., & Lestari, A. 2018. Adherence Of Hypertensive Patients After Giving Pill Card In Hospital X Banjarmasin. JCPS (Journal Of Current Pharmaceutical Sciences), 1(2), 81-88. Retrieved from https://journal.umbjm.ac.id/index.php/jcps/article/view/136

Anonim. 2021. Penyajian Data
http://eprints.binadarma.ac.id/9237/1/PER%2003_Penyajian%20Data%20Statistik%20dan%20Statistika%282020-2021%29Genap_UNIVERSITAS%20BINA%20DARMA.pdf
diakses tanggal 20 Maret 2023

Ariyani, H., Putri N.Y., Fitriana R. 2022. Community Behavior in Consuming Immune-Boosting Vitamins During the Covid-19 Pandemic. Borneo Journal Of Pharmascientech ; Vol 06No 02, Oktober 2022ISSN : 2541 –3651, E-ISSN : 2548 – 3897.https://jurnalstikesborneolestari.ac.id/index.php/borneo/article/view/393

Otok, B.W. & Ratnaningsih, D.J., 2022, Konsep Dasar dalam Pengumpulan dan Penyajian Data
https://pustaka.ut.ac.id/lib/wp-content/uploads/pdfmk/SATS4213-M1.pdf diakses tanggal 20 Maret 2023

Ariyani, H. 2017. Gerakan Bucer "Ibu Cerdas" Melalui Metode Cara Belajar Insan Aktif (Cbia) Sebagai Sarana Mewujudkan Pemilihan Dan Penggunaan Obat Yang Rasionaldi Kelurahan Sungai Andai Banjarmasin, Kalimantan Selatan. UNES Journal of Community Service, 2(2), 105-112. Retrieved from https://ojs.ekasakti.org/index.php/UJCS/article/view/31

# BAB 7
# ANALISIS BIVARIAT, ANALISIS DATA KATEGORIKAL DENGAN KAI KUADRAT, PENYAJIAN DAN INTERPRETASI KAI KUADRAT

*Oleh Ryryn Suryaman Prana Putra*

## 7.1 Analisis Bivariat

Analisis bivariat merupakan salah satu jenis analisis yang digunakan sesuai dengan kondisi jumlah variabel. Analisis yang terkesan sederhana ini mampu menghasilkan pengujian yang sangat bermanfaat. Analisis bivariat merupakan analisis yang dilakukan untuk mengetahui hubungan antara 2 variabel. Dalam analisis ini, dua pengukuran dilakukan untuk masing-masing observasi. Dalam analisis bivariat, sampel yang digunakan bisa saja berpasangan atau masing-masing independen dengan perlakuan tersendiri. Secara umum, dalam analisis bivariat, variabel yang digunakan bisa saja berhubungan atau berdiri sendiri (independen). Saling berhubungan artinya sampel yang sama diberikan 2 pengukuran berbeda. Sedangkan, independen maksudnya adalah pengukuran dilakukan pada kedua kelompok sampel yang berbeda. (Yuvalianda, 2020).

Analisis bivariat adalah analisis statistik di mana dua variabel diamati. Di sini satu variabel tergantung dan yang lain independen. Variabel ini biasanya dilambangkan dengan X dan Y. Jadi, di sini kita menganalisis perubahan yang terjadi antara kedua variabel tersebut dan sejauh mana perubahannya. Analisis bivariat dinyatakan sebagai analisis hubungan konkuren antara dua

variabel atau atribut. Penelitian ini mengeksplorasi hubungan antara kedua variabel dan kedalaman hubungan ini untuk mengetahui apakah ada perbedaan antara kedua variabel dan alasan perbedaan tersebut. Ini adalah analisis statistik tunggal yang digunakan untuk menemukan hubungan yang ada antara dua set nilai. Variabel yang terlibat adalah X dan Y. Analisis data bivariat adalah analisis yang akurat dari dua variabel. (Anonim, 2022).

Analisis Bivariat adalah analisis secara simultan dari dua variabel. Hal ini biasanya dilakukan untuk melihat apakah satu variabel, seperti jenis kelamin, adalah terkait dengan variabel lain, mungkin sikap terhadap pria maupun wanita kesetaraan. Analisis bivariate terdiri atas metode-metode statistik inferensial yang digunakan untuk menganalisis data dua variabel penelitian. Penelitian terhadap dua variabel biasanya mempunyai tujuan untuk mendiskripsikan distribusi data, menguji perbedaan dan mengukur hubungan antara dua variabel yang diteliti.

Dalam Analisis Bivariat hipotesis yang diuji biasanya kelompok yang berbeda dalam ciri khas tertentu dengan koefisien kontigensi yang diberi simbol C. Analisis bivariat menggunakan tabel silang untuk menyoroti dan menganalisis perbedaan atau hubungan antara dua variabel. Menguji ada tidaknya perbedaan/ hubungan antara variabel kondisi pemukian, umur, agama, status migrasi, pendidikan, penghasilan, umur perkawinan pertama, status kerja dan kematian bayi/balita dengan persepsi nilai anak digunakan analisis *chi square*, dengan tingkat kemaknaan=0,05. Hasil yang diperoleh pada analisis *chi square*, dengan menggunakan program SPSS yaitu nilai p, kemudian dibandingkan dengan $\alpha$=0,05. Apabila nilai p< dari $\alpha$ =0,05 maka ada hubungan atau perbedaan antara dua variabel tersebut (Agung, 1993 dalam Novianda, 2013).

## 7.2 Manfaat Analisis Bivariat

Analisis bivariat adalah metode statistik yang penting karena memungkinkan peneliti melihat hubungan antara dua variabel dan menentukan hubungannya. Ini dapat membantu dalam berbagai jenis penelitian, seperti ilmu sosial, kedokteran, pemasaran, dan banyak lagi. Berikut adalah manfaat analisis bivariat (LP2M UMA, 2023):

a. Analisis bivariat membantu mengidentifikasi tren dan pola: Ini dapat mengungkapkan tren dan pola data tersembunyi dengan mengevaluasi hubungan antara dua variabel.
b. Analisis bivariat membantu mengidentifikasi hubungan sebab dan akibat: Ini dapat menilai apakah dua variabel terkait secara statistik, membantu peneliti dalam menetapkan variabel mana yang menyebabkan yang lain.
c. Membantu peneliti membuat prediksi: Ini memungkinkan peneliti memprediksi hasil di masa mendatang dengan memodelkan hubungan antara dua variabel.
d. Membantu menginformasikan pengambilan keputusan: Bisnis, kebijakan publik, dan pengambilan keputusan perawatan kesehatan dapat mengambil manfaat dari analisis bivariat.
e. Kemampuan untuk menganalisis korelasi antara dua variabel sangat penting untuk membuat penilaian yang baik, dan analisis ini sangat membantu tujuan penelitian.

## 7.3 Jenis Analisis Bivariat

Analisis bivariat memiliki beberapa jenis dan dapat digunakan untuk menentukan bagaimana hubungan dua variabel terkait (LP2M UMA, 2023).

a. Scatterplot. Scatterplot adalah grafik yang menunjukkan bagaimana dua variabel terkait satu sama lain. Ini menunjukkan nilai satu variabel pada sumbu x dan nilai variabel lain pada sumbu y. Pola tersebut menunjukkan

hubungan seperti apa yang ada di antara kedua variabel dan seberapa kuat hubungan tersebut.

b. Korelasi. Korelasi adalah ukuran statistik yang menunjukkan seberapa kuat dan ke arah mana dua variabel terkait. Korelasi positif berarti bahwa ketika satu variabel naik, begitu juga yang lain. Korelasi negatif menunjukkan bahwa ketika satu variabel naik, yang lain turun.
c. Regresi. Jenis analisis ini memberi Anda akses ke semua istilah untuk berbagai instrumen yang dapat digunakan untuk mengidentifikasi potensi hubungan antara poin data Anda. Persamaan untuk kurva atau garis tersebut juga dapat diberikan kepada Anda dengan menggunakan analisis regresi. Selain itu, ini mungkin menunjukkan kepada Anda koefisien korelasi.
d. Uji chi-kuadrat. Uji chi-square adalah metode statistik untuk mengidentifikasi perbedaan dalam satu atau lebih kategori antara apa yang diharapkan dan apa yang diamati. Premis utama tes ini adalah untuk menilai nilai data aktual untuk melihat apa yang diharapkan jika hipotesis nol itu valid. Peneliti menggunakan uji statistik ini untuk membandingkan variabel kategori dalam kelompok sampel yang sama. Ini juga membantu memvalidasi atau menawarkan konteks untuk penghitungan frekuensi.
e. Uji-T. Uji-t adalah uji statistik yang membandingkan rata-rata dua kelompok untuk melihat apakah mereka memiliki perbedaan yang besar. Analisis ini sesuai ketika membandingkan rata-rata dua kategori variabel kategori.
f. ANOVA (Analisis Varians). Tes ANOVA menentukan apakah rata-rata lebih dari dua kelompok berbeda satu sama lain secara statistik. Perbandingan rata-rata variabel numerik untuk lebih dari dua kategori variabel kategori sudah tepat.

Yuvalianda (2020) menjelaskan bahwa analisis bivariat dapat dibagi ke dalam 2 klasifikasi berikut :

a. Analisis Deskriptif. Pada analisis deskriptif, analisis bivariat bisa berlaku pada hampir seluruh visualisasi data. Jenis tampilan visualisasi seperti grafik batang, grafik garis, grafik column, dll masih bisa digunakan untuk analisis bivariat. Salah satu visualisasi data menarik yang biasa dilakukan dengan analisis bivariat adalah scatterplot. Scatterplot merupakan visualisasi data dalam bentuk titik-titik yang ditampilkan dalam sumbu x dan ya. Sumbu x dan y mewakili nilai dari masing-masing variabel. Dengan menggunakan scatterplot, kita bisa melihat pola hubungan antara 2 variabel. Hubungan yang terbentuk bisa jadi linier, eksponensial, seasonal, dll sesuai dengan kondisi data. Jangan lupa, scatterplot hanya alat bantu untuk mendeteksi pola hubungan, bukan untuk menarik kesimpulan terhadap pola hubungan antara 2 variabel.
b. Analisis Inferensial. Dengan menggunakan analisis inferensial, anda bisa mengambil kesimpulan yang valid dalam pengujian 2 variabel. Berbicara analisis inferensial, ada banyak jenis uji statistik yang bisa anda lakukan dengan 2 variabel. Berikut sedikit daftar jenis analisis uji yang bisa dilakukan:
    1) Uji McNemar. Uji McNEmar merupakan uji bivariat yang digunakan untuk menguji sebelum dan sesudah perlakuan (Pre-Test dan Post-Test) dimana setiap individu digunakan sebagai pengontrol dirinya sendiri. Uji ini dilakukan untuk pengukuran data nominal dan ordinal. Uji ini digunakan untuk menguji keefektifan suatu perlakukan tertentu terhadap kondisi sampel. Contohnya, uji ini digunakan untuk mengetahui efek perpindahan seseorang dari desa ke kota terhadap preferensi politik.

2) Uji Sign. Uji Sign digunakan untuk mengetahui ada tidaknya perbedaan dari data ordinal yang diperoleh dari sampel yang sama dan berpasangan. Yang perlu diingat dari uji Sign adalah uji ini hanya ampu mengetahui ada tidakknya perbedaan, bukan besar kecilnya perbedaan tersebut. Uji ini dilakukan dengan memberi tanda positf atau negatif dari perbedaan antar pasangan data. Uji Sign bisa digunakan untuk mengidentifikasi kecenderungan seseorang terhadap 2 buah brand poduk. Skala data yang digunakan pada uji ini adalah ordinal.
3) Uji Wilcoxon Matched Pairs. Uji Wilcoxon merupakan uji yang dilakukan untuk mengetahui apakah terdapat hubungan antara dua variabel atau tidak. Skala data yang digunakan pada uji ini adalah ordinal.
4) Uji-t Berpasangan. Uji-t Berpasangan merupakan uji dua variabel yang dilakukan untuk mengetahui apakah terdapat perbedaan rata-rata yang signifikan atau tidak. Contoh penggunaan uji-t berpasangan adalah pengujian apakah terdapat perbedaan rata-rata yang signifikan antara nilai matematika dan nilai kesenian siswa kelas A.
5) Uji Fisher Exact Probability. Uji Fisher Exact Probability merupakan pengujian yang dilakukan untuk mengetahui signfikanasi dari hipotesis komparatif pada dua sampel kecil independen. Uji ini digunakan bila kondisi data bersifat nominal dan ordinal. Dalam perhitungannya, data dalam pengujian ini dikelompokkan menjadi 2 kelompok independen. Misalnya, pria dan wanita, lalu kelompok miskin dan tidak miskin. Nantinya, perhitungan ini akan dikelompokkan dalam table kontingensi 2×2.

6) Uji Chi-Square 2 Sampel. Uji Chi-Square 2 sampel digunakan untuk mengetahui apakah terdapat hubungan antara 2 variabel atau tidak. Pada uji chi-square 2 sampel, skala data yang digunakan adalah skala nominal.
7) Uji Median. Uji ini digunakan untuk menguji hipotesis komparatif dari dua sampel independen. Pada pengujian ini, skala data yang digunakan adalah nominal dan ordinal. Pengujian ini didasarkan atas median sampel yang diambil secara acak. Skala data yang digunakan pada uji ini adalah nominal dan ordinal.
8) Uji Mann-Whitney U-Test. Uji Mann-Whitney U-Test digunakan untuk mengetahui signifikansi perbedaan dari dua populasi. Pada uji ini, skala data yang digunakan adalah ordinal. Contoh pengujian Man-Whitney U-Test adalah seorang guru ingin mengetahui apakah siswa dikelasnya memang memiliki bakat dalam pelajaran matematik atau lebih didominasi karena bantuan les.
9) Uji Kolmogorov Smirnov. Uji Kolmogorov Smirnov merupakan uji yang dilakukan untuk mengetahui apakah dua variabel memiliki distribusi yang sama atau tidak. Uji ini biasa digunakan untuk membuktikan apakah dua variabel yang digunakan berasal dari distribusi yang sama sebelum dilakukan analisis lanjutan. Skala data yang digunakan pada uji ini adalah interval dan rasio
10) Uji Wald-Waldovitz. Uji Wald-Waldovitz merupakan uji yang dilakukan apakah dua variabel yang digunakan berasal dari populasi yang sama atau tidak. Dalam uji ini, setidaknya data yang digunakan memiliki skala ordinal.
11) Uji t–test Independen. Uji-t independen merupakan uji yang dilakukan apakah 2 variabel yang berasal dari

kelompok yang berbeda memiliki rata-rata yang sama atau tidak. Dalam uji ini, skala data yang digunakan adalah interval dan rasio. Contohnya, seorang peneliti ingin membuktikan apakah nilai rata-rata ujian akhir sekolah favorit berbeda signifikan dengan sekolah non favorit.

12) Analisis Korelasi. Analisis korelasi merupakan analisis yang digunakan untuk mengetahui hubungan antara 2 variabel. Dengan analisis korelasi, kita bisa mengetahui apakah 2 variabel memiliki hubungan yang positif ataupun negatif. Penting untuk diingat bahwa korelasi hanyalah analisis yang menjelaskan seberapa kuat hubungan antara 2 variabel. Analisis korelasi tidak bisa dijadikan dasar untuk menyimpulkan hubungan sebab akibat (kausalitas) antara 2 variabel. Contoh penggunaan analisis korelasi adalah hubungan antara tinggi badan dan berat badan siswa.
13) Analisis Regresi Linier Sederhana. Analisis regresi linier sederhana merupakan analisis yang digunakan untuk mengetahui pengaruh dari suatu variabel terhadap variabel lain. Berbeda dengan analisis korelasi, analisis regresi linier sederhana bertujuan untuk menjelaskan hubungan sebab akibat (kausalitas) antara variavel independen dengan variaben dependen. Dengan analisis ini, kita bisa menyimpulkan seberapa jauh suatu variabel memengaruhi variabel lainnya.

## 7.4 Tahapan Analisis Bivariat

Analisis bivariat adalah analisis statistik yang dilakukan untuk menguji hipotesis antara dua variabel, untuk memperoleh jawaban apakah kedua variabel tersebut ada hubungan, berkorelasi, ada perbedaan, ada pengaruh dan sebagainya sesuai

dengan hipotesis yang telah dirumuskan. Adapun tahapan dalam analisis bivariat adalah (Nanda, 2020) :

a. Analisis proporsi atau presentase dengan membandingkan distribusi silang antara dua variabel yang bersangkutan.
b. Hasil analisis dari uji statistik (*chi square test, Z test, t test, Pearson*, dan sebagainya) dapat disimpulkan ada / tidaknya hubungan, korelasi, perbedaan antara kedua variabel tersebut. Bisa saja terjadi secara persentase berhubungan tetapi hasil uji statistik tidak bermakna.
c. Analisis keeratan hubungan antara kedua variabel tersebut dengan melihat Odd Ratio (OR). Besar kecilnya nilai OR menunjukan seberapa erat hubungan kedua variabel, demikian juga rentang OR dibawah angka 1 = faktor protektif dan > 1 = sebagai faktor risiko.

## 7.5 Contoh Analisis Bivariat

Analisis bivariat adalah analisis hubungan konkuren antara dua variabel atau atribut. Penelitian ini akan mengeksplorasi hubungan yang ada antara kedua variabel dan kedalaman hubungan tersebut. Ini membantu untuk mengetahui apakah ada perbedaan antara variabel dan apa alasan perbedaannya. Contoh analisis bivariat yang digunakan adalah mempelajari hubungan antara dua variabel. Mari kita lihat contoh yang mempelajari hubungan antara tekanan darah sistolik dan usia. Di sini, Anda mengambil sampel kelompok usia tertentu. Misalkan Anda mengambil sampel 10 pekerja (Anonim, 2022).

Kolom pertama berisi usia pekerja, dan kolom kedua mencatat tekanan darah sistolik mereka. Tabel ini kemudian perlu ditampilkan dalam format grafik untuk menarik beberapa kesimpulan darinya. Data bivariat biasanya ditampilkan melalui plot pencar. Di sini grafik digambar pada sumbu y kertas kisi relatif terhadap sumbu x, yang membantu untuk memahami hubungan antara kumpulan data yang diberikan. Plot pencar membantu

membangun hubungan antara variabel dan mencoba menjelaskan hubungan antara keduanya. Setelah menerapkan usia ke sumbu y dan tekanan darah sistolik ke sumbu x, Anda akan melihat bahwa kemungkinan ada hubungan linier antara keduanya (Anonim, 2022).

Grafik akan menunjukkan bahwa ada hubungan yang kuat dan positif antara usia dan tekanan darah. Hal ini dikarenakan grafik tersebut memiliki korelasi yang positif. Jadi semakin tua seseorang, semakin tinggi tekanan darah sistoliknya. Garis yang paling sesuai juga membantu untuk memahami kekuatan korelasi. Jika jarak antar titik kecil, korelasinya kuat. Koefisien korelasi atau R adalah angka antara -1 dan 1. Hal ini menunjukkan kekuatan hubungan linier antara kedua variabel. Untuk menggambarkan regresi linier, koefisien ini disebut koefisien korelasi Pearson. Ketika koefisien korelasi mendekati 1, ini menyoroti korelasi positif yang kuat. Ketika koefisien korelasi mendekati -1, ini menunjukkan bahwa ada korelasi negatif yang kuat. Jika koefisien korelasi sama dengan 0, berarti tidak ada hubungan sama sekali. (Anonim, 2022).

Contoh analisis bivariat lainnya seperti sebuah penelitian ingin menyelidiki hubungan antara pendidikan dan pendapatan. Dalam hal ini, salah satu variabelnya adalah tingkat pendidikan (misalnya SMA, perguruan tinggi, sekolah pascasarjana), dan variabel lainnya adalah pendapatan. Analisis bivariat dapat digunakan untuk menentukan apakah ada hubungan yang signifikan antara kedua variabel ini dan, jika ya, seberapa kuat dan ke arah mana hubungan tersebut (LP2M UMA, 2023).

Penelitian lain dengan tujuan ingin menyelidiki hubungan antara penuaan dan tekanan darah. Di sini, usia adalah satu variabel, dan tekanan darah adalah variabel lainnya (sistolik dan diastolik). Dimungkinkan untuk melakukan analisis analisis bivariat untuk menentukan apakah dan seberapa kuat kedua faktor ini terkait dengan pengujian signifikansi statistik. Ini hanya beberapa cara analisis ini dapat digunakan untuk menentukan

bagaimana dua variabel terkait. Jenis data dan pertanyaan penelitian akan menentukan teknik dan uji statistik mana yang digunakan dalam analisis (LP2M UMA, 2023).

## 7.6 Analisis Data Kategorikal dengan Kai Kuadrat/ Chi Square

Chi Square disebut juga dengan Kai Kuadrat. Chi Square adalah salah satu jenis uji komparatif non parametris yang dilakukan pada dua variabel, di mana skala data kedua variabel adalah nominal (Hidayat, 2012). Pengujian dengan menggunakan Chi-Square diterapkan pada kasus dimana akan diuji apakah frekuensi data yang diamati (frekuensi/data observasi) sama atau tidak dengan frekuensi harapan atau frekuensi secara teoritis. Chi Square adalah salah satu jenis uji komparatif yang dilakukan pada dua variabel, di mana skala data kedua variabel adalah nominal. Dalam pengujian Chi square, hal yang dapat diuji antara lain adalah uji independensi, uji Goodness of Fit, uji homogenitas, dan uji varians (PAMU Universitas Esa Unggul, 2019).

a. Uji independensi adalah uji untuk menentukan apakah antara variabel independen dan variabel dependennya terdapat perbedaan (hubungan) yang nyata atau tidak. Misalnya, kita ingin mengamati apakah terdapat perbedaan yang nyata antara pendidikan dengan pekerjaan, maka uji yang tepat dilakukan adalah uji independensi dengan Chi Square (PAMU Universitas Esa Unggul, 2019).

   Ada beberapa syarat yang harus dipenuhi jika akan melakukan pengujian dengan Chi Square.

   1) Tidak ada cell dengan nilai frekuensi kenyataan atau disebut juga Actual Count (Fo) sebesar 0 (Nol).
   2) Apabila bentuk tabel kontingensi 2x2, maka tidak boleh ada 1 cell saja yang memiliki frekuensi harapan atau disebut juga expected count (“Fn”) kurang dari 5.

3) Apabila bentuk tabel lebih dari 2x2, misal 2x3, maka jumlah cell dengan frekuensi harapan yang kurang dari 5 tidak boleh lebih dari 20%.

Ada beberapa rumus yang digunakan untuk menyelesaikan suatu pengujian Chi Square seperti rumus koreksi yates, Fisher Exact Test, dan Pearson Chi Square. Berikut rincian penggunaan rumus-rumusnya.

1) Jika tabel kontingensi berbentuk 2x2, maka rumus yang digunakan adalah "koreksi yates".
2) Apabila tabel kontingensi 2x2, tetapi cell dengan frekuensi harapan kurang dari 5, maka rumus harus diganti dengan rumus "Fisher Exact Test".
3) Rumus untuk tabel kontingensi lebih dari 2x2, rumus yang digunakan adalah "Pearson Chi-Square".

Prosedur uji Chi Square sebagai berikut :

1) Perumusan Hipotesis
   Ho : x = 0
   Ha : x ≠ 0
2) Menetapkan taraf nyata. Menentukan nilai kesalahan = α. Setelah ditetapkan selanjutnya menghitung nilai dari $x^2_{tabel}$ dengan menggunakan tabel $x^2$ dengan db= (r-1)(c-1) dimana db adalah derajat kebebasan, r adalah jumlah baris, dan c adalah jumlah kolom.
3) Menghitung nilai $x^2_{hitung}$. Rumus yang digunakan untuk menghitung nilai dari $x^2_{tabel}$ adalah sebagai berikut.

$$\chi^2_{hitung} = \sum \frac{(f_o - f_e)^2}{f_e}$$

dimana:
fo = nilai frekuensi observasi
fe = nilai frekuensi harapan

4) Penarikan keputusan dan kesimpulan. Kriteria keputusan dari pengujian Chi square adalah :
   Jika $x^2_{hitung} \leq x^2_{tabel}$, maka Ho diterima.
   Jika $x^2_{hitung} > x^2_{tabel}$, maka Ho ditolak.

b. Uji Goodness of Fit (kecocokan) adalah uji untuk menentukan apakah sebuah populasi mengikuti distribusi tertentu atau tidak. Misalnya, kita ingin mengetahui apakah populasi yang diamati berrdistribusi normal atau tidak, atau mungkin populasi yang diamati ternyata berdistribusi poisson, dan seterusnya. Sehingga uji yang digunakan adalah uji Goodness of Fit dengan Chi square. Chi-Square Goodness of Fit dapat digunakan ketika bertemu dengan kondisi sebagai berikut:  Metode sample yang digunakan adalah simple random sampling  Variabel yang digunakan adalah kategorikal  Nilai yang diharapkan pada sampel yang diobservasi minimal 5 dalam setiap level variabel (PAMU Universitas Esa Unggul, 2019).
   Prosedur pengujian Goodness of Fit adalah :
   1) Perumusan Hipotesis
      Ho : x = 0, data berdistribusi tertentu.
      Ha : x ≠ 0, data tidak berdistribusi tertentu.
   2) Menetapkan taraf nyata. Menentukan nilai kesalahan = α. Setelah ditetapkan selanjutnya menghitung nilai dari $x^2_{tabel}$ dengan menggunakan tabel $x^2$ dengan db= r-1 dimana db adalah derajat kebebasan dan r adalah jumlah baris.
   3) Menghitung nilai $x^2_{hitung}$. Rumus yang digunakan untuk menghitung nilai dari $x^2_{tabel}$ adalah sebagai berikut.

$$\chi^2_{hitung} = \sum \frac{(f_o - f_e)^2}{f_e}$$

      dimana:
      fo = nilai frekuensi observasi

fe = nilai frekuensi harapan

4) Penarikan keputusan dan kesimpulan. Kriteria keputusan dari pengujian Chi square adalah :
   Jika $x^2_{hitung} \leq x^2_{tabel}$, maka Ho diterima.
   Jika $x^2_{hitung} > x^2_{tabel}$, maka Ho ditolak.

Dasar dari uji kai kuadrat adalah membandingkan frekuensi yang diamati dengan frekuensi yang diharapkan. Misalnya kalau sebuah uang logam dilambungkan seratus kali permukaan uang tersebut ada dua yaitu M dan B, setelah pelambungan seratus kali kita amati yang keluar permukaan B sebanyak 60 kali. Kalau uang logam tersebut seimbang tentu permukaan B diharapkan keluar adalah 50 kali, maka sebetulnya disini kita mleihat perbedaan antara frekuensi yang diamati (Observed = 0) adalah 60 kali dan yang diharapkan (Expected = E) yakni 50 kali. Jika ada perbedaa antara pengamatan dengan yang diharapkan (O-E), apakah perbedaan itu cukup berarti (bermakna) atau hanya karena faktor variasi sample saja (Purnawinadi, 2016).

Langkah-langkah melakukan Kai Kuadrat/ Chi Square pada aplikasi SPSS adalah sebagai berikut (Purnawinadi, 2016) :

a. Pastikan anda berada pada data editor (SPSS)
b. Dari menu SPSS, klik Analyze, kemudian pilih Descriptive Statistic, lalu pilih Crosstab.
c. Dari menu tersebut, pada kotak Row (s) diisi variabel independen, sedangkan pada kotak Column (s) diisi variabel dependennya.
d. Klik Option Statistics, klik pilihan Chi Square dan klik pilihan Risk.
e. Klik Continue.
f. Klik Option Cells, bawa bagian percentages dan klik Row.
g. Klik Continue.

## 7.7 Penyajian dan Interpretasi Kai Kuadrat/ Chi Square

Penyajian dan interpretasi Kai Kuadrat dapat dilihat pada penelitian untuk melihat hubungan konsumsi Tablet Fe dengan kejadian perdarahan post partum berikut (Nanda, 2020).

**Tabel 7.1.** Hubungan Konsumsi Tablet Fe dengan Kejadian Perdarahan Post Partum

| Konsumsi Tablet Fe | Perdarahan Post Partum | | Total | Nilai P | OR (IK 95%) |
|---|---|---|---|---|---|
| | Ya | Tidak | | | |
| Ya | 7 (20%) | 28 (80%) | 35 (100%) | 0,004 | 3,08 (1,2-6,7_ |
| Tidak | 24 (54%) | 20 (46%) | 44 (100%) | | |
| Total | 31 (39,2%) | 48 (60,8%) | 79 (100%) | | |

Tabel di atas menunjukan bahwa dari 35 responden yang mengkonsumsi Tablet Fe, terdapat 7 orang (20%) yang mengalami Perdarahan Post Partum. Sedangkan dari 44 responden yang tidak mengkonsumsi Tablet Fe, lebih banyak yang mengalami Perdarahan Post Partum yakni sebanyak 24 oran (54%). Hasil uji statistik menunjukan nilai P 0,004 (< 0,05) yang berarti ada hubungan yang bermakna mengkonsumsi Tablet FE dengan Perdarahan Post Partum. Nilai OR (CI 95%) = 3,08 (1,2-6,7) artinya responden yang tidak mengkonsumsi Tablet Fe selama hamil beresiko mengalami Perdarahan Post Partum 3,08 kali lebih besar dibandingkan dengan responden yang mengkonsumsi Tablet Fe (Nanda, 2020).

Meskipun secara statistik ditemukan ada hubungan secara bermakna antara kedua variabel, tidak menjamin kemungkinan bermakna pula secara klinis. Seperti diketahui bahwa semakin

besar sampel yang dianalisis akan semakin besar menghasilkan kemungkinan berbeda/ berhubungan secara bermakna. Dengan sampel besar perbedaan-perbedaan sangat kecil, yang sedikit atau bahkan tidak mempunyai manfaat secara substansi/ klinis dapat berubah menjadi bermakna secara statistik. Dengan demikian peneliti yang melakukan analisis hendaknya jangan hanya melihat dari sudut pandang statistik saja, tetapi harus juga melihat dari segi kegunaan atau manfaat dari sisi klinis juga (Nanda, 2020).

Untuk melihat hubungan antara variable independent dengan dependen maka digunakan Tabel Chi Square Tests untuk melihat hasil yang diharapkan. Misalnya pada contoh di bawah ini pada penelitian untuk melihat hubungan pekerjaan dengan perilaku menyusui. Variabel pekerjaan berisi dua nilai yaitu bekerja dan tidak bekerja, dan variabel menyusui berisi dua nilai yaitu eksklusif dan non eksklusif (Purnawinadi, 2016).

**Chi-Square Tests**

| | Value | df | Asymp. Sig. (2-sided) | Exact Sig. (2-sided) | Exact Sig. (1-sided) |
|---|---|---|---|---|---|
| Pearson Chi-Square | 8.013[a] | 1 | .005 | | |
| Continuity Correction[b] | 6.490 | 1 | .011 | | |
| Likelihood Ratio | 8.244 | 1 | .004 | | |
| Fisher's Exact Test | | | | .010 | .005 |
| Linear-by-Linear Association | 7.853 | 1 | .005 | | |
| N of Valid Cases | 50 | | | | |

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 12.00.

b. Computed only for a 2x2 table

Hasil uji Chi Square dapat dilihat pada tabel Chi Square Tests yang disajikan (Purnawinadi, 2016):

a. Bila pada table 2x2 dijumpai nilai Expected (harapan) kurang dari lima, maka yang digunakan adalah 'Fisher Exact Test.
b. Bila pada tabel 2x2, dan tidak ada nilai E < 5, maka uji yang dipakai sebaiknya 'Continuity Correction.

c. Bila tabelnya lebih dari 2x2, misalnya 3x2, 3x3 dan sebagainya, maka yang digunakan adalah uji 'Pearson Chi Square.
d. Uji 'Likelihood Ratio dan 'Linear-by-Linear Association, biasanya digunakan untuk keperluan lebih spesifik, misalnya analisis stratifikasi pada bidang epidemiologi dan juga untuk mengetahui hubungan linier dua variabel kategorik, sehingga kedua jenis ini jarang digunakan.

Untuk mengetahui adanya nilai E kurang dari 5, dapat dilihat pada foot note di bagian bawah tabel Chi Square Test, dan tertulis nilainya 0 cell (0%) berarti pada tabel silang di aatas tidak ditemukan ada nilai E < 5. Dengan demikian kita menggunakan uji '*Continuity Correction* dengan p value = 0,011 < alpha (0,05), berarti kesimpulannya ada perbedaan yang signifikan antara perilaku menyusui eksklusif antara ibu yang bekerja dengan ibu yang tidak bekerja. Sedangkan untuk mengetahui besar/kekuatan hubungan dua variabel banyak metodenya, tergantung pada latar belakang disiplin ilmu, dan untuk bidang ilmu kesehatan digunakan nilai OR (*Odds Ratio*) untuk penelitian Cross Sectional dan Case Control, dan RR (*Relative Risk*) untuk penelitian Cohort (Purnawinadi, 2016).

**Risk Estimate**

| | Value | 95% Confidence Interval | |
|---|---|---|---|
| | | Lower | Upper |
| Odds Ratio for Status Pekerjaan Ibu (bekerja / tidak kerja) | 5.464 | 1.627 | 18.357 |
| For cohort Status Menyusui Eksklusif = tidak eksklusif | 2.429 | 1.226 | 4.811 |
| For cohort Status Menyusui Eksklusif = eksklusif | .444 | .239 | .827 |
| N of Valid Cases | 50 | | |

Pada tabel di atas nilai OR yaitu 5,464 (95% CI: 1,627 – 18,357). Sedangkan nilai RR dapat dilihat pada baris For Cohort. Data ini merupakan penelitian Cross Sectional, maka kita dapat menginterpretasikan nilai OR = 5,464 bahwa ibu yang tidak bekerja mempunyai peluang 5,46 kali untuk menyusui secara eksklusif dibandingkan dengan ibu yang bekerja (Purnawinadi, 2016).

## 7.8 Keterbatasan Uji Kai Kuadrat/ Chi Square

Teknik uji kai kuadrat adalah memakai data yang diskrit dengan pendekatan distribusi kontinu. Dekatnya pendekatan yang dihasilkan tergantung pada ukuran pada berbagai sel dari tabel kontingensi. Untuk pendekatan yang memadai digunakan aturan dasar bahwa frekuensi harapan tidak boleh terlalu kecil, secara umum ada ketentuan (Kasim dalam Jasaputra & Santosa, 2008) :

1. Tidak boleh ada sel yang mempunyai nilai harapan kecil dari 1 (satu).
2. Tidak boleh lebih dari 20 % sel mempunyai nilai harapan kecil dari 5 (lima).

Hal ini perlu diperhaikan mengingat luasnya pemakaian uji kai kuadrat dan mudahnya melakukan uji. Karena hal ini ditemui didalam suatu tabel kontingensi, teknik yang dianggap dapat menanggulangi permasalahan adalah menggabungkan nilai dari sel yang kecil tadi kepada lainnya (mengcollaps) sel. Artinya kategori dari variabel dikurangi sehingga kategori yang nilai harapannya kecil dapat digabung ke kategori lain. Untuk tabel 2 x 2 hal ini tidak dapat dilakukan, maka solusinya adalah melakukan uji "Fisher Exact" (Kasim dalam Jasaputra & Santosa, 2008).

# DAFTAR PUSTAKA


Anwar Hidayat. 2012. Tutorial Rumus Chi Square Dan Metode Hitung. https://www.statistikian.com/2012/11/rumus-chi-square.html

Anonim. 2022. Pengertian Dan Contoh Analisis Bivariat. https://kolonginfo.com/pengertian-dan-contoh-analisis-bivariat/

Diana Krisanti Jasaputra dan Slamet Santosa. 2008. Metodologi Penelitian Medis Edisi 2. Bandung : PT Danamartha Sejahtera Utama

I Gede Purnawinadi. 2016. Analisis Bivariat Data Kategorik Dan Kategorik Uji Kai Kuadrat (Chi-Square). Universitas Klabat

LP2M Universitas Medan Area. 2023. Mengenal Analisis Bivariat: Pengertian, Jenis dan Contoh. https://lp2m.uma.ac.id/2023/02/28/mengenal-analisis-bivariat-pengertian-jenis-dan-contoh/

Nanda. 2020. Bahan Ajar Metodolgi Riset. https://bahan-ajar.esaunggul.ac.id/mik691/wp-content/uploads/sites/1264/2019/11/Metodologi-Riset-Pertemuan-8.pdf

PAMU Universitas Esa Unggul. 2019. Uji Chi Square Modul Perkuliahan. Jakarta Barat : Pelaksana Akademik Mata Kuliah Umum (PAMU) Universitas Esa Unggul

Reza Novianda. 2013. Analisis Bivariat. https://www.scribd.com/doc/135381487/analisis-bivariat#

Yuvalianda. 2020. Analisis Bivariat: Pengertian Hingga Contoh Lengkap. https://yuvalianda.com/analisis-bivariat/

# BAB 8
# ANALISIS BIVARIAT – T-TEST

*Oleh Putri Rahmadani*

## 8.1 Pendahuluan

Pemilihan uji statistik ditentukan berdasarkan jenis data, apakah data berjenis data numerik atau kategorik. Jika data berbentuk numerik dan kategorik maka analisis yang digunakan adalah uji T Test. Analisis bivariat T-test digunakan untuk mengetahui perbedaan rata-rata antara dua kelompok (Budiarto, 2004). Contohnya:

"Apakah ada perbedaan tekanan darah ibu hamil yang bekerja dengan ibu hamil yang bekerja" atau;

"Apakah ada perbedaan pengetahuan ibu hamil sebelum dan sesudah diberikan penyuluhan"

Uji T-test sering juga disebut sebagai uji beda dua mean. Dua kelompok dari uji t-test apakah merupakan dua kelompok independen/saling bebas atau dua kelompok yang dependen/berpasangan harus diperhatikan terlebih dahulu. Variabel yang dihubungkan dalam uji t-test adalah numerik dan kategorik (hanya dua kelompok).

Uji beda 2 mean dibagi menjadi dua kelompok, yaitu uji T Independen dan uji T dependen.

## 8.2 Uji Normalitas Data

Uji normalitas data dilakukan sebelum analisis data. Uji normalitas dilakukan jika data kita berbentuk data numerik. Oleh karena itu, sebelum melakukan uji T, uji normalitas harus dilakukan terlebih dahulu.

Uji normalitas di SPSS, ditentukan uji Kolmogorov-Smirnov dan uji Shapiro Wilk. Penggunaan dua uji ini ditentukan oleh jumlah sampel penelitian, jika sampel berjumlah <50 sampel maka analisis uji kenormalan data menggunakan uji Shapiro Wilk. Namun, jika sampel berjumlah ≥50, maka uji kenormalan data menggunakan Kolmogorov-Smirnov.

Data dikatakan berdistribusi normal, jika memiliki *p-value* > α (0.05). Dan sebaliknya, jika data memiliki p-value ≤ α (0.05), maka data tidak terdistribusi normal. Uji kenormalan data menentukan analisis yang akan digunakan. Data dengan distribusi normal menggunakan uji parametrik dan data dengan tidak distribusi normal menggunakan uji non parametrik.

## 8.3 Uji T Independen (Uji Beda Dua Mean Independen)

Uji T Independen atau uji beda dua mean independen digunakan untuk mengetahui perbedaan rata-rata antara dua kelompok data yang berbeda (Hastono, 2016). Misal:

1. Untuk mengetahui perbedaan rata-rata tekanan darah ibu hamil yang merokok dan ibu hamil yang tidak merokok.
2. Untuk mengetahui perbedaan rata-rata berat badan bayi dari ibu yang hipertensi dan ibu yang tidak hipertensi.

Adapun syarat yang harus dipenuhi, antara lain:

1. Data harus memiliki distribusi normal/simetris
2. Memiliki dua kelompok yang berbeda (independen)

Jika asumsi tidak terpenuhi, maka dilakukan uji non parametrik yaitu **Mann Whitney Test**.

### 8.3.1 Uji Homogenitas Varians

Uji homogenitas varian bertujuan untuk melihat perbedaan variasi antara dua kelompok data. Uji varian pada uji beda dua mean terdiri dari 2, yaitu varian sama dan varian beda. Sehingga, harus diperhatikan dahulu apakah dua kelompok yang diuji memiliki varian sama atau varian berbeda. Uji yang digunakan untuk mengetahui varians kedua kelompok adalah *Levenest Test* (Riyanto, 2017).

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

Ket:
$df_1 = n_1 - 1$
$df_2 = n_2 - 1$

Pada uji F, varian yang lebih besar berfungsi sebagai pembilang dan varian yang lebih kecil berfungsi sebagai penyebut.
Penentuan:
Jika *p-value levenest test* > α (0.05), maka varian sama
Jika *p-value levenest test* ≤ α (0.05), maka varian berbeda

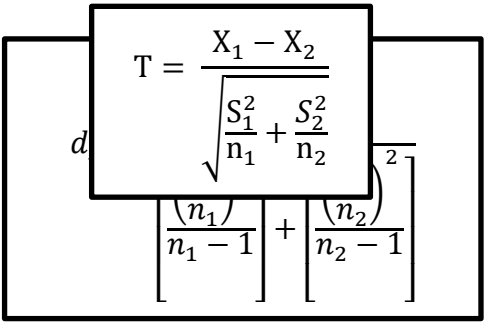
### 8.3.2 Uji T Independen Varian Sama

$$T = \frac{X_1 - X_2}{S_p \sqrt{\frac{1}{n} + \frac{1}{n}}}$$

$$S_p = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

**$df = n_1 + n_2 - 2$**
Keterangan:
$X_1, X_2$ = rata-rata kelompok 1 dan 2
$S_1, S_2$ = standar deviasi kelompok 1 dan 2

### 8.3.3 Uji T Independen Varian Beda



$$T = \frac{X_1 - X_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$$d = \frac{\left(\frac{S_1^2}{n_1}\right)^2}{\left[\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1}\right] + \left[\frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}\right]}$$

## 8.4 Uji T Dependen (*Paired Sample*)

Uji T Dependen atau uji beda dua mean dependen digunakan untuk mengetahui perbedaan rata-rata antara kelompok sebelum dan kelompok setelah diberikan intervensi (Hastono, 2016). Misal:

1. Untuk mengetahui perbedaan rata-rata tekanan darah ibu hamil sebelum dan setelah diberikan obat hipertensi.
2. Untuk mengetahui perbedaan rata-rata pengetahuan ibu sebelum dan setelah diberikan penyuluhan.

Adapun syarat yang harus dipenuhi, antara lain:

1. Data harus memiliki distribusi normal/simetris
2. Memiliki dua kelompok yang sama (dependen)

Jika asumsi tidak terpenuhi, maka dilakukan uji non parametrik yaitu **Wilcoxon Sign Test.**

$$T = \frac{d}{\frac{S}{\sqrt{n}}}$$

df = n – 1

Keterangan:
d = rata-rata deviasi *atau* selisih sampel 1 dengan 2
S = standar deviasi dari nilai d

## 8.5 Studi Kasus

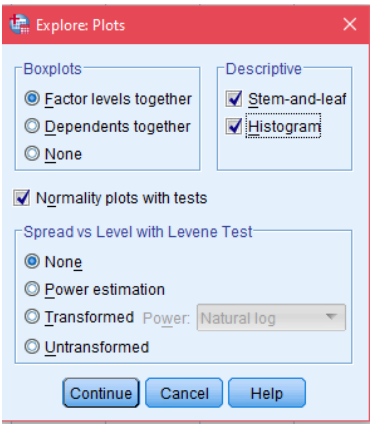
### 8.5.1 Uji T Independen

Studi kasus yang akan dilakukan analisis :
"untuk mengetahui perbedaan rata-rata tekanan darah ibu hamil yang bekerja dengan ibu yang tidak bekerja"

Sebelumnya, lakukan uji normalitas terlebih dahulu, langkah-langkahnya sebagai berikut :

a. Pada menu utama SPSS, pilih '*analyze',* kemudian pilih '*descriptive statistics',* dan pilih '*explore*'.
b. Muncul jendela '*explore'*, masukkan variabel numerik yaitu tekanan darah ke '*dependent list'* dan variabel kategorik yaitu pekerjaan ke '*factor list'.*


c. Pada menu '*plots',* centang '*histogram*' dan '*normality plots with tests',* continue dan OK.




**Hasil uji kenormalan data**

**Tests of Normality**

| | status pekerjaan ibu hamil | Kolmogorov-Smirnov[a] | | | Shapiro-Wilk | | |
|---|---|---|---|---|---|---|---|
| | | Statistic | df | Sig. | Statistic | df | Sig. |
| tekanan darah ibu hamil | Tidak Bekerja | .194 | 10 | .200* | .898 | 10 | .207 |
| | Bekerja | .220 | 10 | .188 | .969 | 10 | .882 |

*. This is a lower bound of the true significance.

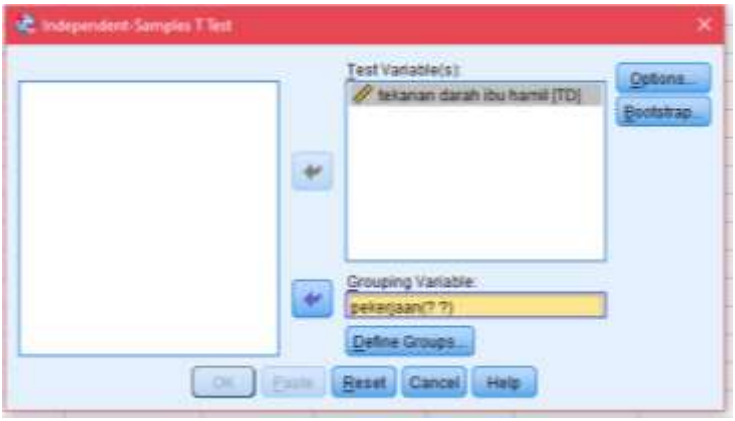
a. Lilliefors Significance Correction

Berdasarkan hasil output, didapatkan hasil uji kenormalan data yaitu uji Kolmogorov-Smirnov dan uji Shapiro-Wilk. Karena jumlah sampel sebanyak 20, maka uji kenormalan data yang dipakai adalah uji 'Shapiro Wilk'. Dari hasil uji 'Shapiro-Wilk', didapatkan p-value 0.207 untuk yang ibu tidak bekerja dan p-value 0.882 untuk ibu

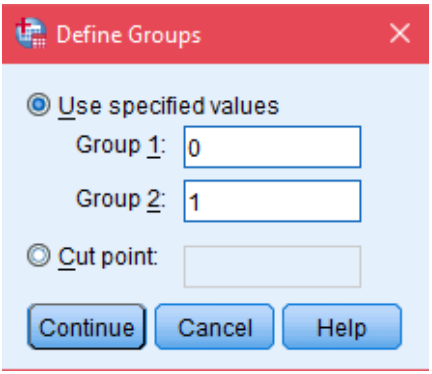
yang bekerja. Dapat disimpulkan bahwa data memiliki distribusi normal, karena memiliki p-value > α (α=0.05).

Setelah dilakukan uji normalitas data, kesimpulannya data terdistribusi normal. Oleh karena itu, kita menggunakan uji *Independent Sample T Test.* Langkah-langkah analisisnya :

d. Pada menu utama SPSS, pilih '*analyze*', kemudian pilih '*compare means*', dan pilih '*Independent Sample T Test*'.
e. Muncul jendela '*Independent Samples T Test*', kemudian masukkan variabel numerik ke kotak '*Test Variables*' dan variabel kategorik ke kotak '*Grouping Variable*'. Jadi masukkan variabel 'tekanan darah' ke '*Test Variables*' dan variabel 'pekerjaan' masukkan ke kotak '*Grouping Variable*'.




f. Pilih '*define grup*', kemudian isikan '*group 1*' dan '*group 2*'. Dalam studi kasus ini, kode '0' untuk ibu yang tidak bekerja dan kode '1' untuk ibu yang bekerja. Sehingga, masukkan '0' ke 'group 1' dan '1' ke 'group 2', kemudian 'continue' dan 'OK'




**Hasil analisis T Independen :**

**Group Statistics**

| | status pekerjaan ibu hamil | N | Mean | Std. Deviation | Std. Error Mean |
|---|---|---|---|---|---|
| tekanan darah ibu hamil | Tidak Bekerja | 10 | 129.80 | 7.857 | 2.485 |
| | Bekerja | 10 | 139.70 | 6.111 | 1.932 |

**Independent Samples Test**

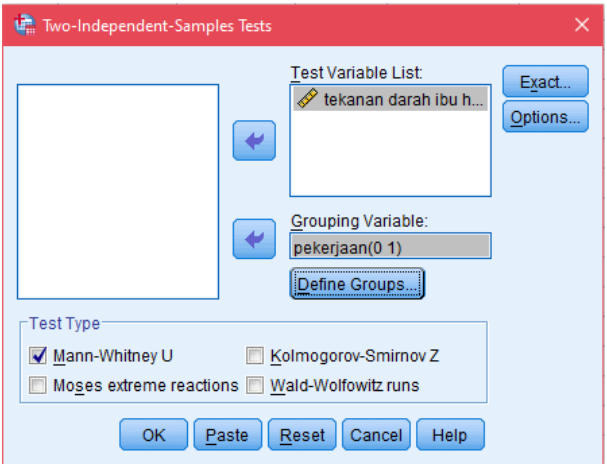
| | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | F | Sig. | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference | |
| | | | | | | | | | Lower | Upper |
| tekanan darah ibu hamil | Equal variances assumed | .712 | .410 | -3.145 | 18 | .006 | -9.900 | 3.148 | -16.513 | -3.287 |
| | Equal variances not assumed | | | -3.145 | 16.972 | .006 | -9.900 | 3.148 | -16.542 | -3.258 |

Berdasarkan tabel di atas, deskriptif data dapat dilihat pada tabel '*Group Statistics*' yaitu nilai rata-rata, standar deviasi, dan standar error tekanan darah ibu hamil pada masing-masing kategori.

Hasil uji T dapat dilihat pada tabel '*Independent Samples Test*'. Sebelum itu perhatikan dahulu hasil uji *Levene's Test* untuk mengetahui apakah varian kedua kelompok sama (*equal variances assumed*) atau varian kedua kelompok berbeda (*equal variances not assumed*). Lihat p-value *Levene's Test*, jika *p-value* > α (0.05), maka varian sama dan jika *p-value* ≤ α (0.05), maka varian beda. Pada tabel didapatkan nilai *Levene's Test*=0.410, berarti tidak ada perbedaan varian (varian sama) antara dua kelompok. Maka untuk melihat *p-value* hasil analisis uji T, dilihat pada *equal variances assumed* di kolom *Sig. (2 tailed)*, yaitu sebesar p=0.006, artinya terdapat perbedaan rata-rata tekanan darah ibu hamil yang bekerja dan ibu hamil yang tidak bekerja.

Jika kita mengasumsikan **data** diatas **tidak berdistribusi normal**, maka pilih uji non parametrik yaitu *Mann Whitney Test*. Langkah-langkahnya:

a. Pada menu utama SPSS, pilih '*analyze*', kemudian '*non parametric test*', '*legacy dialogs*', dan pilih '2 *Independent Samples*'.
b. Muncul jendela '*Two-Independent-Samples Test*', masukkan variabel 'tekanan darah' ke kotak '*Test Variables*' dan variabel 'pekerjaan ibu' ke kotak '*Grouping Variables*'. Kemudian pilih '*define group*', masukkan '0' ke 'group 1' dan '1' ke 'group 2'.
c. Pastikan test type yang tercentang adalah '*Mann Whitney U*', kemudian '*continue*' dan 'OK'.



**Hasil analisis Mann Whitney :**

**Ranks**

| | status pekerjaan ibu hamil | N | Mean Rank | Sum of Ranks |
|---|---|---|---|---|
| tekanan darah ibu hamil | Tidak Bekerja | 10 | 7.25 | 72.50 |
| | Bekerja | 10 | 13.75 | 137.50 |
| | Total | 20 | | |

**Test Statistics[a]**

| | tekanan darah ibu hamil |
|---|---|
| Mann-Whitney U | 17.500 |
| Wilcoxon W | 72.500 |
| Z | -2.475 |
| Asymp. Sig. (2-tailed) | .013 |
| Exact Sig. [2*(1-tailed Sig.)] | .011[b] |

a. Grouping Variable: status pekerjaan ibu hamil

b. Not corrected for ties.

Berdasarkan tabel di atas, deskriptif data dapat dilihat pada tabel '*Ranks*' yaitu nilai rata-rata tekanan darah ibu hamil pada masing-masing kategori.

Hasil uji T dapat dilihat pada tabel '*Test Statistics*' pada nilai '*Asymp. Sig (2 tailed)*'. Jika *p-value* ≤ α (0.05), maka Ho ditolak, dan jika *p-value* > α (0.05), maka Ho diterima. Pada tabel didapatkan hasil *p-value* 0.013 (p ≤ α), artinya terdapat perbedaan rata-rata tekanan darah antara ibu hamil yang bekerja dengan ibu hamil yang tidak bekerja.

**Penyajian dan Interpretasi Data**

Hasil analisis data di SPSS tidak boleh di*copy* dan disajikan langsung ke dalam penulisan laporan. Oleh karena itu, harus dibuat tabel baru untuk menampilkan hasil ouput analisis.

**Tabel 8.3.** Pengaruh Tekanan Darah Ibu Hamil Menurut Pekerjaan di Kota Y Tahun 2023

| Pekerjaan | Mean | SD | *p-value* |
|---|---|---|---|
| Tidak Bekerja | 129.80 | 7.857 | 0.006 |
| Bekerja | 139.70 | 6.111 | |

Berdasarkan Tabel 8.1 didapatkan bahwa rata-rata tekanan darah ibu hamil yang bekerja lebih yaitu 139.7 dengan variasi 6.11 mg, dibandingkan dengan ibu yang tidak bekerja yaitu rata-ratanya 129.8 mmHg dengan variasi 7.86 mmHg. Hasil analisis uji T didapatkan *p-value* 0.006, artinya terdapat perbedaan rata-rata tekanan darah ibu hamil yang bekerja dengan ibu hamil yang tidak bekerja.

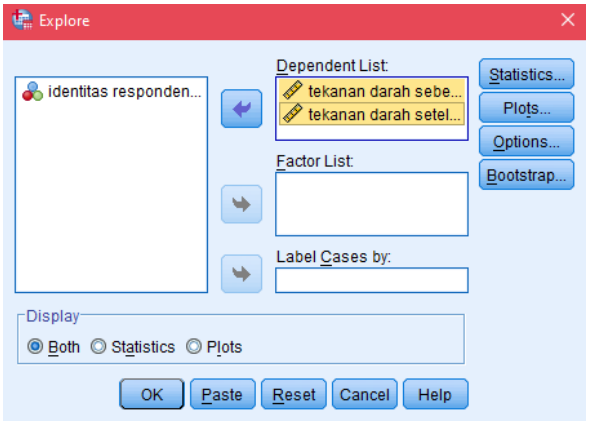
**8.4.2 Uji T Dependen**

Studi kasus yang akan dilakukan analisis :

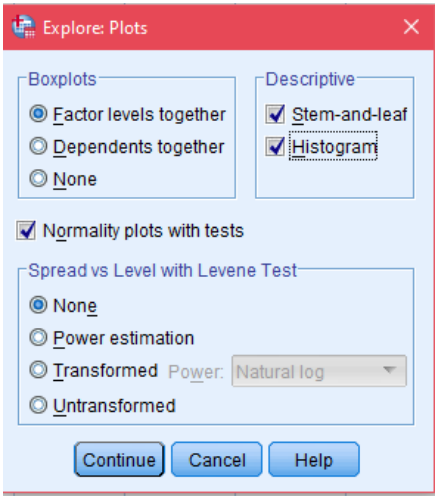
"untuk mengetahui perbedaan tekanan darah sebelum dan sesudah diberikan obat hipertensi"

Sebelumnya, lakukan uji normalitas terlebih dahulu, langkah-langkahnya sebagai berikut :

a. Pada menu utama SPSS, pilih '*analyze',* kemudian pilih '*descriptive statistics',* dan pilih '*explore*'.
b. Muncul jendela '*explore'*, masukkan variabel numerik yaitu tekanan darah sebelum dan sesudah ke '*dependent list'.*




c. Pada menu '*plots'*, centang '*histogram*' dan '*normality plots with tests'*, continue dan OK.



**Hasil uji kenormalan data**

**Tests of Normality**

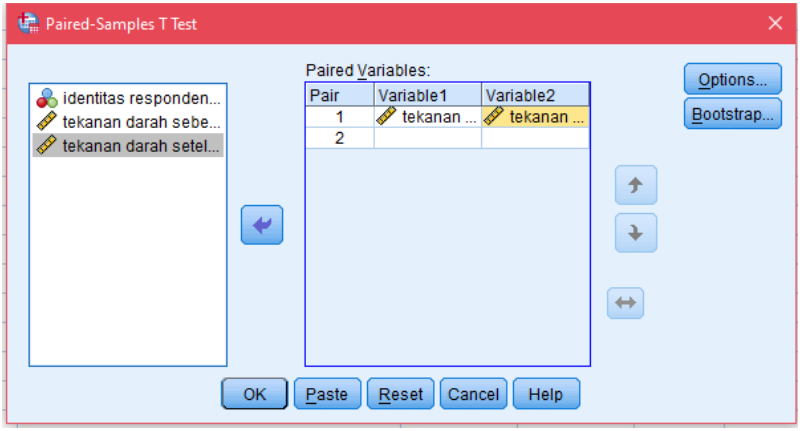
| | Kolmogorov-Smirnov[a] | | | Shapiro-Wilk | | |
|---|---|---|---|---|---|---|
| | Statistic | df | Sig. | Statistic | df | Sig. |
| tekanan darah sebelum intervensi | .149 | 15 | .200* | .941 | 15 | .397 |
| tekanan darah setelah intervensi | .236 | 15 | .024 | .914 | 15 | .154 |

*. This is a lower bound of the true significance.
a. Lilliefors Significance Correction

Berdasarkan hasil output, didapatkan hasil uji kenormalan data yaitu uji *Kolmogorov-Smirnov* dan uji *Shapiro-Wilk*. Karena jumlah sampel sebanyak 30, maka uji kenormalan data yang dipakai adalah uji '*Shapiro Wilk*'. Dari hasil uji '*Shapiro-Wilk*', didapatkan p-value tekanan darah sebelum intervensi 0.397 dan p-value tekanan darah setelah intervensi 0.154. Dapat disimpulkan bahwa data memiliki distribusi normal, karena memiliki p-value > α (α=0.05).

Setelah dilakukan uji normalitas data, kesimpulannya data terdistribusi normal. Oleh karena itu, kita menggunakan uji *Paired Sample T Test.* Langkah-langkah analisisnya :

a. Pada menu utama SPSS, pilih '*analyze*', kemudian pilih '*compare means*', dan pilih '*Paired Sample T Test*'.
b. Muncul jendela '*Paired-Samples T Test*', kemudian masukkan variabel 'tekanan darah sebelum' ke 'Variable 1' dan variabel 'tekanan darah setelah' ke 'Variable 2' dan 'OK'.




**Hasil analisis T Dependen :**

**Paired Samples Statistics**

| | | Mean | N | Std. Deviation | Std. Error Mean |
|---|---|---|---|---|---|
| Pair 1 | tekanan darah sebelum intervensi | 133.80 | 15 | 9.237 | 2.385 |
| | tekanan darah setelah intervensi | 126.53 | 15 | 9.935 | 2.565 |

Berdasarkan hasil output pada tabel '*Paired Samples Statistics*' didapatkan nilai rata-rata, standar deviasi, dan standar error mean masing-masing sebelum dan sesudah intervensi. Rata-rata tekanan darah pada sebelum intervensi adalah 133.8 mmHg dengan standar deviasi 9.24 mmHg. Rata-rata tekanan darah setelah intervensi adalah 126.5 mmHg dengan standar deviasi 9.94 mmHg.

**Paired Samples Correlations**

| | | N | Correlation | Sig. |
|---|---|---|---|---|
| Pair 1 | tekanan darah sebelum intervensi & tekanan darah setelah intervensi | 15 | .757 | .001 |

Berdasarkan hasil output pada tabel '*Paired Samples Correlations*' didapatkan nilai korelasi antara nilai sebelum dan sesudah intervensi.

**Paired Samples Test**

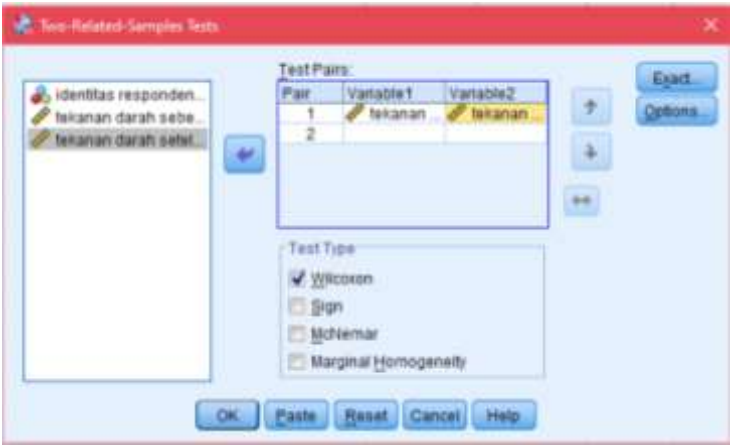
| | | Paired Differences | | | | | t | df | Sig. (2-tailed) |
|---|---|---|---|---|---|---|---|---|---|
| | | Mean | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference | | | | |
| | | | | | Lower | Upper | | | |
| Pair 1 | tekanan darah sebelum intervensi - tekanan darah setelah intervensi | 7.267 | 6.713 | 1.733 | 3.549 | 10.984 | 4.192 | 14 | .001 |

Berdasarkan hasil output pada tabel '*Paired Samples Test*' didapatkan hasil perbedaan rata-rata tekanan darah sebelum dan sesudah intervensi. Didapatkan perbedaan mean antara tekanan darah sebelum dan sesudah intervensi adalah 7.27

mmHg. Hasil analisis menunjukkan *p-value* 0.001, artinya ada perbedaan rata-rata tekanan darah sebelum dan sesudah diberikan intervensi.

Jika **data tidak berdistribusi normal**, maka pilih uji non parametrik yaitu *Wilcoxon Test*. Langkah-langkahnya:

a. Pada menu utama SPSS, pilih '*analyze'*, kemudian '*non parametric test*', '*legacy dialogs*', dan pilih '2 *Related Samples*'.
b. Muncul jendela '*Two-Related-Samples Test'*, masukkan variabel 'tekanan darah sebelum' ke kotak '*Variable 1'* dan variabel 'tekanan darah setelah' ke kotak '*Variable 2'*.
c. Pastikan test type yang tercentang adalah '*Wilcoxon*' dan 'OK'.



**Hasil analisis Wilcoxon:**

**Ranks**

| | | N | Mean Rank | Sum of Ranks |
|---|---|---|---|---|
| tekanan darah setelah intervensi - tekanan darah sebelum intervensi | Negative Ranks | 14[a] | 7.50 | 105.00 |
| | Positive Ranks | 0[b] | .00 | .00 |
| | Ties | 1[c] | | |
| | Total | 15 | | |

a. tekanan darah setelah intervensi < tekanan darah sebelum intervensi
b. tekanan darah setelah intervensi > tekanan darah sebelum intervensi
c. tekanan darah setelah intervensi = tekanan darah sebelum intervensi

Berdasarkan hasil output tabel '*Ranks*', didapatkan deskriptif perbedaan rata-rata tekanan darah sebelum dan sesudah diberikan obat hipertensi. '*Negative Ranks*' menunjukkan jika rata-rata tekanan setelah lebih rendah dibandingkan dengan rata-rata tekanan darah sebelum pemberian obat intervensi. Nilai '*Positive Ranks*' menunjukkan rata-rata tekanan darah setelah pemberian obat lebih tinggi dibandingkan dengan rata-rata tekanan darah sebelum pemberian obat. '*Ties*' menunjukkan rata-rata tekanan darah setelah pemberian obat sama dengan rata-rata tekanan darah sebelum pemberian obat.

**Test Statistics**[a]

| | tekanan darah setelah intervensi - tekanan darah sebelum intervensi |
|---|---|
| Z | -3.302[b] |
| Asymp. Sig. (2-tailed) | .001 |

a. Wilcoxon Signed Ranks Test
b. Based on positive ranks.

Berdasarkan hasil output tabel '*Test Statistics*', didapatkan hasil analisis wilcoxon yaitu pada '*Asymp Sig (2 tailed)*' = 0.001. Artinya ada perbedaan rata-rata tekanan darah ibu sebelum dan sesudah pemberian obat.

**Penyajian dan Interpretasi Data**

Hasil analisis data di SPSS tidak boleh di*copy* dan disajikan langsung ke dalam penulisan laporan. Oleh karena itu, harus dibuat tabel baru untuk menampilkan hasil ouput analisis.

**Tabel 8.4.** Pengaruh Pemberian Obat Hipertensi terhadap Tekanan Darah Ibu di Kota Y Tahun 2022

| **Pemberian Obat Hipertensi** | **Mean** | **SD** | ***p-value*** |
|---|---|---|---|
| Sebelum | 133.8 | 9.24 | 0.001 |
| Sesudah | 126.5 | 9.94 | |

Berdasarkan Tabel 8.2 didapatkan bahwa pemberian obat hipertensi dapat menurunkan tekanan darah sebesar 7.3 mmHg (sebelum pemberian obat) mengalami penurunan menjadi 126.5 mmHg (sesudah pemberian obat). Hasil analisis menunjukkan *p-value* 0.001, artinya ada perbedaan tekanan darah antara sebelum dan sesudah pemberian obat.

# DAFTAR PUSTAKA

Budiarto, E. 2004. *Biostatistik untuk Kedokteran dan Kesehatan Masyarakat*. Jakarta: EGC.

Hastono, S. P. 2016. *Analisis Data Bidang Kesehatan*. Jakarta: Rajawali Press.

Riyanto, A. 2017. *Statistik Inferensial Untuk Kesehatan*. Yogyakarta: Nuha Medika.

# BAB 9
# ANALISIS DATA KATAGORIKAL DENGAN ANOVA, PENYAJIAN INTERPRETASI ANALISIS ANOVA

*Oleh Joni Wilson Sitopu*

## 9.1 Pendahuluan

Analisis bivariat memungkinkan kita mempelajari hubungan yang ada antara dua variabel. Hal ini banyak memiliki kegunaan dalam kehidupan sehari-hari. Hal ini dapat membantu untuk mengetahui apakah ada hubungani antara variabel dan jika ya, lalu apa kekuatan hubungan tersebut?

Analisis bivariat membantu untuk menguji hipotesis ada atau tidak ada hubungan antara dua variabel. yaitu membantu memprediksi nilai variabel dependen berdasarkan perubahan pada variabel independen.

Analisis bivariat berarti analisis data bivariat. Ini adalah analisis statistik tunggal yang digunakan untuk mengetahui hubungan antara dua kumpulan nilai, yaitu variabel X dan Y.

1. Analisis univariat adalah bila hanya satu variabel yang dianalisis.
2. Analisis data bivariat adalah ketika tepat dua variabel dianalisis.
3. Analisis multivariat adalah ketika lebih dari dua variabel dianalisis.

Hasil yang diperoleh dari analisis bivariat disimpan dalam tabel data dengan dua kolom. Analisis bivariat tidak boleh

dikacaukan dengan analisis data dua sampel dimana variabel x dan y tidak berhubungan secara langsung.

Jenis analisis bivariat tergantung pada jenis atribut dan variabel yang digunakan untuk menganalisis data. Variabel mungkin ordinal, kategorikal, atau numerik. Variabel independen bersifat kategoris, seperti merek pena.

## 9.2 Analisis Data Katagorikal Dengan Anova

*Analysis of variance* (ANOVA) adalah teknik statistik untuk menganalisis variabilitas data untuk menyimpulkan ketidaksetaraan antara rata-rata populasi. Teknik ANOVA memperluas apa yang dapat dilakukan oleh uji t sampel independen ke berbagai cara. Hipotesis nol yang diperiksa oleh uji t sampel independen adalah bahwa dua rata-rata populasi adalah sama. Jika lebih dari dua rata-rata dibandingkan, penggunaan berulang uji t sampel independen akan menghasilkan tingkat kesalahan Tipe I yang lebih tinggi (level α berdasarkan eksperimen) daripada level α yang ditetapkan untuk setiap uji t. Pendekatan yang lebih baik daripada uji t adalah dengan mempertimbangkan semua rata-rata dalam satu hipotesis nol—yaitu, memeriksa kemungkinan hipotesis nol dengan uji statistik tunggal. Dengan demikian, peneliti tidak hanya menghemat waktu dan energi, tetapi dapat melakukan kontrol yang lebih baik terhadap kemungkinan salah menyatakan perbedaan yang signifikan di antara sarana (R. A. Fisher). Statistic yang digunakan dalam ANOVA disebut statistik F. Statistik F adalah rasio. Pembilang dan penyebutnya keduanya adalah perkiraan.

Jika pada dua variabel dalam data bivariat berada dalam bentuk statis, data diinterpretasikan, dan pernyataan serta prediksi dibuat tentangnya. Pada saat penelitian, analisis akan membantu menentukan penyebab dan dampak untuk menyimpulkan bahwa variabel yang diberikan bersifat kategorikal.

### 9.2.1 One way test

Distribusi normal tidak cukup untuk menguji hipotesis tentang mean dalam sampel data kecil dengan varians yang tidak diketahui.

Ditulis, jika X ~ N (μ, σ2), maka $\bar{x}$ ~ N (μ, σ2/n) di mana

$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ yang berarti

$Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

statistik uji Z adalah distribusi normal standar. Namun, dalam prakteknya, varians populasi $\sigma^2$ sering tidak diketahui, sehingga varians sampel $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_1 - \bar{x})^2$ harus digunakan pada tempatnya.

Varians sampel digunakan sebagai pengganti varians populasi sebenarnya, statistik uji ;

$T = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \backsim t_{n-1}$

di mana $t_{df}$ menunjukkan distribusi t Student (Student, 1908) dengan df derajat kebebasan df = n-1

a) Untuk menginterpretasikan t-test terlebih dahulu harus ditentukan :
   4) Nilai signifikansi α
   5) df (degree of freedom) = n-k, khusus untuk one sample t-test
      df = n - 1
b) Bandingkan nilai $t_{hit}$ dengan $t_{tab}$, dimana $t_{tab} = t_{\frac{\alpha}{2}n-1}$
c) Jika : thit> ttab, maka ada berbeda secara signifikansi ($H_0$ ditolak)
   Jika thit < ttab, maka tidak berbeda secara signifikansi ($H_0$ diterima)

### 9.2.2 Independent t-test

Independent t-test digunakan untuk membandingkan dua kelompok mean dari dua sampel yang berbeda (independent).

Rumus independent sample t-test

$$t = \frac{\bar{\bar{x}}_1 - \bar{x}_2}{\overline{Sx_1 - x_2}}$$

t = Nilai t hitung

$\bar{x}_1$ =mean kelompok 1

$\bar{x}_2$ =mean kelompok 2

$Sx_1 - x_2$ =Standard error kedua kelompok

Untuk mengintepretasikan t-test terlebih dahulu harus ditentukan :

a) Nilai α
b) Untuk independent sample t-test df=N-2
c) Bandingkan nilai $t_{hit}$ dengan $t_{tab}$, dimana $t_{tab} = t_{\frac{\alpha}{2}n-2}$
d) Apabila : thit> ttab, maka berbeda secara signifikansi ($H_0$ ditolak)
   Jika thit < ttab, maka Tidak berbeda secara signifikansi ($H_0$ diterima)

### 9.2.3 Paired t-test

Digunakan untuk membandingkan mean dari suatu sampel yang berpasangan (*paired*).

Sampel berpasangan adalah sebuah kelompok sampel dengan subyek yang sama namun mengalami dua perlakuan atau pengukuran yang berbeda.

Rumus paired sample t-test

$$t = \frac{\bar{x}}{s/\sqrt{n}}$$

t = Nilai t hitung

$\bar{x}$ =mean selisih pengukuran 1 & 2

$s$ =Standar deviasi selisih pengukuran 1 & 2

n =Jumlah sample

Untuk mengintepretasikan t-test terlebih dahulu harus ditentukan :

1) Nilai α
2) df (degree of freedom) = n-k
   a) Untuk paired sample t-test df = n-1
   b) Bandingkan nilai t-hit dengan nilai t-tab
   c) Jika :
3) t-hit > t-tab maka Berbeda secara signifikan ($H_0$ Ditolak)
4) t-hit < t-tab maka Tidak berbeda secara signifikan ($H_0$ Diterima)

### 9.2.4 One Way ANOVA

Anava digunakan untuk melakukan analisis komparasi (perbandingan, perbedaan) multivariable.

Uji t hanya efektif untuk mencari perbedaan 2 mean (rata-rata)

Menguji perbedaan mean melalui perhitungan berdasarkan varians.

Jenis data yg digunakan :

1. Variabel bebas : nominal atau ordinal
2. Variabel terikat: interval atau rasio

Dengan ketentuan :

1. Distribusi data normal
2. Pengambilan sampel dilakukan secara random
3. Setiap kelompok berasal dari populasi yg sama dgn variansi yg sama pula
4. Menguji kesamaan/perbedaan rerata dari 2 kelompok atau lebih. Variabel terikat (Interval dan ratio) ditinjau dari variabel bebas (nominal atau ordinal) dengan > 2 tingkatan.

Ketentuan pengujian :

1. Data awal homogeny
2. Data berdistribusi normal

3. Data dipilih secara acak (random)

**a. Pengukuran variabilitas**

Variabilitas Total (*Total sum of squares*) :

Jumlah kuadrat penyimpangan total : jumlah kuadrat selisih antara individual dengan mean totalnya ($JK_T$)

$JK_T = \sum x^2 - \frac{(\sum x_T)^2}{N}$ atau $JK_T = JK_a + JK_d$

Keterangan:

$\sum x^2$ : Jumlah x kuadrat setiap nilai

$N_i$ : banyak sampel setiap kelompok

$\sum x_T$ : total X keseluruhan

N : banyak sampel keseluruhan

**b. Antar kelompok (between treatmens)**

Variansi mean kelompok sampel terhadap mean total (perbedaan perlakuan antar kelompok)

$JK_a = \sum \frac{(\sum x)^2}{n_i} - \sum \frac{(\sum x_T)^2}{N}$

Keterangan:

$\sum X_i$ : total X setiap kelompok

$n_i$ : banyak sampel setiap kelompok

$\sum X_T$ : total X keseluruhan

N : banyak sampel keseluruhan

**c. Dalam kelompok (*within treatmens*)**

Variansi mean yang ada dalam setiap kelompok sampel terhadap mean total (tdk terpengaruh perbedaan perlakuan antar kelompok)

$JK_d = JK_T - JK_a$

Keterangan:

$JK_d$ : Jumlah kuadrat dalam kelompok

$JK_T$ : Jumlak kuadrat total

$JK_a$ : Jumlah kuadrat antar kelompok

**d. Hipotesis yang diuji:**

$H_a$ : Ada perbedaan mean tingkat penjualan produk yang signifikan antara kemasan A, B, dan C.

$H_0$ : Tidak ada perbedaan mean tingkat penjualan produk yang signifikan antara kemasan A, B, dan C.

**Hipotesis statistik yg diuji :**

$H_a : x_A \neq x_B \neq x_C$

$H_0 : x_A = x_B = x_C$

Derajat kebebasan

$dk_T = n - 1$

$dk_a = k - 1$

$dk_d = n - k$

dimana k adalah jumlah kelompok kategori dan n jumlah data dalam kelompok kategori

n : banyak sampel keseluruhan

k : banyak kelompok

**Tabel 9.1.** Tabel Ringkasan Anova

| Jumlah Variasi | JK | dk | RK | $F_h$ | $F_t$ |
|---|---|---|---|---|---|
| Antar Dalam | $JK_a$ | k-1 | $RK_a=JK_a/k-1$ | Fh= $RK_a/ RK_d$ | Ft |
| Dalam kelompok | $JK_d$ | n-k | $RK_d=JKd/n-k$ | | |
| Total | $JK_T$ | n-1 | | | |

Jika $F_h > F_t$ maka $H_o$ ditolak ($H_0 : x_A = x_B = x_C$), artinya ada perbedaan mean dan sebaliknya. Jika $F_h < F_t$ maka $H_o$ diteriama ($H_0 : x_A = x_B = x_C$), artinya tidak ada perbedaan mean. Perhitungan dengan aplikasi spss adalah sebagai berikut ;

jika sig. < 0,05 maka $H_o$ ditolak dan $H_a$ diterima, artinya ada perbedaan mean, dan sebaliknya jika sig. > 0,05 maka $H_o$ ditterima dan $H_a$ ditolak, artinya tidak ada perbedaan mean.

Selanjutnya dengan,

**e. Uji Lanjut (Tukey HSD test) (Honestly significant different)**

Rumus

$HSD = q \sqrt{(s^2_d / n)}$

Keterangan:

$S^2_d$ : mean jumlah kuadrat (RJK) dalam kelompok

n : jumlah kasus dalam tiap kelompok

q : nilai statistik (*the Studentized range statistic*), lihat tabel berikut ini ;

Jumlah kasus (n) tidak selalu sama. Gunakan *mean harmonik*

$n = k / (1/n_1 + 1/n_2 + \ldots + 1/n_k)$

k = Banyak kelompok

$n_k$ = jumlah kasus dalam kelompok

**Hipotesis:**

$H_a : x_1 \neq x_2 \neq x_3$

$H_0 : x_1 = x_2 = x_3$

Keterangan:

n : banyak sampel keseluruhan

k : banyak kelompok

Tabel Ringkasan ANOVA seperti tabel 1 diatas,

Jika $F_h > F_t$ maka $H_0$ ditolak. Jadi terdapat perbedaan mean yang signifikan antara ketiga variabel, dan sebaliknya Jika $F_h < F_t$ maka $H_0$ diterima. Jadi tidak terdapat perbedaan mean yang signifikan antara ketiga variabel

## 9.3 Penyajian Interpretasi Analisis Anova

Untuk contoh penyajian interpretasi analisis anova sbb :

### 9.3.1 One way test

Seorang peneliti dalam penelitiannya ingin mengetahui apakah ada perbedaan nilai ujian antara kelas A dan Kelas B pada Program Studi Manajemen. Universitas P.

Untuk pengujian ini jumlah kelompok responden tidak harus sama antara A dan B.

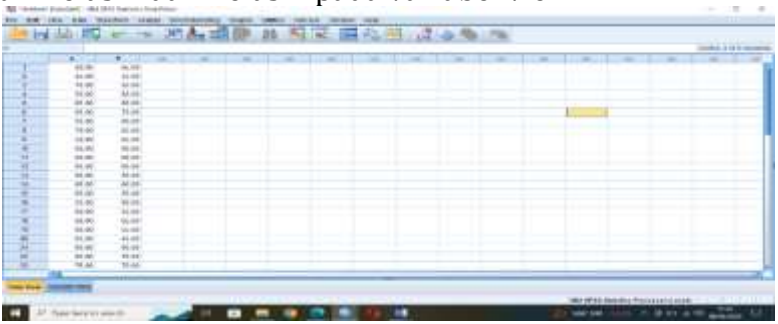
Data yang diperoleh :

| No. Responden | Nilai Ujian | |
|---|---|---|
| | Kelas A | Kelas B |
| 1 | 60 | 65 |
| 2 | 45 | 55 |
| 3 | 70 | 55 |
| 4 | 50 | 60 |
| 5 | 65 | 65 |
| 6 | 65 | 70 |
| 7 | 55 | 60 |
| 8 | 70 | 65 |
| 9 | 50 | 65 |
| 10 | 60 | 60 |
| 11 | 60 | 60 |
| 12 | 65 | 65 |
| 13 | 60 | 50 |
| 14 | 65 | 60 |
| 15 | 65 | 55 |
| 16 | 55 | 60 |
| 17 | 60 | 55 |
| 18 | 60 | 65 |
| 19 | 60 | 55 |

| | | |
|---|---|---|
| 20 | 65 | 45 |
| 21 | 60 | 60 |
| 22 | 60 | 70 |
| 23 | 75 | 70 |
| 24 | 60 | 60 |
| 25 | 65 | 65 |
| 26 | 65 | 65 |
| 27 | 60 | 70 |
| 28 | 70 | 65 |
| 29 | 60 | 65 |
| 30 | 60 | 55 |

Penyelesaiannya dengan SPSS :

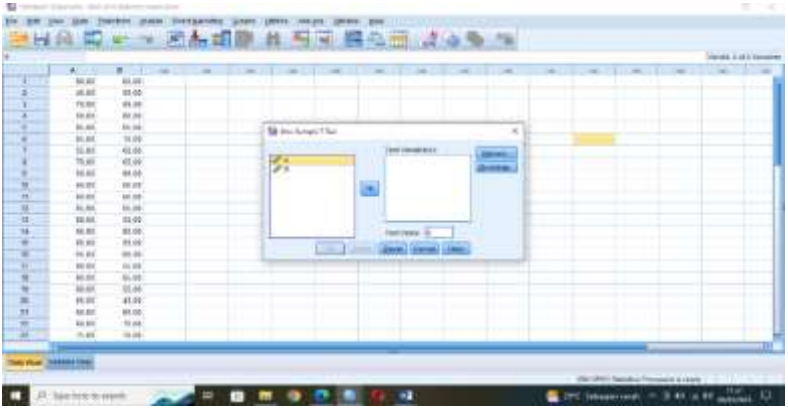
1. Buka aplikasi SPSS
2. Masukkan data (data view) kelas A dan B
3. Ketik Kelas A dan Kelas B pada variabel view



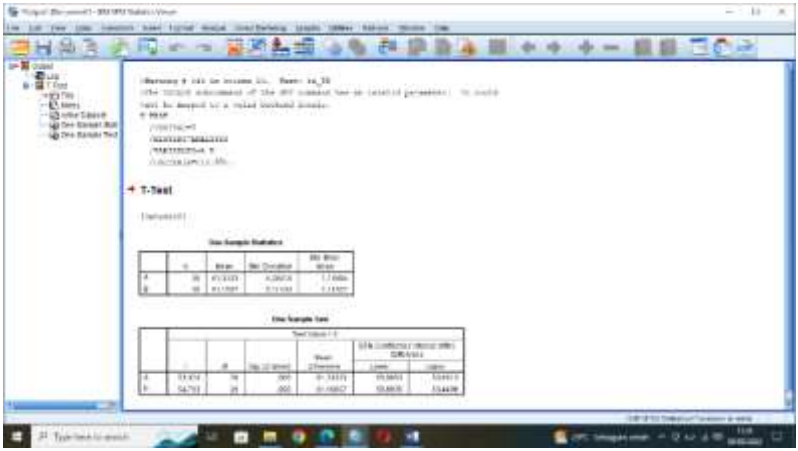
4. Klik Analyze – Compare Means – Sample T Test

5. KLIK Sample T Test

6. Klik A dan B ke test variabel



7. Klik Oke.

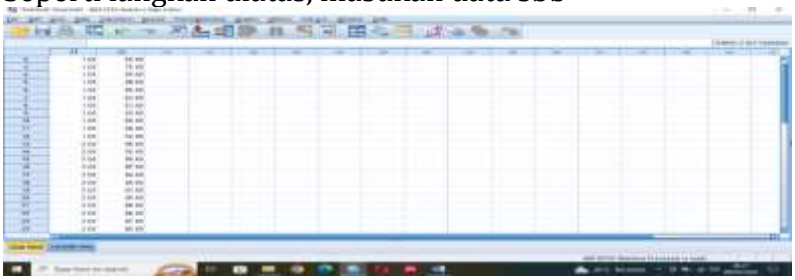


Interpretasi :

Hasil uji beda mean satu sampel. Diperoleh nilai t sebesar 23,424 dengan sig (2 tailed) 0,000. Atau jika $t_{hit} < 0,05$ maka Ho ditolak. Artinya ada perbedaan nilai ujian antara kelas A dan Kelas B pada Program Studi Manajemen. Universitas P.

### 9.3.2 Independent t-test

a) Seperti langkah diatas, masukan data sbb



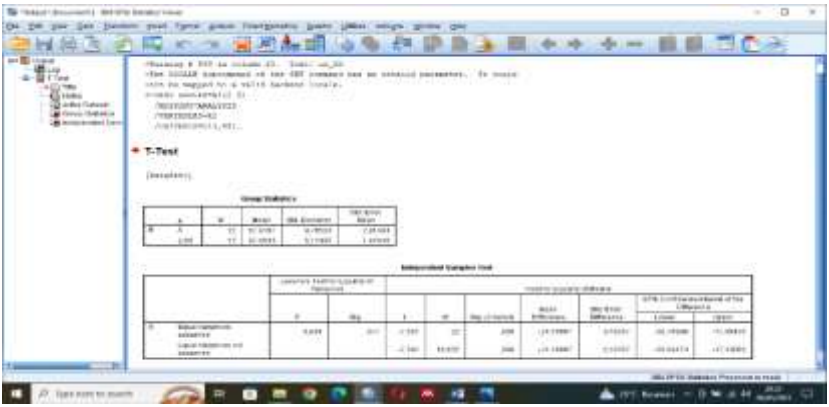
b) Klik label A value 1, label B value 2

c) kita masuk ke langkah sbb, Klik Analyze – Compare Means – Independence Sample T Test

d) Klik Independent test

e) Lakukan analisis dengan menggunakan menu Analyze ◊ Compare Means ◊ Independent Samples t test ≅ Masukkan variabel X2 ke Test Variables dan X1 ke Grouping Variable ≅ Klik tombol Define Groups lalu isikan 1 pada kotak Group 1 dan isikan 2 pada kotak Group 2 lalu klik Continue, sehingga akan terlihat seperti berikut:

f) Klik oke



Interpretasi :

Pengujian homogenitas varians (Levene's test for equality of variances). Jika nilai signifikansi pengujian F=3,434 > 0,05, atau nilai sig (0,077) > 0,05 maka varians kedua kelompok homogen. Uji yang digunakan adalah pooled t test.

Hasil uji t ditemukan nilai t sebesar -7,581 dengan sig (2-tailed) 0,077. Oleh karena nilai sig < 0,05 maka ada perbedaan mean nilai ujian antara kelas A dan Kelas B pada Program Studi Manajemen. Universitas P.

### 9.3.3 Paired t-test

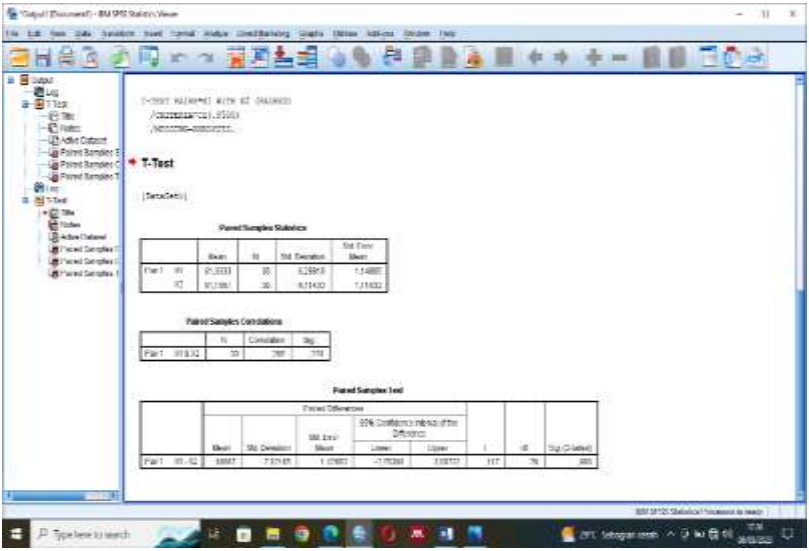
1. Masukan data

2. Klik Compare Means –sample t test

3. Klik Variabel X1 dan X1

4. Klik option-continue

5. Klik oke

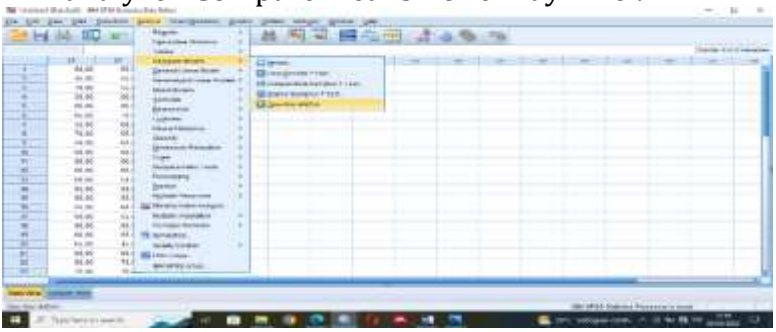



Interpretasi :

Hasil uji beda mean antara nilai pre test dan post test. Diperoleh Nilai $t_h$ = 0,117 dengan sig (2 tailed) 0,908, atau sig > 0,05. Maka Ho ditolak, artinya ada perbedaan mean nilai ujian antara kelas A dan Kelas B pada Program Studi Manajemen. Universitas P.

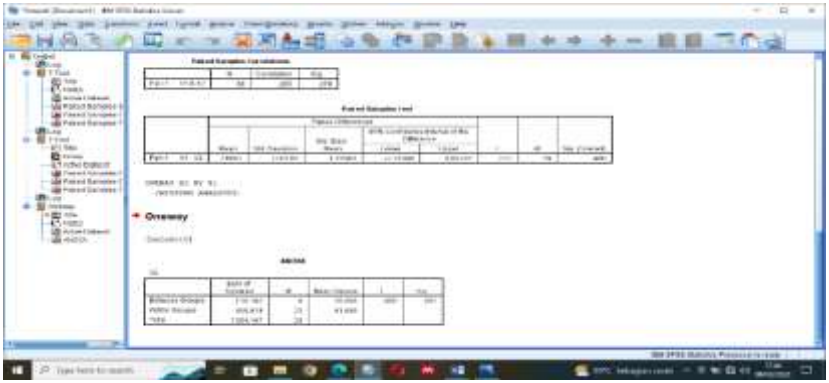
### 9.3.4 One Way ANOVA

1. Klik analyze - Compare Means - One Way ANOVA



2. Klik X1 dependent, X1 faktor

3. Klik oke



Interpretasi :

Hasil uji $F_h$ sebesar 0,500 dengan sig = 0,801. maka nilai sig > 0,05 maka Ho diterima sehingga tidak ada perbedaan mean nilai ujian antara kelas A dan Kelas B pada Program Studi Manajemen. Universitas P

# DAFTAR PUSTAKA


Anderson, David R., Sweeney, Dennis J., & Williams, Thomas A. 2011. *Statistics For Business and Economics*, Eleventh Edition.Oklohama: South-Westrn, Cengage Learning.

Hair, J.F., et.al. 2010. Multivariate Data Analysis, 7 th Edition. New York: Pearson Prentice Hall

Ho, Robert. 2006. Handbook of univariat and multivariate data analysis and interpretation with SPSS. New York: Taylor & Francis Group

Howell, David C. 2014. Fundamental Statistics for the Behavioral Sciences. Belmont, CA: Cengage Learning.

Leech, N., Karen Barrett, George A Morgan. 2005. SPSS for Intermediate Statistics Use and Interpretation. New Jersey: Lawrence Erlbaum Associates, Inc.

Mason, Robert D. & Lind, Douglas A. 1996. Teknik Statistika Untuk Bisnis & Ekonomi, Jilid I. (Alih Bahasa: Ellen Gunawan Sitompul, dkk). Jakarta: Erlangga.

McClave, et.al. 2011. Statistik untuk Bisnis dan Ekonomi Jilid 1 (Edisi kesebelas). (Alih bahasa: Bob Sabran). Jakarta: Erlangga Norusis, M.J. (1986) SPSS/PC+ for the IBM PC/XT/AT. California: SPSS Inc.

Nazariah, Noviyanti, Dita Kurniawati, Roni Priyanda, Khairul Alim, Joni Wilson Sitopu, Joko Sabtohadi, Nurul Hidayah, P.A. 2022. *Statistik Dasar*. 1st edn, *statistik dasar*. 1st edn. PT Get Press.

Sitopu, J.W., Purba, I.R. and Sipayung, T. 2021. 'Pelatihan Pengolahan Data Statistik Dengan Menggunakan Aplikasi SPSS', *Dedikasi Sains dan Teknologi*, 1(2), pp. 82–87. doi:10.47709/dst.v1i2.1068.

# BAB 10
# ANALISIS REGRESI LINIER SEDERHANA, PENYAJIAN INTERPRETASI ANALISIS

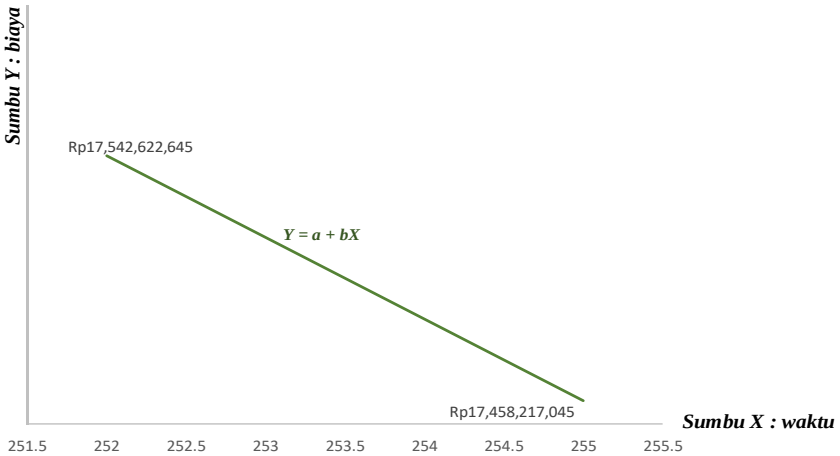
*Oleh Sediyanto*

## 10.1 Pendahuluan

Analisis regresi linear sederhana merupakan suatu proses perhitungan dengan muara membentuk persamaan garis lurus yang menggambarkan keterkaitan antara tujuan penelitian (Y) dengan masalah dalam suatu penelitian (X) (Afifah Muhartini et al., 2021). Regresi linier sederhana mengkaji kekuatan dari keterkaitan antara satu variabel bebas (X) dengan variable terikat yang merupakan tujuan dari penelitian (Y) (Refiantoro et al., 2022). Pada penelitian, biasanya variable X yang merupakan suatu permasalahan yang akan dikaji, ditentukan secara mandiri oleh peneliti, misalnya keterlambatan proyek, penyebab *variation order*, dan sebagainya (Bhirawa, 2020). Regresi linear sederhana merupakan bagian dari jenis analisis prediktif yang biasa digunakan dalam data kuantitatif dengan skala interval atau skala rasio) (Ilmi, 2019). Regresi linear sederhana juga membantu dalam melihat kapasitas kontribusi yang diberikan oleh suatu variable (Marbun *et al.*, 2020).

## 10.2 Permodelan Regresi Linier Sederhana

Permodelan ini menggambarkan suatu bentuk persamaan garis lurus yang menjelaskan keterkaitan antara tujuan penelitian (Y) dengan masalah dalam suatu penelitian (X) (Yuliara, 2016) seperti disajikan pada Gambar 10.1.




**Gambar 10.1.** Grafik regresi linier sederhana

Menurut (Zunaidhi *et al.,* 2012), notasi regresi linier sederhana berupa grafik garis lurus disajikan sebagai berikut.

$$\mathbf{Y = a + bX}$$

dimana Y adalah variable yang diprediksi, X adalah variabel yang diketahui, a adalah konstanta, dan b adalah koefisien (Afifah Muhartini et al., 2021)(Refiantoro et al., 2022).

Menurut (Ayuni & Fitrianah, 2019), persamaan untuk mendapatkan nilai a dan b sebagai berikut:

$$a = \frac{(\sum Y)(\sum X\,2)-(\sum X)(\sum XY)}{n(\sum X\,2)-(\sum X)^2}$$

$$b = \frac{n(\sum XY)-(\sum X)(\sum XY)}{n(\sum X\,2)-(\sum X)^2}$$

Maksud dari analisis regresi ini adalah untuk menemukan suatu garis yang dapat secara optimal merepresentasikan atau

mendekati sekelompok titik data yang ada (Trinanda Syahputra *et al.,* 2018).

## 10.3 Algoritma Regresi Linier Sederhana

Proses analisis dalam suatu penelitian harus mengikuti algoritma regresi linier sederhana yaitu instruksi untuk membentuk aliran terstruktur dari suatu proses agar dapat mencapai hasil yang diharapkan (Setyoningrum *et al.*, 2022).

Regresi linier sederhana adalah teknik statistik yang digunakan untuk menguji seberapa besar hubungan antara variabel yang mempengaruhi dengan variabel yang dipengaruhi. Variabel kausal atau penyebab biasanya dengan simbol huruf X, sedangkan variabel efek atau akibat dengan simbol huruf Y (Ginting et al., 2019).

Algoritma atau langkah analisis regresi linier sederhana sebagai berikut :

1. Menentukan tujuan analisis, yakni untuk memahami hubungan antara variabel yang diwakili dalam sebuah persamaan matematika.
2. Identifikasi variabel kausal ( variabel x) dan variabel efek (variabel y).
3. Melengkapi pengumpulan data yang diperlukan untuk analisis regresi linier, kemudian dibuat data primernya dalam bentuk tabel.
4. Menentukan nilai total atau sigma dari x, y, $x^2$, xy
5. Menentukan nilai konstanta a dan koefisien b
6. Proses pembuatan model regresi linier sederhana melibatkan penciptaan sebuah persamaan garis dengan menggunakan notasi dasar y = a + bx.
7. Memberi kesimpulan sebagai interprestasi dari hasil analisis yang diperoleh berupa prediksi terhadap variabel X atau variabel Y.(Ilmi, 2019)

## 10.4 Penyajian Interpretasi Analisis

Studi kasus yang akan dianalisis menggunakan regresi linier sederhana mengenai keterkaitan antara intensitas hujan dan waktu lamanya turun hujan (Bunganaen, 2013) dengan mengikuti algoritma regresi linier yang telah dijelaskan sebelumnya sebagai berikut :

1. Maksud dari melakukan analisis regresi linier sederhana adalah untuk memperoleh model persamaan yang didasarkan pada hubungan antara intensitas hujan dan waktu lamanya turun hujan.
2. Variabel X adalah waktu lamanya turun hujan dan variabel Y adalah intensitas hujan.
3. Tabel data primer penelitian yang diperoleh berdasarkan rata-rata curah hujan tahunan di daerah Z selama 10 tahun terakhir.

Intensitas hujan dan waktu lamanya turun hujan biasanya memiliki keterkaitan langsung. Apabila waktu lamanya turun hujan bertambah maka intensitas hujan akan bertambah. Rumus untuk menggambarkan keterkaitan hubungan intensitas hujan terhadap waktu lamanya turun hujan dinyatakan sebagai berikut: $H = k\ t^n$ .....rumus 1 (Soewarno dalam Bunganaen, 2013).

Keterangan :
H = intensitas hujan (satuan milimeter)
t = waktu lamanya turun hujan (satuan waktu yang digunakan adalah menit)
k = angka yang menggambarkan tingkat keterkaitan antara dua variabel dalam suatu analisis

n = pangkat dari sebuah bilangan riil positif yang memiliki nilai lebih kecil dari satu, biasanya berada dalam rentang antara 0,20 hingga 0,50.

Rumus pertama diatas kemudian dikonversi ke dalam bentuk logaritma :

log H =log k + n log t .....rumus 2

Kemudian dilakukan proses konversi lagi ke persamaan regresi linier sederhana yang dituliskan dalam bentuk ;

Y = a + bX .....rumus 3

Keterangan :

Y = Log H

A = Log k

bX = n.Log t

b = n, sehingga

X = Log t .....lihat rumus 2 dan 3

dimana Tabel 10.1 menampilkan hasil dari analisis yang telah dilakukan.

**Tabel 10.1.** Data primer rata-rata curah hujan tahunan di daerah Z
selama 10 tahun terakhir

| No | Waktu lamanya turun hujan | X = Log ti | Intensitas Hujan | Y = Log Hi |
|---|---|---|---|---|
| | Ti | (X) | Hi | (Y) |
| 1 | 5 | 0,70 | 13,87 | 1,14 |
| 2 | 10 | 1,00 | 20,56 | 1,31 |
| 3 | 15 | 1,18 | 29,40 | 1,47 |
| 4 | 30 | 1,48 | 42,44 | 1,63 |
| 5 | 45 | 1,65 | 48,74 | 1,69 |
| 6 | 60 | 1,78 | 57,19 | 1,76 |
| 7 | 120 | 2,08 | 67,82 | 1,83 |
| 8 | 180 | 2,26 | 70,72 | 1,85 |

| No | Waktu lamanya turun hujan | X = Log ti | Intensitas Hujan | Y = Log Hi |
|---|---|---|---|---|
| | Ti | (X) | Hi | (Y) |
| 9 | 360 | 2,56 | 79,50 | 1,90 |
| 10 | 720 | 2,86 | 86,42 | 1,94 |
| 11 | 1440 | 3,16 | 115,34 | 2,06 |

4. Menentukan nilai total atau sigma dari X, Y, X2, XY dari data pada Tabel 10.1.

| No | X | Y | XY | $X^2$ | $Y^2$ |
|---|---|---|---|---|---|
| 1 | 0,70 | 1,14 | 0,80 | 0,49 | 1,30 |
| 2 | 1,00 | 1,31 | 1,31 | 1,00 | 1,72 |
| 3 | 1,18 | 1,47 | 1,73 | 1,38 | 2,16 |
| 4 | 1,48 | 1,63 | 2,40 | 2,18 | 2,65 |
| 5 | 1,65 | 1,69 | 2,79 | 2,73 | 2,85 |
| 6 | 1,78 | 1,76 | 3,12 | 3,16 | 3,09 |
| 7 | 2,08 | 1,83 | 3,81 | 4,32 | 3,35 |
| 8 | 2,26 | 1,85 | 4,17 | 5,09 | 3,42 |
| 9 | 2,56 | 1,90 | 4,86 | 6,53 | 3,61 |
| 10 | 2,86 | 1,94 | 5,53 | 8,16 | 3,75 |
| 11 | 3,16 | 2,06 | 6,51 | 9,98 | 4,25 |
| Total | 20,69 | 18,58 | 37,04 | 45,03 | 32,16 |

5. Menentukan nilai konstanta a dan koefisien b

$$a = \frac{(\sum Y)(\sum X2) - (\sum X)(\sum XY)}{n(\sum X2) - (\sum X)^2}$$

$$a = 1{,}04$$

dan,

$$b = \frac{n(\sum XY) - (\sum X)(\sum XY)}{n(\sum X2) - (\sum X)^2}$$

$$b = 0{,}34$$

6. Proses pembuatan model regresi linier sederhana melibatkan sebuah persamaan garis dengan menggunakan notasi dasar Y = a + bX.

   Berdasarkan hasil analisis pada langkah 1 sampai 5, diperoleh persamaan regresi linier sederhana:

$$Y = 1{,}04 + 0{,}34X$$

7. Nilai k diperoleh dari rumus 2, dimana
   a = Log k maka,
   k = antiLog a
   k = antiLog 1,04
   k = 11, 03
   Persamaan regresi linier sederhana kemudian diubah kembali menjadi bentuk rumus untuk menggambarkan keterkaitan hubungan intensitas hujan terhadap waktu lamanya turun hujan:

$$H = k\, t^{n}$$

$$H = 11{,}03\, t^{0{,}34}$$

   Model persamaan garis ini dapat digunakan untuk memperkirakan intensitas hujan berdasarkan waktu lamanya turun hujan yang diukur. Dalam situasi dimana alat pengukur intensitas hujan mengalami kerusakan dalam kurun waktu tertentu, model persamaan ini dapat digunakan sebagai pengganti.

# DAFTAR PUSTAKA


Afifah Muhartini, A., Sahroni, O., Dwi Rahmawati, S., Febrianti, T., Mahuda, I., Saintek, F., & Bina Bangsa, U. 2021. Analisis Peramalan Jumlah Penerimaan Mahasiswa Baru Dengan Menggunakan Metode Regresi Linear Sederhana. *Jurnal Bayesian : Jurnal Ilmiah Statistika Dan Ekonometrika*, *1*(1), 17–23. https://bayesian.lppmbinabangsa.id/index.php/home/article/view/2

Ayuni, G. N., & Fitrianah, D. 2019. Penerapan metode Regresi Linear untuk prediksi penjualan properti pada PT XYZ. *Jurnal Telematika*, *14*(2), 79–86. https://journal.ithb.ac.id/telematika/article/view/321

Bhirawa, W. T. 2020. Proses Pengolahan Data Dari Model Persamaan Regresi Dengan Menggunakan Statistical Product and Service Solution (SPSS). *Statistika*, 71–83. http://journal.universitassuryadarma.ac.id/index.php/jmm/article/download/528/494

Bunganaen, W. 2013. *Analisis Hubungan Tebal Hujan dan Durasi Hujan Pada Stasiun Klimatologi Lasiana Kota Kupang*. *II*(2), 181–190.

Ginting, F., Buulolo, E., & Siagian, E. R. 2019. Implementasi Algoritma Regresi Linear Sederhana Dalam Memprediksi Besaran Pendapatan Daerah (Studi Kasus: Dinas Pendapatan Kab. Deli Serdang). *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, *3*(1), 274–279. https://doi.org/10.30865/komik.v3i1.1602

Ilmi, U. 2019. Studi Persamaan Regresi Linear Untuk Penyelesaian Persoalan Daya Listrik. *Teknika*, *11*(1), 1083–1088.

Marbun, N. J., Mahmud, S. F., & ... 2020. Strategi Pemasaran Alat Olahraga Toko Yonex Sport Di Kota Dumai. *Jurnal ARTI (Aplikasi Teknik Industri)*, *16*(1), 100–107. https://ejurnal.sttdumai.ac.id/index.php/arti/article/view/199%0Ahttps://ejurnal.sttdumai.ac.id/index.php/arti/article/download/199/143

Refiantoro, R. F., Nugroho, C. R., & Hapsari, Y. T. 2022. Analisis Regresi Sederhana Pada Nilai UAS Menggunakan Microsoft Excel Dan IBM SPSS. *Jurnal ARTI (Aplikasi Rancangan Teknik Industri)*, *17*(2), 107–116. https://ejurnal.sttdumai.ac.id/index.php/arti/article/view/396

Setyoningrum, N. R., Rahimma, P. J., Teknologi, S. T., Tanjungpinang, I., & Tanjungpinang, K. 2022. Implementasi Algoritma Regresi Linear Dalam Sistem Prediksi Pendaftar Mahasiswa Baru Sekolah Tinggi Teknologi Indonesia Tanjungpinang. *Prosiding Seminar Nasional Ilmu Sosial Dan Teknologi (SNISTEK)*, *4*, 13–18. https://ejournal.upbatam.ac.id/index.php/prosiding/article/view/5200

Trinanda Syahputra, Jufri Halim, & Karunia Perangin-Angin. 2018. Penerapan Data Mining Dalam Memprediksi Tingkat Kelulusan Uji Kompetensi ( UKOM ) Bidan Pada STIKes Senior Medan Dengan Menggunakan Metode Regresi Linier Berganda. *Sains Dan Komputer (SAINTIKOM)*, *17*(1), 1–7.

Yuliara, I. M. 2016. Modul Regresi Linier Sederhana. *Universitas Udayana*, 1–10. https://simdos.unud.ac.id/uploads/file_pendidikan_1_dir/3218126438990fa0771ddb555f70be42.pdf

Zunaidhi, R., Saputra, W. S. ., & Sari, N. K. 2012. Aplikasi Peramalan Penjualan Menggunakan Metode Regresi Linier. *Scan*, *7*(3), 41–45.

# BAB 11
# STRUCTURAL EQUATION MODELING (SEM)

***Oleh Nova Suryani***

## 11.1 Pendahuluan

*Structural equation modeling* digunakan untuk penelitian yang berfungsi untuk mengonfirmasi kebenaran suatu teori. Teori yang kuat pada *structural equation model* menjadi sangat penting dan mutlak. Analisis SEM bersifat menguji model, sehingga memerlukan teori yang mapan atau teori yang kuat. Oleh karena hal tersebut, maka peneliti harus membangun dan membuat model penelitian yang didasarkan pada teori yang kuat untuk diuji atau diteliti menggunakan structural equation modeling. Structural equation modeling ini diperuntukkan untuk menguji teori-teori yang dapat dikonfirmasi melalui data empirik.

Pemodelan persamaan struktural adalah analisis multivariat generasi kedua. Analisis SEM (*structural equation modeling*) merupakan gabungan dari perspektif ekonometrika dan psykonometrika. Ekonometrika berfokus pada prediksi dan psykonometrika berfokus pada konsep model penelitian dengan variabel laten. Variabel laten merupakan variabel yang tidak dapat diukur dan dinilai secara langsung, sehingga untuk mengukur variabel laten harus menggunakan variabel manifest.

SEM menggabungkan analisis jalur dan analisis faktor. Analisis ini memungkinkan peneliti mengestimasi dan menguji multiple variabel laten independen dan multiple variabel laten dependen secara simultan dengan banyak indikator (*variabel manifest*). Selain itu, SEM juga dapat menguji model dengan efek

mediator ataupun moderator. SEM dapat digunakan untuk melakukan analisis menggunakan variabel laten dengan model penelitian yang kompleks. Variabel laten dalam suatu model penelitian dapat diestimasi dan dievaluasi menggunakan variabel manifest. Analisis ini dapat digunakan untuk menganalisis pola hubungan antara variabel laten dan variabel manifest, variabel laten dengan variabel laten lainnya serta kesalahan pengukuran secara langsung.

SEM berfungsi untuk menjelaskan secara menyeluruh hubungan antar variabel dalam penelitian, memeriksa serta memperbaiki model penelitian. Sehingga Syarat utama menggunakan analisis SEM adalah dengan adanya model hipotesis yang dibangun peneliti. Model penetian menggunakan structural equation modeling terdiri dari model struktural dan model pengukuran yang berbentuk diagram jalur.

## 11.2 Variabel Laten dan Variabel Manifest

Variabel sangat penting dalam suatu penelitian. Variabel marupakan sesuatu yang dapat dipelajari, diamati sehingga dapat ditarik suatu kesimpulan. Variabel dalam structural equation modeling merupakan sesuatu yang dapat diamati dan dinilai baik secara langsung maupun tidak langsung.

Variabel dalam SEM terdiri dari dua variabel yaitu variabel laten dan variabel manifest. Variabel laten dapat dikatan sebagai konsep yang abstrak, misalnya perilaku, perasaan dan motivasi. Variabel laten tidak bisa diamati secara langsung sehingga tidak dapat diukur secara langsung. Maka perlu dilakukan operasionalisasi dari variabel laten untuk merepresentasikan variabel laten tersebut. Operasionalisasi dari variabel laten dinamakan dengan variabel manifest. Sehingga variabel manifest merupakan indikator dari variabel laten. Dengan kata lain variabel laten hanya dapat diamati secara tidak langsung atau tidak sempurna melalui efeknya yaitu pada variabel manifest. Sedangkan

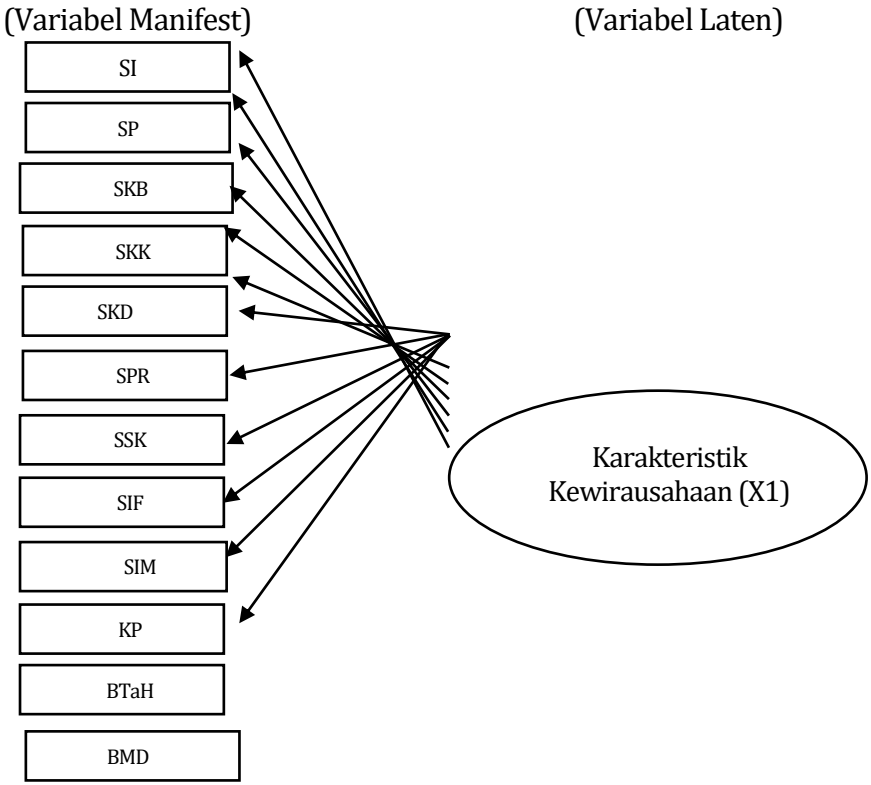
variabel manifest sendiri pada structural equation modeling merupakan variabel yang dapat secara langsung diukur secara empiris atau sering dikenal dengan indikator. Variabel manifest merupakan ukuran dari variabel laten yang harus dirumuskan berdasarkan teori.

Variabel laten digambarkan dalam bentuk oval atau ellips. Sementara itu, variabel manifest digambarkan dalam bentuk bujur sangkar. Variabel laten terbagi menjadi dua yaitu variabel eksogen dan variabel endogen. Variabel eksogen sama dengan variabel bebas, variabel endogen sama dengan variabel terikat. Variabel eksogen adalah variabel yang tidak dipengaruhi oleh variabel lain. Sedangkan variabel endogen adalah adalah variabel yang yang dipengaruhi oleh variabel lain. Sebagaimana yang dijelaskan oleh Byrne (2010) yang menjelaskan bahwa eksogen dan endogen dalam structural equation model berguna untuk membantu membedakan variabel laten yang bersifat eksogen dan variabel laten yang bersifat endogen. Variabel laten eksogen dapat disebut juga dengan variabel independen. Variabel independen dapat memberikan dampak terhadap nilai variabel laten endogen pada suatu model penelitian. Sedangkan variabel laten endogen disebut juga dengan variabel dependen atau variabel terikat, yaitu varibel yang dapat dipengaruhi oleh variabel laten eksogen dalam suatu model, secara langsung maupun secara tidak langsung.

Notasi matematik dari laten eksogen dalah ξ (“ksi“) dan variabel laten endogen ditandai dengan $\eta$ (“eta“). Variabel eksogen adalah  adalah variabel yang tidak terdapat penyebab-penyebab eksplisitnya. Variabel eksogen pada diagram diagram tidak terdapat anak panah yang mengarah pada variabel tersebut.

Variabel laten dapat dibentuk oleh variabel manifest atau indikator yang bersifat formatif dan reflektif. Indikator yang bersifat reflektif berarti bahwa variabel laten merupakan eksogen yang indikatornya merupakan endogen. Sebaliknya jika bersifat

formatif maka berarti bahwa varabel laten merupakan endogen dengan indikatornya yang eksogen.

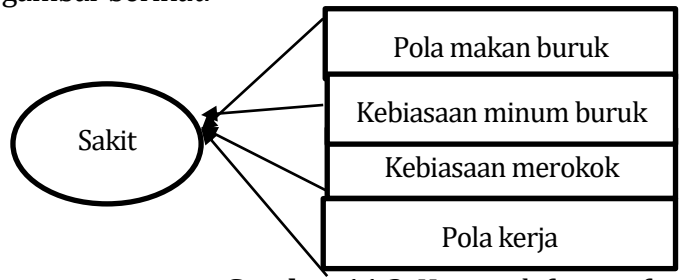



**Gambar 11.1.** Konstruk reflektif

Karakteristik kewirausahaan merupakan variabel laten karena *karakteristik* kewirausahaan tidak dapat diamati langsung. Maka, oleh sebab itu dibutuhkan indikator (variable manifestt) untuk mengukur karakteristik kewirausahaan seorang pelaku usaha seperti sifat instrumental (SI), sifat prestatif (SP), sifat keluwesan bergaul (SKB), sifat kerja keras (SKK), sifat keyakinan

diri (SKD), sifat penambilan resiko yang diperhitungkan (SPR), sifat swakendali (SSK), Sifat inovatif (SIF), sifat mandiri (SM), kepemimpinan (KP), berorientasi tugas dan hasil (BTH), dan berorientasi masa depan (BMD). Gambar 1 merupakan contoh konstruk reflektif karena karakteristik kewirausahaan dapat dilihat atau dicirikan dengan sifat instrumental (SI), sifat prestatif (SP), sifat keluwesan bergaul (SKB), sifat kerja keras (SKK), sifat keyakinan diri (SKD), sifat penambilan resiko yang diperhitungkan (SPR), sifat swakendali (SSK), Sifat inovatif (SIF), sifat mandiri (SM), kepemimpinan (KP), berorientasi tugas dan hasil (BTH), dan berorientasi masa depan (BMD).

*Contoh* konstruk yang bersifat formatif dapat dilihat pada gambar berikut:

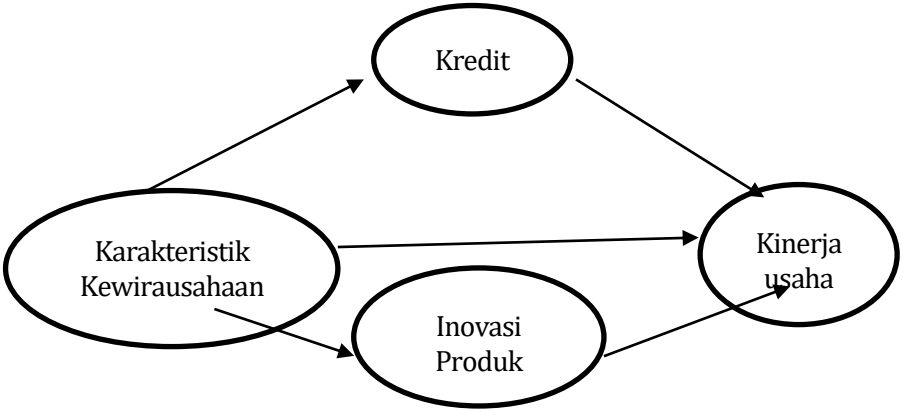



**Gambar 11.2.** Konstruk formatif

Pada gambar 11.2 dapat dijelaskan bahwa konstruk sakit ditentukan, dipengaruhi atau dibentuk oleh komponen pola makan yang buruk, kebiasaan minum buruk, kebiasaan merokok, dan pola kerja. Suatu konstruk dapat berbentuk reflektiif atau berbentuk formatif tergantung dari fenomena yang akan diteliti.

Sesama indikator (variabel manifest) pada structural equation modeling diasumsikan atau dianggap tidak berkolerasi sehingga tidak diperlukan uji konsistensi atau reabilitas internalnya. Apabila salah satu indikatornya hilang maka dapat mengakibatkan perubahan makna dari konstruk. Sementara itu

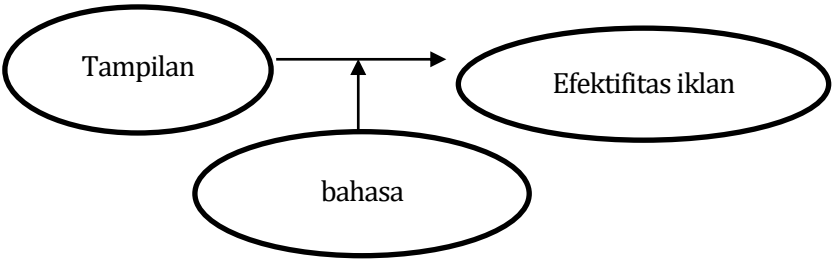
pada indikator yang bersifat formatif, tidak diperlukan kovarian antar indikator.




**Gambar 11.3.** Model dengan variabel mediasi

Pada Gambar 11.3 menjelaskan pengaruh karakteristik kewirausaan, inovasi produk dan kredit terhadap kinerja suatu usaha. Pada model di gambar 11.3, variabel kredit dan inovasi produk merupakan variabel mediasi atau intervening variabel. Variabel mediator atau variabel intervening adalah variabel yang secara teoritis mempengaruhi hubungan antar variabel independen dengan variabel dependen menjadi hubungan yang tidak langsung, sehingga tidak dapat diamati serta diukur. Variabel kredit dan inovasi produk sebagai variabel mediastor yang memediasi pengaruh antara karakteristik kewirausahaan terhadap kinerja usaha. Pada model tersebut menunjukkan bahwa variabel kinerja usaha merupakan variabel endogen, sedangkan variabel karakteristik kewirausahaan, kredit dan inovasi produk merupakan variabel eksogen. Variabel kredit, inovasi produk, dan kinerja usaha berperan sebagai variabel eksogen dan berperan sebagai variabel endogen dari variabel karakteristik kewirausahaan.

Selain variabel mediasi juga terdapat variabel moderator pada SEM. Variabel moderator terbagi menjadi variabel moderator nonmetrik dan variabel moderator metrik.




**Gambar 11.4.** Model dengan variabel moderator non metrik

Pada gambar 11.4 dapat dijelaskan bahwa tinggi rendahnya kualitas tambilan iklan yang dihasilkan berpengaruh positif dan tinggi terhadap efektifitas iklan jika didorong oleh bahasa yang digunakan dalam iklan adalah bahasa yang mayoritas digunakan oleh masyarakat. Begitu pula sebaliknya kualitas tambilan iklan yang dihasilkan tidak berpengaruh positif dan rendah terhadap efektifitas iklan jika Bahasa yang digunakan pada iklan adalah Bahasa minoritas. Variabel Bahasa dalam model merupakan variabel moderator nonmetrik.

Analisis efek moderating pada variabel moderator metrik, dapat dilakukan dengan multigroup SEM. Alternative lain untuk menguji efek moderasi pada variabel moderator dengan memodelkan suat interaksi (seperti pada model regresi moderasi).

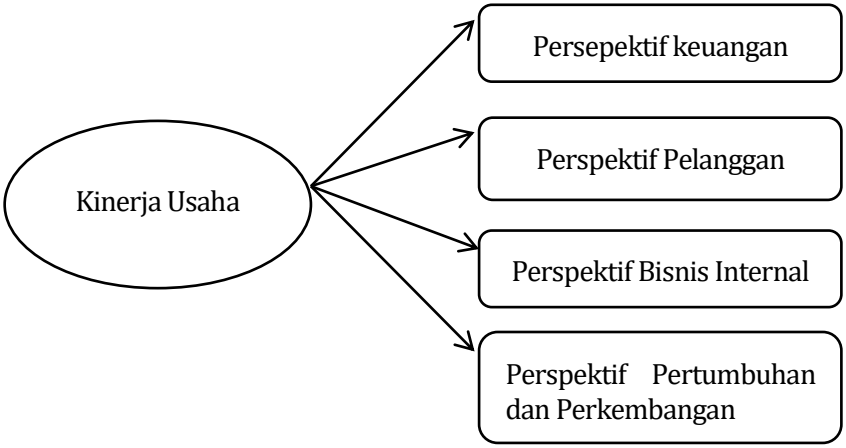
## 11.3 Model Pengukuran dan Model Structural

Analisis SEM terdiri dari dua model, yaitu model pengukuran (*mesurement model*) atau disebut juga dengan outer model dan struktural model atau inner model. Model pengukuran menunjukkan bagaimana variabel manifest atau indikator dapat

merepresentasikan variabel laten. Sedangkan model struktural menunjukkan kekuatan ekstimasi antar variabel laten.

Model pengukuran (*measurement model*) digunakan untuk menguji apakah indikator yang digunakan oleh peneliti mampu dan valid untuk mengukur variabel laten. Jumlah indikator yang cukup sangat diperlukan dalam penelitian. Hal disebabkan karena apabila suatu indikator ternyata tidak valid maka indikator tersebut harus di hapus. Haryanto (2017) yang menjelaskan bahwa model pengukuran adalah suatu teknik untuk menentukan signifikansi hubungan antar indikator yang diukur atau observed variabel dalam membentuk variabel laten atau unobserved variabel yang tidak dapat diukur secara langsung, sehingga harus dikur melalui indikator indikatornya.

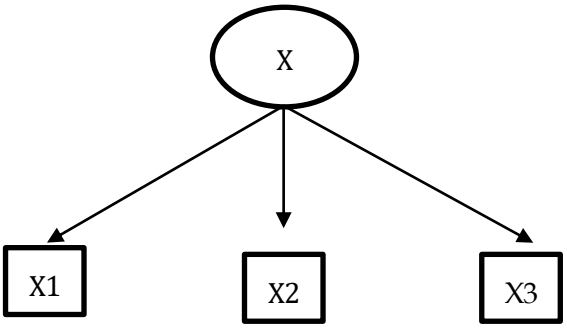
*Confimatory factor analysis* digunakan untuk menguji apakah indikator yang digunakan dapat dengan tepat menjelaskan variabel laten serta untuk mengukur apakah variabel benar-benar valid untuk digunakan. Sehingga sangat penting untuk membangun indikator berdasarkan teori.



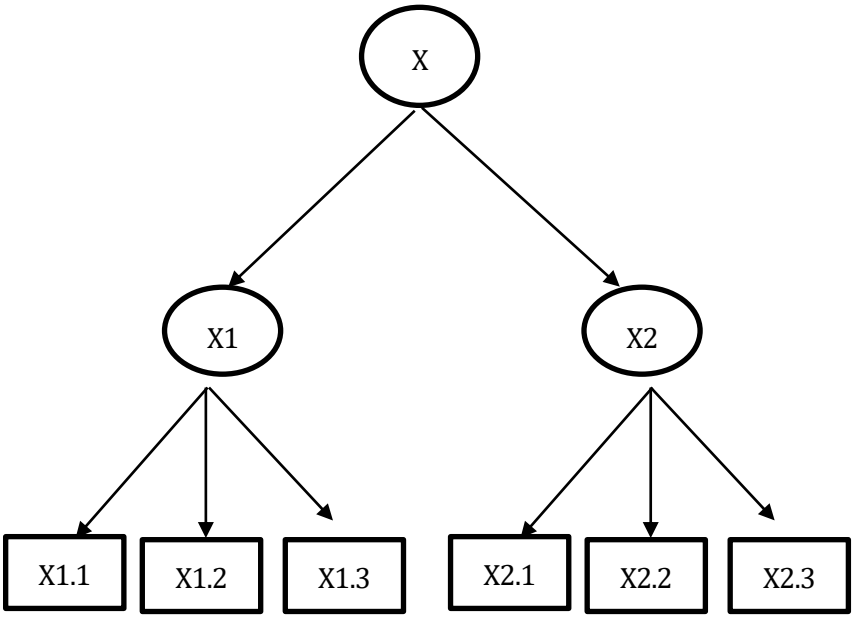

**Gambar 11.5.** Contoh model pengukuran

Misalnya kinerja usaha, dalam hal ini adalah variabel laten, sedangkan perspektif keuangan, perspektif pelanggan, persepektif bisnis internal dan perspektif pertumbuhan dan perkembangan merupakan variabel manifest (indikator) dalam model yang bersumber ari teori. Untuk mengetahui apakah perspektif keuangan, perspektif pelanggan, persepektif bisnis internal dan perspektif pertumbuhan dan perkembangan merupakan alat ukur yang kuat dan tepat untuk mengukur kinerja usaha, maka dilakukan *cofirmatory factor analysis*. Namun, jika terjadi salah satu indikator tidak kuat ataupun tidak akurat untuk merepresentasikan variabel laten, maka indikator itu harus dihapus dari model penelitian.

Model pengukuran yang variabel latennya dapat diukur langsung melalui indikator (unidimentional) disebut dengan *first order*. Variabel laten yang tidak dapat dinilai atau diukur dengan indikator secara langsung maka dilihat terlebih dahulu dimensi dari variabel laten (multidimentional) disebut dengan *second order*.




**Gambar 11.6.** Model pengukuran unidimentional dengan Indikator *first order*.

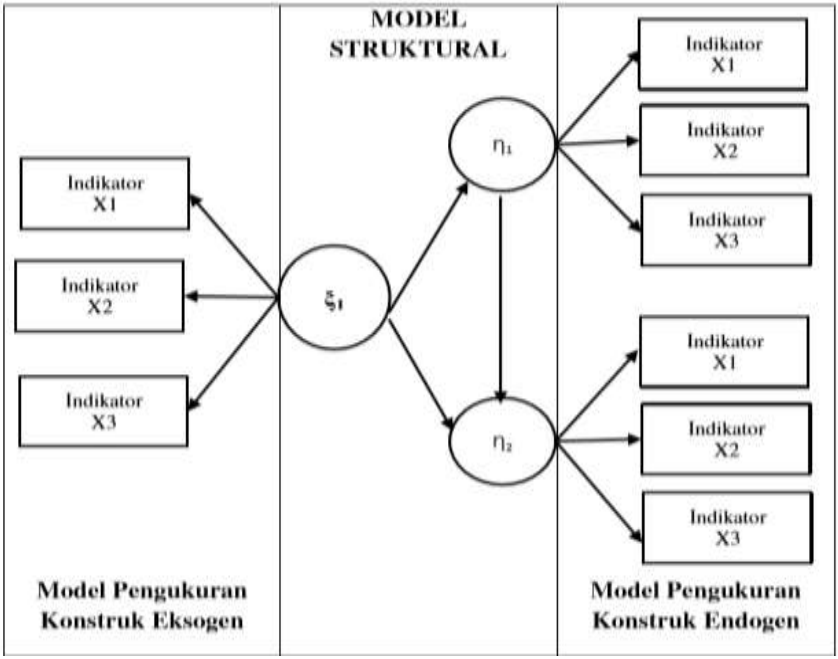



**Gambar 11.7.** Model pengukuran multidimentional dengan Indikator *second order*

Apabila peneliti meneliti konstruk berbentuk unidimentional maka dapat menggunakan *first order confirmatory factor analysis*. Sebaliknya, apabila konstruk berbentuk muldimentional maka dapat menggunakan *second order confirmatory factor analysis.*

Model struktural bertujuan untuk melihat signifikansi hubungan antar konstruk atau variabel. Hal ini dapat dilihat dari koefisien jalur yang menggambarkan kekuatan-kekuatan hubungan antar variabel. Tanda atau arah pada jalur musti sesuai dengan teori yang dihipotesiskan oleh peneliti. Menurut Haryanto (2017) menjelaskan bahwa model regresi simultan atau persamaan struktural yang terdiri dari beberapa konstruk

(variabel) eksogen, variabel intervening, varabel moderating/mediasi, dan variabel endogen.




**Gambar 11.8.** Model structural
Sumber : Latan (2013)

# DAFTAR PUSTAKA

Byrne, B. M. 2010. *Structural Equation Medeling with Amos*. In Structural Equation Modeling with Amos Basic Censepts, Aplikations, and Programming. Taylor & Francis Group. https://doi.org/10.4324/9781410600219

Latan, H. 2013. *Model Persamaan Structural Teori dan Implikasi*. Bandung: Alfabeta.

Haryono, S. 2017. *Metode SEM untuk Penelitian Manajemen AMOS Lisrel PLS*. Bekasi : PT. Intermedia Personalia Utama.

# BAB 12
# APLIKASI REGRESI LINIER GANDA

## *Oleh Abdul Wahab*

Analisis regresi ganda adalah suatu alat analisis peramalan nilai pengaruh dua variabel bebas atau lebih terhadap variabel-variabel terikat, untuk membuktikan ada atau tidaknya hubungan kausal antara dua variabel bebas atau lebih ($X_1$), ($X_2$), ($X_3$), .... , ($X_n$) dengan suatu variabel terikat.

Asumsi dan arti persamaan regresi sederhana berlaku pada regresi ganda, tetapi bedanya terletak pada rumusnya, sedangkan analisis regresi ganda dapat dihitung dengan manual atau menggunakan software statistik komputer seperti Statistikal Product and Service Solution (SPSS) versi terbaru, MINITAB, SAS, Sigmastat dan lain-lain. Apabila terdapat lebih dari dua variabel, hubungan linear dapat dinyatakan dalam persamaan regresi linear berganda sebagai berikut:

a. Dua Variabel Bebas

$$\hat{Y} = a + b_1 X_1 + b_2 X_2$$

b. Tiga Variabel Bebas

$$\hat{Y} = a + b_1 X_1 + b_2 X_2 + b_3 X_3$$

c. Ke-n Variabel Bebas

$$\hat{Y} = a + b_1 X_1 + b_2 X_2 + b_3 X_3 + .... + b_n X_n$$

Dimana: Y = Nilai observasi (data hasil pencatatan)

$\hat{Y}$ = Nilai regresi

Berarti ada satu variabel tidak bebas (*dependent variable*) yaitu $\widehat{Y}$ dan sebanyak n variabel bebas (*independent variable*), yaitu $X_1, X_2, X_3, ..., X_n$

**Langkah-langkah pengujian hipotesis:**

1. Membuat Ho dan $H_1$ dalam bentuk kalimat dan bentuk statistik
2. Menentukan uji statistik yang cocok
3. Menentukan taraf nyata
4. Menentukan daerah kritik
   Jika F hit ≤ F tab, maka terima Ho
   Jika F hit > F tab, maka tolak Ho
5. Perhitungan Statistik
   a) Menghitung nilai-nilai persamaan $b_1$, $b_2$, dan a dengan rumus:

   **Cara I: Matriks**

$$\sum Y = a.n + b_1 \sum X_1 + b_2 \sum X_2 + \ldots + b_n \sum X_n$$
$$\sum X_1 Y = a \sum X_1 + b_1 \sum X_1^2 + b_2 \sum X_1 X_2 + \ldots + b_n \sum X_1 X_n$$
$$\sum X_2 Y = a \sum X_2 + b_1 \sum X_1 X_2 + b_2 \sum X_2^2 + \ldots + b_n \sum X_2 X_n$$
$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots$$
$$\sum X_n Y = a \sum X_n + b_1 \sum X_n X_1 + b_2 \sum X_n X_2 + \ldots + b_n \sum X_n^2$$

   Dengan kata lain jika persamaan ini kita pecahkan, maka akan diperoleh nilai $b_1$, $b_2$, dan a. Kemudian dapat dibentuk persamaan regresi linear berganda. Apabila persamaan regresi sudah diperoleh, barulah kita dapat meramalkan nilai (Y) dengan syarat nilai $(X_1)$, $(X_2)$, $(X_3)$, .... , $(X_n)$ sebagai variabel bebas diketahui.

   Untuk n = 3, $\widehat{Y} = a + b_1 X_1 + b_2 X_2$, satu variabel tak bebas (Y) dan dua variabel bebas ($X_1$ dan $X_2$). $b_1$, $b_2$, dan a dihitung dari persamaan berikut:

$$
\begin{aligned}
a.n \quad &+ b_1 \sum X_1 \quad + b_2 \sum X_2 \quad = \sum Y \\
a \sum X_1 &+ b_1 \sum X_1^2 \quad + b_2 \sum X_1 X_2 \quad = \sum X_1 Y \\
a \sum X_2 &+ b_1 \sum X_1 X_2 + b_2 \sum X_2^2 \quad = \sum X_2 Y
\end{aligned}
$$

Persamaan tersebut dapat dinyatakan dalam persamaan matriks sebagai berikut:

$$
\underbrace{\begin{bmatrix} n & \sum X_1 & \sum X_2 \\ \sum X_1 & \sum X_1^2 & \sum X_1 X_2 \\ \sum X_2 & \sum X_1 X_2 & \sum X_2^2 \end{bmatrix}}_{A} \quad \underbrace{\begin{bmatrix} a \\ b_1 \\ b_2 \end{bmatrix}}_{B} = \underbrace{\begin{bmatrix} \sum Y \\ \sum X_1 Y \\ \sum X_2 Y \end{bmatrix}}_{H}
$$

Dengan : A = Matriks diketahui

H = Vektor Kolom (diketahui)

B = Vektor Kolom (tidak diketahui), dapat dicari dengan:

AB = H

B = $A^{-1}H$

$A^{-1}$ = invers dari A

Cara memecahkan persamaan lebih dari dua variabel, diantaranya dengan menggunakan determinan. Jika terdapat tiga persamaan dengan tiga variabel yang tak diketahui nilainya, yaitu $b_1$, $b_2$, dan a, maka dapat dibuat:

$$
\left.\begin{aligned}
a_{11}a &+ a_{12}b_1 + a_{13}b_2 = h_1 \\
a_{21}a &+ a_{22}b_1 + a_{23}b_2 = h_2 \\
a_{31}a &+ a_{32}b_1 + a_{33}b_2 = h_3
\end{aligned}\right\} \Rightarrow \underbrace{\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}}_{A} \quad \underbrace{\begin{bmatrix} a \\ b_1 \\ b_2 \end{bmatrix}}_{B} = \underbrace{\begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix}}_{H}
$$

$$
\det(A) = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{32}a_{21} - a_{31}a_{22}a_{13} - a_{21}a_{12}a_{33} - a_{11}a_{23}a_{32}
$$

Maka, $a = \frac{\det(A_1)}{\det(A)}$; $b_1 = \frac{\det(A_2)}{\det(A)}$; $b_2 = \frac{\det(A_3)}{\det(A)}$

Dimana,

$$A_1 = \begin{bmatrix} h_1 & a_{12} & a_{13} \\ h_2 & a_{22} & a_{23} \\ h_3 & a_{32} & a_{33} \end{bmatrix}; \quad A_2 = \begin{bmatrix} a_{11} & h_1 & a_{13} \\ a_{21} & h_2 & a_{23} \\ a_{31} & h_3 & a_{33} \end{bmatrix}; \quad A_3 = \begin{bmatrix} a_{11} & a_{12} & h_1 \\ a_{21} & a_{22} & h_2 \\ a_{31} & a_{32} & h_3 \end{bmatrix}$$

$$\det(A_1) = h_1a_{22}a_{33} + a_{12}a_{23}h_3 + a_{13}h_2a_{32} - a_{12}h_2a_{33} - h_1a_{23}a_{32} - a_{13}a_{22}h_3$$
$$\det(A_2) = a_{11}h_2a_{33} + h_1a_{23}a_{31} + a_{13}a_{21}h_3 - h_1a_{21}a_{33} - a_{11}a_{23}h_3 - a_{13}h_2a_{31}$$
$$\det(A_3) = a_{11}a_{22}h_3 + a_{12}h_2a_{31} + h_1a_{21}a_{32} - a_{12}a_{21}h_3 - a_{11}h_2a_{32} - h_1a_{22}a_{31}$$

**Cara II: Tabel Penolong :**

| No. | $X_1$ | $X_2$ | Y | $X_1^2$ | $X_2^2$ | $Y^2$ | $X_1Y$ | $X_2Y$ | $X_1X_2$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | .. | .. | .. | .. | .. | .. | .. | .. | .. |
| 2 | .. | .. | .. | .. | .. | .. | .. | .. | .. |
| . | .. | .. | .. | .. | .. | .. | .. | .. | .. |
| N | .. | .. | .. | .. | .. | .. | .. | .. | .. |
| Statistik | $\sum X_1$ | $\sum X_2$ | $\sum Y$ | $\sum X_1^2$ | $\sum X_2^2$ | $\sum Y^2$ | $\sum X_1Y$ | $\sum X_2Y$ | $\sum X_1X_2$ |

*a.* $\sum x_1^2 = \sum X_1^2 - \frac{\left(\sum X_1\right)^2}{n}$

*b.* $\sum x_2^2 = \sum X_2^2 - \frac{\left(\sum X_2\right)^2}{n}$

*c.* $\sum y^2 = \sum Y^2 - \frac{\left(\sum Y\right)^2}{n}$

$$d.\quad \sum x_1 y = \sum X_1 Y - \frac{\left(\sum X_1\right).\left(\sum Y\right)}{n}$$

$$e.\quad \sum x_2 y = \sum X_2 Y - \frac{\left(\sum X_2\right).\left(\sum Y\right)}{n}$$

$$f.\quad \sum x_1 x_2 = \sum X_1 X_2 - \frac{\left(\sum X_1\right).\left(\sum X_2\right)}{n}$$

Masukkan hasil dari nilai–nilai statistik kedalam rumus, kemudian masukkan hasil dari jumlah kuadrat ke persamaan $b_1$, $b_2$, dan a :

$$b_1 = \frac{\left(\sum x_2^2\right).\left(\sum x_1 y\right) - \left(\sum x_1 x_2\right).\left(\sum x_2 y\right)}{\left(\sum x_1^2\right).\left(\sum x_2^2\right) - \left(\sum x_1 x_2\right)^2}$$

$$b_2 = \frac{\left(\sum x_1^2\right).\left(\sum x_2 y\right) - \left(\sum x_1 x_2\right).\left(\sum x_1 y\right)}{\left(\sum x_1^2\right).\left(\sum x_2^2\right) - \left(\sum x_1 x_2\right)^2}$$

$$a = \frac{\sum Y}{n} - b_1.\left(\frac{\sum X_1}{n}\right) - b_2.\left(\frac{\sum X_2}{n}\right)$$

b) Mencari korelasi ganda dengan rumus:

$$\left(R_{x_1, x_2, y}\right) = \sqrt{\frac{b_1.\sum x_1 y + b_2.\sum x_2 y}{\sum y^2}}$$

c) Mencari nilai kontribusi korelasi ganda dengan rumus:

$$KP = \left(R_{x_1, x_2, y}\right)^2.100\%$$

d) Menguji signifikansi dengan membandingkan $F_{hitung}$ dengan $F_{tabel}$, Dengan rumus:

$$F_{hitung} = \frac{R^2\left(n - m - 1\right)}{m.\left(1 - R^2\right)}$$

Dimana : n = Jumlah responden
m = Jumlah variabel bebas

Kaidah pengujian signifikansi: jika $F_{hitung} \geq F_{tabel}$ , maka tolak $H_0$ artinya signifikan dan jika $F_{hitung} < F_{tabel}$ , maka terima $H_0$ artinya tidak signifikan Pada taraf signifikansi: $\alpha = 0{,}01$ atau $\alpha = 0{,}05$

Carilah nilai $F_{tabel}$ menggunakan Tabel F dengan rumus:

$F_{tabel} = F_{\{(1-\alpha)\,(dk\ pembilang = m),\,(dk\ penyebut = n-m-1)\}}$

6. Membuat Kesimpulan

   Membuat kesimpulan berarti menerima atau menolak hipotesis

**Contoh 1.**

Sebuah penelitian dilakukan untuk mengetahui apakah terdapat pengaruh yang signifikan antara pendapatan perminggu ($X_1$) dan jumlah anggota rumah tangga ($X_2$) terhadap pengeluaran untuk pembelian barang-barang tahan lama perminggu (Y), dengan data sebagai berikut:

| No. | $X_1$ | $X_2$ | Y |
|---|---|---|---|
| 1. | 10 | 7 | 23 |
| 2. | 2 | 3 | 7 |
| 3. | 4 | 2 | 15 |
| 4. | 6 | 4 | 17 |
| 5. | 8 | 6 | 23 |
| 6. | 7 | 5 | 22 |
| 7. | 4 | 3 | 10 |
| 8. | 6 | 3 | 14 |
| 9. | 7 | 4 | 20 |
| 10. | 6 | 3 | 19 |

*Pertanyaan:*

a. Tentukan persamaan regresi ganda?
b. Buktikan apakah ada pengaruh yang signifikan antara pendapatan perminggu dan jumlah anggota rumah tangga terhadap pengeluaran untuk pembelian barang–barang tahan lama perminggu?

*Penyelesaian*

a. Persamaan Regresi Ganda

| No. | $X_1$ | $X_2$ | Y | $X_1^2$ | $X_2^2$ | $Y^2$ | $X_1Y$ | $X_2Y$ | $X_1X_2$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 10 | 7 | 23 | 100 | 49 | 529 | 230 | 161 | 70 |
| 2 | 2 | 3 | 7 | 4 | 9 | 49 | 14 | 21 | 6 |
| 3 | 4 | 2 | 15 | 16 | 4 | 225 | 60 | 30 | 8 |
| 4 | 6 | 4 | 17 | 36 | 16 | 289 | 102 | 68 | 24 |
| 5 | 8 | 6 | 23 | 64 | 36 | 529 | 184 | 138 | 48 |
| 6 | 7 | 5 | 22 | 49 | 25 | 484 | 154 | 110 | 35 |
| 7 | 4 | 3 | 10 | 16 | 9 | 100 | 40 | 30 | 12 |
| 8 | 6 | 3 | 14 | 36 | 9 | 196 | 84 | 42 | 18 |
| 9 | 7 | 4 | 20 | 49 | 16 | 100 | 140 | 80 | 28 |
| 10 | 6 | 3 | 19 | 36 | 9 | 361 | 114 | 57 | 18 |
| Jmh | 60 | 40 | 170 | 406 | 182 | 3162 | 1122 | 737 | 267 |

Persamaan :

$$a.n + b_1\sum X_1 + b_2\sum X_2 = \sum Y \quad Maka: \; 10a + 60b_1 + 40b_2 = 170$$

$$a\sum X_1 + b_1\sum X_1^2 + b_2\sum X_1X_2 = \sum X_1Y \; Maka: \; 60a + 406b_1 + 267b_2 = 1122$$

$$a\sum X_2 + b_1\sum X_1X_2 + b_2\sum X_2^2 = \sum X_2Y \; Maka: \; 40a + 267b_1 + 182b_2 = 737$$

Maka persamaan matriksnya:

$$\underbrace{\begin{bmatrix} 10 & 60 & 40 \\ 60 & 406 & 267 \\ 40 & 267 & 182 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} a \\ b_1 \\ b_2 \end{bmatrix}}_{B} = \underbrace{\begin{bmatrix} 170 \\ 1122 \\ 737 \end{bmatrix}}_{H}$$

Det(A)=(10)(406)(182)+(60)(267)(40)+(40)(60)(267)-(60)(60)(182)-(10)(267)(267)-(40)(406)(40) = 2830

$$A_1 = \begin{bmatrix} 170 & 60 & 40 \\ 1122 & 406 & 267 \\ 737 & 267 & 182 \end{bmatrix}; \quad A_2 = \begin{bmatrix} 10 & 170 & 40 \\ 60 & 1122 & 267 \\ 40 & 737 & 182 \end{bmatrix}; \quad A_3 = \begin{bmatrix} 10 & 60 & 170 \\ 60 & 406 & 1122 \\ 40 & 267 & 737 \end{bmatrix}$$

Det $(A_1)$ = (170)(406)(182) + (60)(267)(182) + (40)(1122)(267) - (60)(1122)(182) - (170)(267) (267) - (40)(406)(737) = 11090

Det $(A_2)$ = (10)(1122)(182) + (170)(267)(40) + (40)(60)(737) - (170)(60)(182) - (10)(267)(737) - (40)(1122)(40) = 7050

Det $(A_3)$ = (10)(406)(737) + (60)(1122)(40) + (170)(60)(267) - (60)(60)(737) - (10)(1122)(267) - (170)(406)(40) = - 1320

Jadi,

$$a = \frac{\det(A_1)}{\det(A)}; \quad b_1 = \frac{\det(A_2)}{\det(A)}; \quad b_2 = \frac{\det(A_3)}{\det(A)}$$

$$= \frac{11090}{2830} \qquad = \frac{7050}{2830} \qquad = \frac{-1320}{2830}$$

$$= 3{,}92 \qquad = 2{,}49 \qquad = -0{,}47$$

Persamaan Regresi Linear ganda :

$$\hat{Y} = 3{,}92 + 2{,}49X_1 - 0{,}47X_2$$

b. Pengujian Signifikansi

1. Membuat $H_1$ dan Ho dalam bentuk kalimat dan dalam bentuk statistik

   $H_0$ : Tidak terdapat pengaruh yang signifikan pendapatan perminggu dan jumlah anggota rumah tangga terhadap pengeluaran untuk pembelian barang–barang tahan lama perminggu.

   $H_1$ : Terdapat pengaruh yang signifikan pendapatan perminggu dan jumlah anggota rumah tangga terhadap pengeluaran untuk pembelian barang–barang tahan lama perminggu.

   $H_0$ : R = 0 lawan $H_1$ : R ≠ 0
2. Menentukan uji statistik yang cocok: Uji F
3. Menentukan taraf nyata : α = 5 % = 0,05
   dk pembilang = m = 2

dk penyebut = n – m – 1 = 10 – 2 – 1 = 7

$F_{tabel} = F_{0,05} = 4,74$

4. Menentukan daerah kritik

   Penerimaan $H_0$ : $F_{hit} \leq 4,74$

   Penolakan $H_0$ : $F_{hit} > 4,74$

5. Perhitungan statistik

   a. Mencari korelasi ganda dengan rumus :

$$\left(R_{x_1,x_2,y}\right) = \sqrt{\frac{b_{1.}\sum x_1 y + b_2 . \sum x_2 y}{\sum y^2}}$$

$$a. \quad \sum y^2 = \sum Y^2 - \frac{\left(\sum Y\right)^2}{n} = 3162 - \frac{(170)^2}{10} = 272$$

$$b. \quad \sum x_1 y = \sum X_1 Y - \frac{\left(\sum X_1\right).\left(\sum Y\right)}{n} = 1122 - \frac{(60)(170)}{10} = 102$$

$$c. \quad \sum x_2 y = \sum X_2 Y - \frac{\left(\sum X_2\right).\left(\sum Y\right)}{n} = 737 - \frac{(40)(170)}{10} = 57$$

$$\left(R_{x_1,x_2,y}\right) = \sqrt{\frac{b_{1.}\sum x_1 y + b_2 . \sum x_2 y}{\sum y^2}}$$

$$= \sqrt{\frac{(2,49)(102) + (-0,47)(57)}{272}} = \sqrt{0,84} = 0,915$$

   b. Mencari nilai kontribusi korelasi ganda dengan rumus:

$$KP = \left(R_{x_1,x_2,y}\right)^2 .100\% = (0,915)^2 .100\% = 83,72\%$$

   c. Menguji signifikansi dengan membandingkan $F_{hitung}$ dengan $F_{tabel}$

   dengan rumus :

$$F_{hitung} = \frac{R^2(n-m-1)}{m.(1-R^2)} = \frac{(0,915)^2(10-2-1)}{2(1-0.915^2)} = 17,89$$

6. Membuat kesimpulan
   Karena $F_{hitung}$=17,89 > $F_{tabel}$ =4,74, maka tolak $H_0$ dan terima $H_1$, artinya terdapat pengaruh yang signifikan pendapatan perminggu dan jumlah anggota rumah tangga terhadap pengeluaran untuk pembelian barang–barang tahan lama perminggu.

Analisis data (MINITAB)

**Regression Analysis: Y versus X1, X2**

The regression equation is

**Y = 3.92 + 2.49 X1 - 0.47 X2**

| Predictor | Coef | SE Coef | T | P | VIF |
|---|---|---|---|---|---|
| Constant | 3.919 | 2.418 | 1.62 | 0.149 | |
| X1 | 2.4912 | 0.7029 | 3.54 | 0.009 | 3.6 |
| X2 | -0.466 | 1.016 | -0.46 | 0.660 | 3.6 |

S = 2.52099 R-Sq = 83.6% R-Sq(adj) = 79.0%
PRESS = 96.7469 R-Sq(pred) = 64.43%

Analysis of Variance

| Source | DF | SS | MS | F | P |
|---|---|---|---|---|---|
| Regression | 2 | 227.51 | 113.76 | 17.90 | 0.002 |
| Residual Error | 7 | 44.49 | 6.36 | | |
| Lack of Fit | 6 | 31.99 | 5.33 | 0.43 | 0.823 |
| Pure Error | 1 | 12.50 | 12.50 | | |
| Total | 9 | 272.00 | | | |

8 rows with no replicates

| Source | DF | Seq SS |
|---|---|---|
| X1 | 1 | 226.17 |
| X2 | 1 | 1.34 |

Durbin-Watson statistic = 1.51603
No evidence of lack of fit (P >= 0.1).

**Interpretasi Data**

**Uji Hipotesis**

Hasil analisis variansi (seperti pada hasil analisis di atas) adalah sebagai berikut:

Analysis of Variance

| Source | DF | SS | MS | F | P |
|---|---|---|---|---|---|
| Regression | 2 | 227.51 | 113.76 | 17.90 | 0.002 |
| Residual Error | 7 | 44.49 | 6.36 | | |
| Total | 9 | 272.00 | | | |

Hasil perhitungan pengujian hipotesis dengan taraf keberartian $\alpha$=0,05 yang menguji Ho: tidak terdapat pengaruh yang signifikan pendapatan perminggu (X1) dan jumlah anggota rumah tangga (X2) terhadap pengeluaran untuk pembelian barang–barang tahan lama perminggu (Y) lawan $H_1$: terdapat pengaruh yang signifikan pendapatan perminggu (X1) dan jumlah anggota rumah tangga (X2) terhadap pengeluaran untuk pembelian barang–barang tahan lama perminggu (Y) dengan ketentuan tolak Ho jika $p < \alpha$ ($H_1$ diterima) dan sebaliknya tolak $H_1$ jika $p > \alpha$ (Ho diterima).

Uji keberartian persamaan regresi menghasilkan $F_{hitung}$ = 17,90 > $F_{tabel}$ = 4,74 atau p = 0,002 < $\alpha$= 0,05. Hal ini menunjukkan bahwa persamaan regresi Y atas X1, X2 adalah berarti *(Signifikan)*. Dengan kata lain secara simultan terdapat pengaruh yang signifikan pendapatan perminggu (X1) dan jumlah anggota rumah tangga (X2) terhadap pengeluaran untuk pembelian barang–barang tahan lama perminggu (Y).

Demikian pula pengaruh sendiri-sendirinya, pendapatan perminggu (X1) diperoleh Nilai T = 3,54 atau p= 0,009 < $\alpha$= 0,05 yang menunjukkan bahwa terdapat pengaruh pendapatan perminggu (X1) terhadap pengeluaran untuk pembelian barang–barang tahan lama perminggu (Y) dengan memperhatikan jumlah anggota rumah tangga (X2). Namun jumlah anggota rumah tangga (X2) diperoleh Nilai T = -0,46 atau p= 0,660 > $\alpha$= 0,05 yang menunjukkan bahwa tidak terdapat pengaruh jumlah anggota rumah tangga (X2) terhadap pengeluaran untuk pembelian barang–barang tahan lama perminggu (Y) dengan memperhatikan pendapatan perminggu (X1).

Koefisien determinasi ($R^2$) diperoleh 0,836. Hal ini menunjukkan bahwa 83,6% total variansi pengeluaran untuk pembelian barang–barang tahan lama perminggu (Y) dapat dijelaskan oleh pendapatan perminggu (X1) dan jumlah anggota rumah tangga (X2).

**Contoh 2.**

Diberikan Judul penelitian "Pengaruh Umur dan Tinggi terhadap Berat Badan di Rumah Sakit Hasan Sadikin Bandung". Data dianggap memenuhi asumsi dan persyaratan analisis; data dipilih secara random; berdistribusi normal; berpola linear; data homogen; dengan data sebagai berikut:

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Umur ($X_1$) thn | 9 | 12 | 6 | 10 | 9 | 10 | 7 | 8 | 11 | 6 | 10 | 8 | 12 | 10 |
| Tinggi ($X_2$) cm | 125 | 137 | 99 | 122 | 129 | 128 | 96 | 104 | 132 | 95 | 114 | 101 | 146 | 132 |
| Berat (Y) kg | 37 | 41 | 34 | 39 | 39 | 40 | 37 | 39 | 42 | 35 | 41 | 40 | 43 | 38 |

*Pertanyaan:*

a. Tentukan persamaan regresi ganda?
b. Buktikan apakah ada pengaruh yang signifikan antara umur dan tinggi terhadap berat badan di rumah sakir Hasan Sadikin Bandung?

*Penyelesaian:*
Analisis data (MINITAB)
**Regression Analysis: Y versus X1, X2**
The regression equation is
**Y = 33.6 + 1.98 X1 - 0.107 X2**

| Predictor | Coef | SE Coef | T | P | VIF |
|---|---|---|---|---|---|
| Constant | 33.565 | 2.692 | 12.47 | 0.000 | |
| X1 | 1.9758 | 0.4115 | 4.80 | 0.001 | 6.3 |
| X2 | -0.10712 | 0.04775 | -2.24 | 0.046 | 6.3 |

S = 1.15870 R-Sq = 82.6% R-Sq(adj) = 79.4%
PRESS = 24.9731 R-Sq(pred) = 70.60%

Analysis of Variance

| Source | DF | SS | MS | F | P |
|---|---|---|---|---|---|
| Regression | 2 | 70.160 | 35.080 | 26.13 | 0.000 |
| Residual Error | 11 | 14.768 | 1.343 | | |
| Total | 13 | 84.929 | | | |

No replicates.
Cannot do pure error test.

| Source | DF | Seq SS |
|---|---|---|
| X1 | 1 | 63.403 |
| X2 | 1 | 6.758 |

Durbin-Watson statistic = 1.42042
No evidence of lack of fit (P >= 0.1).

**Interpretasi Data**

**Uji Hipotesis**

Hasil analisis variansi (seperti pada hasil analisis di atas) adalah sebagai berikut:

Analysis of Variance

| Source | DF | SS | MS | F | P |
|---|---|---|---|---|---|
| Regression | 2 | 70.160 | 35.080 | 26.13 | 0.000 |
| Residual Error | 11 | 14.768 | 1.343 | | |
| Total | 13 | 84.929 | | | |

Hasil perhitungan pengujian hipotesis dengan taraf keberartian $\alpha$=0,05 yang menguji Ho: tidak terdapat pengaruh yang signifikan umur (X1) dan tinggi (X2) terhadap berat badan (Y) lawan $H_1$: terdapat pengaruh yang signifikan umur (X1) dan tinggi (X2) terhadap berat badan (Y) dengan ketentuan tolak Ho jika $p < \alpha$ ($H_1$ diterima) dan sebaliknya tolak $H_1$ jika $p > \alpha$ (Ho diterima).

Uji keberartian persamaan regresi menghasilkan p=0,000< $\alpha$= 0,05. Hal ini menunjukkan bahwa persamaan regresi Y atas X1, X2 adalah berarti *(Signifikan)*. Dengan kata lain secara simultan terdapat pengaruh yang signifikan umur (X1) dan tinggi (X2) terhadap berat badan (Y).

Demikian pula pengaruh sendiri-sendirinya, pendapatan perminggu (X1) diperoleh Nilai T = 4,80 atau p= 0,001 < $\alpha$= 0,05 yang menunjukkan bahwa terdapat pengaruh umur (X1) terhadap berat badan (Y) dengan memperhatikan tinggi badan (X2). Namun jumlah anggota rumah tangga (X2) diperoleh Nilai T = -0,24 atau p= 0,0460 < $\alpha$= 0,05 yang menunjukkan bahwa terdapat pengaruh tinggi badan (X2) terhadap berat badan (Y) dengan memperhatikan umur (X1).

Koefisien determinasi ($R^2$) diperoleh 0,826. Hal ini menunjukkan bahwa 82,6% total variansi berat badan (Y) dapat dijelaskan oleh umur (X1) dan tinggi badan (X2).

# DAFTAR PUSTAKA


Darmadi, H. 2011. *Metode Penelitian Pendidikan*. Bandung: Alfabeta.

Draper, M.R & H. Smith. 1981. (Alih bahasa Bambang Sumantri, 1992). *Applied Regression Analysis.* Second Edition. Institut Pertanian Bogor.

Iriawan, N. dan Astuti,S.P.2006. *Mengolah Data Statistik dengan Menggunakan Minitab 14*. Yogyakarta: Andi Offset.

*Minitab Statistical Software, Release 14*. State College, Pa.: Minitab, Inc. 2000.

Purwanto, 2011. *Statistika untuk Penelitian*. Yogyakarta : Pustaka Pelajar.

Sembiring, R.K. 1995. *Analisis Regresi*. ITB Bandung.

Trihendradi, S. 2009. *Step by Step SPSS 16 Analisis Data Statistik*. Yogyakarta: Andi Offset.

Walpole, R.E. & Myers, R.H. 1989. (Terjemahan R.K. Sembiring, 1995. *Ilmu Peluang dan Statistika untuk Insinyur dan Ilmuan*. Edisi ke-4. ITB Bandung.

Wibisono, Y. 2005. *Metode Statistik*. Gadjah Mada University Press: Yogyakarta.

# BAB 13
# ANALISIS MULTIVARIAT, ANALISIS REGRESI LOGISTIK PENYAJIAN DAN INTERPRETASI ANALISIS

*Oleh Adhar Arifuddin*

## 13.1 Pengertian dan Konsep Dasar Analisis Regresi Logistik

Regresi logistik adalah suatu teknik analisis statistik yang digunakan untuk memodelkan hubungan antara satu atau lebih variabel prediktor (variabel independen) dan variabel respon biner (variabel dependen). Regresi logistik memiliki tujuan untuk memprediksi kemungkinan kejadian suatu peristiwa (misalnya, apakah seorang pasien akan terkena penyakit tertentu atau tidak) berdasarkan beberapa variabel prediktor seperti usia, jenis kelamin, riwayat kesehatan, dan lain-lain (Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X, 2013). Artinya, Regresi logistik adalah sebuah metode statistik yang digunakan untuk menganalisis hubungan antara variabel dependen biner (dichotomous) dengan satu atau lebih variabel independen. Variabel dependen biner adalah variabel yang hanya memiliki dua kemungkinan nilai, misalnya ya atau tidak, sukses atau gagal, hidup atau mati (Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X, 2013).

Dalam regresi logistik, variabel dependen biasanya direpresentasikan sebagai variabel dummy, yaitu variabel yang bernilai 1 jika kejadian terjadi dan 0 jika kejadian tidak terjadi.

Tujuan dari regresi logistik adalah untuk mempelajari faktor-faktor yang mempengaruhi kemungkinan terjadinya suatu kejadian (Agresti, A, 2015). Model regresi logistik menghasilkan nilai probabilitas untuk setiap kategori pada variabel respon dan digunakan untuk memprediksi kategori variabel respon dari nilai-nilai variabel predictor (Kleinbaum, D. G., & Klein, M, 2010).

Regresi logistik sering digunakan dalam berbagai bidang, termasuk ekonomi, kesehatan, ilmu sosial, dan pemasaran. Contoh penggunaan regresi logistik termasuk memprediksi kemungkinan seseorang membeli produk tertentu, kemungkinan seorang pasien terkena penyakit tertentu, dan kemungkinan keberhasilan atau kegagalan suatu proyek (Agresti, A, 2019).

## 13.2 Model Analisis Regresi Logistik

Konsep dasar regresi logistik adalah untuk memodelkan hubungan antara variabel prediktor (biasanya berupa variabel independen) dan variabel target (biasanya berupa variabel dependen biner atau kategorikal). Dalam regresi logistik, variabel target dianggap sebagai variabel acak dengan distribusi Bernoulli, yang hanya memiliki dua kemungkinan nilai: 0 dan 1.

Model regresi logistik menggunakan logit transformasi dari probabilitas kejadian variabel target sebagai variabel respons, dan kemudian mencari hubungan linier antara logit transformasi ini dan variabel prediktor. Model ini kemudian diestimasi dengan menggunakan metode *maximum likelihood estimation* (MLE) yang bertujuan untuk memaksimalkan probabilitas peristiwa yang diamati.

Model regresi logistik didasarkan pada fungsi logistik atau sigmoid, yang mengubah nilai dari -∞ hingga +∞ menjadi nilai antara 0 dan 1. Fungsi sigmoid digunakan untuk

menghitung probabilitas target biner (y) berdasarkan variabel prediktor (x), yaitu:

p = 1 / (1 + e^-(b0 + b1x1 + b2x2 + ... + bnxn))

di mana p adalah probabilitas target biner, e adalah bilangan konstan yang merupakan basis logaritma natural, dan b0, b1, b2, ..., bn adalah koefisien regresi yang mempengaruhi hubungan antara variabel prediktor dan target biner.

Koefisien regresi ini dapat dihitung menggunakan metode maksimum likelihood, di mana model regresi logistik yang meminimalkan kesalahan antara nilai probabilitas prediksi dan nilai aktual target biner ditemukan. Dalam prakteknya, regresi logistik sering digunakan untuk melakukan klasifikasi biner pada data, seperti prediksi apakah suatu email adalah spam atau bukan, atau apakah pelanggan akan membeli produk tertentu atau tidak.

Koefisien regresi yang dihasilkan dapat digunakan untuk menghitung prediksi probabilitas terjadinya peristiwa tersebut dengan menggunakan persamaan:

logit(p) = b0 + b1x1 + b2x2 + ... + bnxn

di mana:
p adalah probabilitas terjadinya peristiwa
b0 adalah konstanta
b1, b2, ..., bn adalah koefisien regresi untuk setiap variabel independen x1, x2, ..., xn

Untuk memperoleh hasil prediksi, persamaan di atas perlu diubah kembali ke bentuk probabilitas menggunakan rumus:

p = exp(logit(p)) / (1 + exp(logit(p)))
dengan exp() merupakan fungsi eksponensial.

Persamaan ini menunjukkan bahwa logit (yaitu, logaritma dari odds) dari probabilitas hasil yang mungkin terjadi adalah fungsi linear dari nilai-nilai prediktor. Untuk membuat prediksi, nilai-nilai prediktor dimasukkan ke dalam persamaan dan dihitung nilai logit-nya, kemudian nilai logit dikonversi kembali menjadi probabilitas menggunakan fungsi eksponensial.

## 13.3 Asumsi Analisis Regresi Logistik

Beberapa asumsi yang perlu dipenuhi dalam analisis regresi logistik adalah:

1. Hubungan linear: asumsi ini menunjukkan bahwa variabel prediktor memiliki hubungan linier dengan variabel target biner. Variabel prediktor harus memiliki hubungan linier dengan logit dari variabel respons. Ini biasanya diasumsikan dalam regresi linear dan dapat diuji dengan menggunakan plot scatter dan metode lainnya.
2. Independence: asumsi ini menunjukkan bahwa observasi dalam dataset saling independen satu sama lain. Artinya, tidak ada hubungan antara observasi yang satu dengan yang lainnya.
3. Variabel independen harus memiliki nilai yang bervariasi: Variabel independen harus memiliki variasi yang cukup besar sehingga dapat mempengaruhi variabel target biner.
4. Ketergantungan non-linier antara variabel independen: Asumsi ini menunjukkan bahwa variabel independen memiliki efek non-linear pada variabel target biner. Oleh karena itu, jika asumsi linearitas terpenuhi, akan memudahkan dalam memahami bagaimana variabel independen mempengaruhi variabel target biner.

5. Tidak ada multicollinearity: Asumsi ini menunjukkan bahwa variabel prediktor tidak saling terkait satu sama lain, sehingga memastikan bahwa koefisien regresi logistik dapat diinterpretasikan secara benar. Tidak ada multikolinearitas antara variabel prediktor. Multikolinearitas terjadi ketika dua atau lebih variabel prediktor memiliki hubungan linier yang kuat satu sama lain, dan hal ini dapat menyebabkan masalah dalam menafsirkan hasil regresi. Oleh karena itu, sebelum melakukan analisis regresi logistik, penting untuk mengevaluasi keberadaan multikolinearitas.
6. Model-fit: Asumsi ini menunjukkan bahwa model regresi logistik yang digunakan adalah model yang sesuai dengan data yang diamati.
7. Tidak ada heteroskedastisitas dalam kesalahan model. Heteroskedastisitas terjadi ketika varian kesalahan model tidak konstan di seluruh rentang nilai prediktor. Hal ini dapat diuji dengan menggunakan plot residual dan metode lainnya.
8. Tidak ada pengaruh outliers yang signifikan pada model. Outliers adalah nilai yang jauh dari nilai-nilai lain dalam sampel, dan dapat mempengaruhi hasil analisis regresi. Oleh karena itu, penting untuk mengevaluasi keberadaan outliers dan mengambil tindakan yang sesuai jika ditemukan.
9. Variabel prediktor harus tidak berkorelasi dengan kesalahan model. Korelasi antara variabel prediktor dan kesalahan model dapat menyebabkan bias dalam estimasi parameter dan dapat mempengaruhi kesimpulan yang diambil dari analisis regresi. Oleh karena itu, perlu untuk mengevaluasi korelasi antara variabel prediktor dan kesalahan model.

Jika asumsi-asumsi tersebut tidak terpenuhi, maka hasil analisis regresi logistik dapat menjadi tidak valid atau menyesatkan. Oleh karena itu, sangat penting untuk memeriksa dan memenuhi asumsi-asumsi tersebut sebelum melakukan analisis regresi logistik.

## 13.4 Jenis Analisis Regresi Logistik

Beberapa jenis analisis regresi logistik yang umum digunakan antara lain:

1. Regresi Logistik Binomial: digunakan ketika variabel dependen hanya memiliki dua kategori (biner) dan tidak ada kelompok perbandingan.
2. Regresi Logistik Multinomial: digunakan ketika variabel dependen memiliki tiga atau lebih kategori dan tidak ada kelompok perbandingan.
3. Regresi Logistik Ordinal: digunakan ketika variabel dependen memiliki tiga atau lebih kategori dan ada kelompok perbandingan yang teratur (ordinal).
4. Regresi Logistik Proporsional Hazard: digunakan ketika mempelajari pengaruh variabel independen terhadap kejadian suatu peristiwa dalam waktu tertentu.
5. Regresi Logistik Berulang: digunakan ketika mempelajari pengaruh variabel independen terhadap kejadian suatu peristiwa yang terjadi berkali-kali pada subjek yang sama.

Semua jenis analisis regresi logistik ini bertujuan untuk mempelajari pengaruh variabel independen terhadap variabel dependen dan memberikan prediksi probabilitas terjadinya peristiwa. Pilihan jenis analisis regresi logistik yang tepat tergantung pada jenis data dan tujuan penelitian. Pada pembahasan chapter ini kita akan membahas salah satu analisis regresi lofgistik yaitu regresi logistik binomial.

## 13.5 Analisis Regresi Logistik Binomial

Analisis Regresi Logistik Binomial adalah salah satu metode statistik yang digunakan untuk memprediksi probabilitas kejadian suatu peristiwa berdasarkan beberapa variabel prediktor. Metode ini sering digunakan dalam penelitian medis, sosial, ekonomi, dan bisnis. Pada analisis regresi logistik binomial, variabel terikat adalah variabel binomial, yaitu variabel yang hanya memiliki dua nilai mungkin (0 atau 1). Sedangkan variabel bebas atau prediktor dapat berupa variabel numerik maupun kategorik.

Langkah-langkah dalam melakukan analisis regresi logistik binomial adalah sebagai berikut:

1. Mengumpulkan data yang akan dianalisis.
2. Menentukan variabel terikat (*response variable*) dan variabel bebas (*predictor variable*) yang akan digunakan.
3. Melakukan uji asumsi dasar, seperti uji normalitas, homogenitas, dan independensi data.
4. Melakukan analisis regresi logistik binomial dengan menggunakan metode *Maximum Likelihood Estimation* (MLE) untuk mendapatkan model regresi yang optimal.
5. Menginterpretasikan hasil analisis regresi, yaitu koefisien regresi dan odds ratio untuk menentukan pengaruh variabel bebas terhadap variabel terikat.
6. Melakukan uji signifikansi terhadap koefisien regresi dan odds ratio untuk mengetahui apakah variabel bebas memiliki pengaruh signifikan terhadap variabel terikat.
7. Melakukan uji akurasi model dengan menggunakan confusion matrix, ROC curve, dan AUC score.
8. Memvalidasi model dengan menggunakan data uji (validation data) untuk menguji kemampuan model dalam memprediksi variabel terikat.

## 13.6 Uji Signifikansi Regresi Logistik Binomial

Uji signifikansi dalam analisis regresi logistik digunakan untuk mengevaluasi signifikansi model secara keseluruhan dan setiap variabel independen secara individual dalam memprediksi variabel dependen. Terdapat beberapa metode uji signifikansi yang umum digunakan dalam analisis regresi logistik, yaitu:

1. **Uji likelihood ratio (LR)**: Uji ini membandingkan nilai log-*likelihood* dari model yang menggunakan variabel independen dengan model yang tidak menggunakan variabel independen. Uji ini menghasilkan statistik uji chi-square dengan derajat kebebasan yang sama dengan jumlah variabel independen yang diuji. Uji *likelihood ratio* (LR) merupakan salah satu metode untuk menguji signifikansi secara keseluruhan dari analisis regresi logistik. Uji ini dilakukan dengan membandingkan model regresi logistik yang lengkap dengan model regresi logistik yang hanya terdiri dari intercept.

   Langkah-langkah untuk melakukan uji likelihood ratio (LR) adalah sebagai berikut:

   a. Hitung likelihood dari model regresi logistik yang lengkap dan model regresi logistik yang hanya terdiri dari intercept.
   b. Hitung nilai likelihood ratio (LR) dengan menghitung selisih antara log-likelihood model lengkap dan log-likelihood model yang hanya terdiri dari intercept. Nilai LR dihitung dengan rumus:

      LR = -2*(log-likelihood model intercept - log-likelihood model lengkap)

      Dimana -2 adalah konstanta yang digunakan untuk menghasilkan distribusi chi-square.
   c. Hitung derajat kebebasan (df) dari uji LR dengan mengurangi jumlah parameter dalam model lengkap

dengan jumlah parameter dalam model yang hanya terdiri dari intercept.

d. Tentukan nilai p dengan menggunakan tabel distribusi chi-square dengan derajat kebebasan yang sudah dihitung sebelumnya.
e. Jika nilai p kurang dari alpha (tingkat signifikansi yang ditentukan), maka model regresi logistik dianggap signifikan secara keseluruhan.

Jadi, uji likelihood ratio (LR) dapat digunakan untuk menguji signifikansi secara keseluruhan dari analisis regresi logistik dengan membandingkan model regresi logistik yang lengkap dengan model regresi logistik yang hanya terdiri dari intercept. Semakin tinggi nilai LR dan semakin rendah nilai p, semakin signifikan model regresi logistik tersebut.

2. **Uji Wald**: Uji ini menguji apakah koefisien regresi pada setiap variabel independen berbeda secara signifikan dari nol. Uji ini menghasilkan statistik uji z dengan nilai yang dihitung berdasarkan perbedaan antara koefisien regresi dan nilai nol, dibagi dengan standar error koefisien. Uji Wald merupakan salah satu teknik yang digunakan untuk menguji signifikansi secara individual dari setiap variabel dalam analisis regresi logistik. Uji ini dilakukan dengan menghitung nilai z-score dari koefisien regresi dan kemudian menghitung nilai chi-square dari z-score tersebut.

   Berikut adalah langkah-langkah untuk melakukan uji Wald pada analisis regresi logistik secara individual:

   a. Hitung koefisien regresi (β) dari masing-masing variabel yang digunakan dalam model regresi logistik.

b. Hitung standar error (SE) dari masing-masing koefisien regresi menggunakan rumus SE(β) = (varian residual model) / (n-1).
c. Hitung nilai z-score untuk masing-masing koefisien regresi dengan cara membagi nilai koefisien regresi dengan nilai standar error. Rumusnya adalah z = β / SE(β).
d. Hitung nilai chi-square untuk setiap variabel dengan cara mengkuadratkan nilai z-score dan membaginya dengan satu. Rumusnya adalah chi-square = z^2 / 1.
e. Hitung nilai p-value untuk setiap variabel dengan menggunakan tabel distribusi chi-square dengan derajat kebebasan satu dan tingkat signifikansi α yang ditentukan. P-value dapat dihitung dengan cara mencari nilai probabilitas pada tabel yang lebih besar dari nilai chi-square yang ditemukan.
f. Interpretasikan hasil uji Wald dengan melihat nilai p-value yang ditemukan. Jika nilai p-value kurang dari tingkat signifikansi α yang ditentukan, maka variabel tersebut dianggap signifikan secara individual dalam model regresi logistik.

## 13.7 Uji Kecocokan Model Regresi Logistik Binomial

Uji kecocokan model regresi logistik binomial digunakan untuk menguji apakah model regresi logistik binomial yang telah dibuat cocok dengan data yang ada. Untuk melakukan uji ini, dapat menggunakan rumus uji Pearson's chi-square (χ^2) atau uji likelihood ratio (G^2). Berikut adalah rumus uji kecocokan model regresi logistik binomial dengan menggunakan uji Pearson's chi-square:

$$\chi^2 = \sum [ (O_i - E_i)^2 / E_i ]$$

dimana:

$\chi^2$ adalah nilai uji Pearson's chi-square
$O_i$ adalah jumlah observasi aktual dalam kategori i
$E_i$ adalah jumlah prediksi model dalam kategori i

Berikut adalah langkah-langkah uji kecocokan model regresi logistik binomial dengan menggunakan uji Pearson's chi-square:

1. Buat model regresi logistik binomial berdasarkan data yang telah tersedia.
2. Hitung jumlah observasi yang masuk dalam masing-masing sel kategori dari model yang dihasilkan.
3. Tentukan nilai harapan untuk setiap sel kategori dengan menggunakan rumus berikut:
   nilai harapan = (jumlah baris total x jumlah kolom total) / jumlah total observasi
4. Hitung nilai Pearson's chi-square dengan menggunakan rumus berikut:
   chi-square = Σ ((jumlah observasi aktual - nilai harapan)^2 / nilai harapan)
5. Tentukan derajat kebebasan dari uji chi-square dengan rumus:
   derajat kebebasan = (jumlah baris - 1) x (jumlah kolom - 1)
6. Tentukan nilai p dengan menggunakan tabel distribusi chi-square. Carilah nilai kritis chi-square dengan derajat kebebasan yang sama dengan yang dihitung pada langkah 5. Kemudian, tentukan nilai p dengan mencari area di bawah kurva chi-square yang lebih besar dari nilai kritis yang telah dicari.

7. Lakukan pengujian hipotesis dengan membandingkan nilai p yang telah dihitung dengan nilai alpha yang telah ditetapkan. Jika nilai p kurang dari atau sama dengan alpha, maka hipotesis nol ditolak, sehingga dapat disimpulkan bahwa terdapat perbedaan signifikan antara model regresi dan data aktual. Jika nilai p lebih besar dari alpha, maka hipotesis nol diterima, sehingga dapat disimpulkan bahwa tidak terdapat perbedaan signifikan antara model regresi dan data aktual.

## 13.8 Metode Pemilihan Variabel Bebas Model Regresi Logistik Binomial

Metode pemilihan variabel bebas untuk model regresi logistik binomial dapat dilakukan dengan tiga metode, yaitu metode enter, metode forward, dan metode backward.

1. Metode Enter
   Metode Enter adalah metode yang paling sederhana, di mana semua variabel bebas dimasukkan ke dalam model regresi secara bersamaan. Metode ini cocok digunakan jika peneliti memiliki keyakinan bahwa semua variabel bebas yang terpilih benar-benar memiliki pengaruh terhadap variabel terikat.

2. Metode Forward
   Metode Forward adalah metode yang memulai dengan tidak ada variabel bebas dalam model dan kemudian secara bertahap menambahkan variabel bebas yang paling berpengaruh terhadap variabel terikat, hingga tidak ada lagi variabel bebas yang dapat ditambahkan. Metode ini lebih baik digunakan jika peneliti tidak memiliki keyakinan bahwa semua variabel bebas memiliki pengaruh terhadap variabel terikat.

3. Metode Backward
   Metode Backward adalah kebalikan dari metode forward, dimana semua variabel bebas dimasukkan ke dalam model dan kemudian secara bertahap dihapus satu per satu variabel bebas yang kurang signifikan terhadap variabel terikat, hingga tidak ada lagi variabel bebas yang dapat dihapus. Metode ini cocok digunakan jika peneliti memiliki keyakinan bahwa semua variabel bebas memiliki pengaruh terhadap variabel terikat, namun tidak yakin variabel mana yang paling signifikan.

**Contoh kasus analisis regresi binomial**

Peneliti ingin mengetahui faktor risiko stunting pada balita di kota x, untuk keperluan penelitian diambil sampel sebanyak 150 ibu yang memiliki balita secara random dengan data sebagai berikut :

**Tabel 13.1.** Faktor Risiko Stunting pada Balita di Kota X

| No | Umur Ibu | Pendapatan | Jarak Kelahiran | PMT | Pengetahuan Ibu | Stunting |
|---|---|---|---|---|---|---|
| 1 | 35 | 1 | 1 | 0 | 0 | 1 |
| 2 | 33 | 1 | 0 | 0 | 1 | 1 |
| 3 | 35 | 0 | 0 | 0 | 1 | 0 |
| 4 | 35 | 0 | 0 | 1 | 0 | 0 |
| 5 | 25 | 1 | 0 | 1 | 1 | 0 |
| 6 | 25 | 1 | 1 | 1 | 0 | 1 |
| 7 | 35 | 0 | 1 | 0 | 1 | 1 |
| 8 | 23 | 1 | 1 | 0 | 0 | 1 |
| 9 | 33 | 0 | 0 | 0 | 1 | 0 |
| .. | ... | ... | ... | ... | ... | ... |
| 147 | 35 | 0 | 1 | 0 | 1 | 1 |
| 148 | 23 | 1 | 1 | 0 | 0 | 1 |
| 149 | 33 | 0 | 0 | 0 | 1 | 0 |
| 150 | 35 | 0 | 0 | 1 | 0 | 0 |

**Keterangan :**
Pendapatan : 0 = Pendapatan Kurang, 1 = Pendapatan Cukup
Jarak Kelahiran : 0 = Risiko Tinngi, 1 = Risiko Rendah
PMT : 0 = Tidak mengkonsumsi PMT, 1 = Mengkonsumsi PMT
Pengetahuan Ibu : 0 = Pengetahuan Rendah, 1 = Pengetahuan Cukup
Stunting : 0 = Stunting, 1 = Tidak Stunting

Tujuan Penelitian :

1. Menganalisis pengaruh umur terhadap kejadian stunting pada balita di Korta X
2. Menganalisis pengaruh pendapatan terhadap kejadian stunting pada balita di Kota X
3. Menganalisis pengaruh jarak kelahiran terhadap kejadian stunting pada balita di Kota X
4. Menganalisis pengaruh PMT terhadap kejadian stunting pada balita di Kota X
5. Menganalisis pengaruh pengetahuan ibu terhadap kejadian stunting pada balita di Kota X
6. Membentuk model faktor risiko terhadap kejadian stunting pada balita di Kota X

Pada tabel tersebut, terdapat beberapa variabel independen, yaitu Umur Ibu, Pendapatan, Jarak Kelahiran, Status PMT, dan Pengetahuan Ibu. Variabel dependennya adalah Kejadian Stunting pada balita, yang nilainya hanya bisa "Stunting" atau "Tidak Stunting". Dari data tersebut, dapat dilakukan analisis regresi logistik binomial untuk mengetahui faktor-faktor yang berpengaruh pada kejadian Stunting pada Balita.

Berikut adalah langkah-langkah untuk melakukan analisis regresi logistik binomial di SPSS :

1. Buka program SPSS dan masukkan data yang akan dianalisis.
2. Pilih menu "*Analyze*" di atas layar, kemudian pilih "Regression" dan "Binary Logistic".
3. Pada jendela "*Binary Logistic Regression*", pilih variabel dependen (yaitu variabel yang ingin diprediksi) dan masukkan ke kotak "Dependent" dengan mengklik tanda panah di tengah.
4. Pilih variabel independen (yaitu variabel yang digunakan untuk memprediksi variabel dependen) dan masukkan ke kotak "Covariates" dengan mengklik tanda panah di tengah.
5. Jika ingin menambahkan interaksi antara variabel independen, pilih "Interaction" dan masukkan variabel yang ingin digunakan untuk interaksi ke kotak "Interaction".
6. Pilih "Method" untuk menentukan metode estimasi yang ingin digunakan. Pada contoh ini, gunakan metode "Enter" yang mengharuskan semua variabel independen dimasukkan ke dalam model.
7. Klik "Options" untuk menentukan opsi tambahan. Di sini, kita dapat memilih opsi seperti "Odds Ratio" untuk menampilkan rasio peluang dan "Classification Plots" untuk menampilkan grafik ROC.
8. Klik "OK" untuk menjalankan analisis. Output akan muncul di jendela yang terpisah.
9. Interpretasikan hasil analisis regresi logistik binomial yang muncul di output SPSS. Hal ini meliputi tabel koefisien regresi, tabel kesalahan standar, tabel nilai z, tabel nilai Signifikansi, tabel odds ratio, dan grafik ROC. Interpretasi dilakukan dengan memeriksa nilai-nilai signifikansi dan koefisien masing-masing variabel independen serta interpretasi odds ratio yang berkaitan dengan masing-masing variabel independen.

Setelah melakukan analisis regresi logistik binomial di SPSS, kita dapat melihat output yang terdiri dari beberapa tabel. Untuk memahami tabel output SPSS analisis regresi logistik binomial, ada beberapa informasi penting yang harus diperhatikan. Berikut adalah penjelasan untuk setiap bagian dari tabel output tersebut:

1. Tabel *Case Processing Summary* menjelaskan bahwa kasus secara keseluruhan teramati, artinya tidak terdapat satu pun data yang tidak teramati. Data keseluruhan yang teramati sebanyak 150 responden.

**Case Processing Summary**

| Unweighted Cases[a] | | N | Percent |
|---|---|---|---|
| Selected Cases | Included in Analysis | 150 | 100.0 |
| | Missing Cases | 0 | .0 |
| | Total | 150 | 100.0 |
| Unselected Cases | | 0 | .0 |
| Total | | 150 | 100.0 |

a. If weight is in effect, see classification table for the total number of cases.

2. Tabel Dependent Variable Encoding menggambarkan hasil proses input data yang telah digunakan pada variable dependen, yaitu tidak stunting diberi kode 0 dan stunting diberi kode 1.

**Dependent Variable Encoding**

| Original Value | Internal Value |
|---|---|
| Tidak Stunting | 0 |
| Stunting | 1 |

3. Table *Categorical Variables Coding* menunjukkan hasil proses input data yang digunakan pada variable dependen, yaitu tidak menderita stunting diberi kode 0 dan menderita stunting diberi kode 1.

**Categorical Variables Codings**

| | | Frequency | Parameter coding (1) |
|---|---|---|---|
| Pengetahuan Ibu | Cukup | 60 | .000 |
| | Kurang | 90 | 1.000 |
| Jarak Kelahiran | Risiko Rendah | 89 | .000 |
| | Risiko Tinggi | 61 | 1.000 |
| Pemberian Makanan Tambahan | Mengkonsumsi PMT | 65 | .000 |
| | Tidak Mengkonsumsi PMT | 85 | 1.000 |
| Pendapatan Keluarga | Cukup | 76 | .000 |
| | Kurang | 74 | 1.000 |

Tabel *Categorical Variables Coding* menjelaskan bahwa software secara otomatis melakukan perubahan kode sebagai contoh variabel pengetahuan ibu kategori cukup diberi kode 0 dan kurang diberi kode1, variabel jarak kelahiran kategori risiko rendah diberi kode 0 dan risiko tinggi diberi kode 1, variabel PMT kategori mengkonsumsi PMT diberi kode 0 dan tidak mengkonsumsi PMT diberi kode 1, dan variabel pendapatan Keluarga kategori cukup diberi kode 0 dan kurang diberi kode 1. Perubahan ini terjadi karena pada saat melakukan perintah analisis regresi logistic, kita melakukan prosedur kategorikal.

**Metode enter**

Blok 1 adalah tahap memasukkan variabel independen ke dalam model penelitian, cara memasukkan variabel independen ini menggunakan metode enter, yaitu semua variabel independen secara bersama-sama dimasukkan ke dalam model.

4. Model Summary: Bagian ini memberikan informasi tentang kualitas model secara umum. Terdapat beberapa metrik evaluasi model, termasuk Log-Likelihood, Chi-square, dan Cox & Snell R Square, Nagelkerke R Square. Semakin tinggi nilai R Square, semakin baik modelnya dalam menjelaskan varian dalam data.

**Model Summary**

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|---|---|---|---|
| 1 | 65.679[a] | .606 | .812 |

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

Nilai Nagelkerke R Square menunjukkan nilai koefisien determinasi. Didapatkan nilainya 0.878 yang artinya 87.8% pengaruh seluruh variable independen terhadap variable dependen. Dengan demikian dapat kita menfsirkan bahwa proporsi varians diagnose kejadian stunting yang bias dijelaskan oleh umur, jarak kelahiran, PMT dan pengetahuan ibu sebesar 87.8%. perlu diketahui bahwa interpretasi ini adalah nilai pendekatan saja seperti pada koefisien determinasi, karena dalam regresi logistic koefisien determinasi tidak bias dihitung seperti pada regresi linear.

5. Hosmer and Lemeshow Test merupakan uji kelayakan model, dimana hipotesisnya Ho : Model layak atau model telah cukup menjelaskan data (*Goodness of Fit*) dan H1 : Model tidak layak atau model tidak cukup menjelaskan data.

**Hosmer and Lemeshow Test**

| Step | Chi-square | df | Sig. |
|---|---|---|---|
| 1 | 3.452 | 8 | .903 |

Hasil uji menunjukkan nilai sig 0.903 yang artinya model layak atau Fit.

6. Variables in the Equation: Bagian ini memberikan informasi tentang variabel yang termasuk dalam model dan koefisien regresi logistik binomial yang terkait. Pada tabel ini terdapat dua kolom, yaitu "B" (beta) dan "SE" (standard error). Koefisien beta menunjukkan arah dan besar pengaruh variabel terhadap variabel respon, sedangkan standard error menunjukkan seberapa akurat perkiraan koefisien beta.

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) | 95.0% C.I.for EXP(B) | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Lower | Upper |
| Step 1[a] | Umur | -.423 | .086 | 24.063 | 1 | .000 | .655 | .553 | .776 |
| | Pendapatan (1) | .379 | .735 | .266 | 1 | .606 | 1.461 | .346 | 6.163 |
| | Jarak_Kelahiran(1) | 2.737 | .802 | 11.661 | 1 | .001 | 15.445 | 3.210 | 74.322 |
| | PMT(1) | 1.940 | .660 | 8.655 | 1 | .003 | 6.960 | 1.911 | 25.351 |
| | Pengetahuan_Ibu(1) | 1.700 | .776 | 4.796 | 1 | .029 | 5.474 | 1.196 | 25.062 |
| | Constant | 9.570 | 2.540 | 14.200 | 1 | .000 | 1.433E4 | | |

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) | 95.0% C.I.for EXP(B) | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Lower | Upper |
| Step 1[a] | Umur | -.423 | .086 | 24.063 | 1 | .000 | .655 | .553 | .776 |
| | Pendapatan (1) | .379 | .735 | .266 | 1 | .606 | 1.461 | .346 | 6.163 |
| | Jarak_Kelahiran(1) | 2.737 | .802 | 11.661 | 1 | .001 | 15.445 | 3.210 | 74.322 |
| | PMT(1) | 1.940 | .660 | 8.655 | 1 | .003 | 6.960 | 1.911 | 25.351 |
| | Pengetahuan_Ibu(1) | 1.700 | .776 | 4.796 | 1 | .029 | 5.474 | 1.196 | 25.062 |

a. Variable(s) entered on step 1: Umur, Pendapatan, Jarak_Kelahiran, PMT, Pengetahuan_Ibu.

Angka di kolom sig menunjukkan signifikan atau tidak pengaruh sebuah variable, sebuah variable berpengaruh terhadap stunting jika nilainya < 0.05. terlihat bahwa yang berpengaruh signifikan terhadap stunting adalah umur, jarak kelahira, PMT dan pengetahuan ibu.

7. Odds Ratio: Bagian ini menunjukkan nilai odds ratio untuk setiap variabel dalam model. Odds ratio mengukur perubahan rasio odds (yaitu probabilitas kejadian suatu kejadian dibagi dengan probabilitas tidak terjadi) ketika variabel independen meningkat satu satuan. Nilai odds ratio yang lebih besar dari 1 menunjukkan hubungan positif antara variabel independen dan variabel respon, sedangkan nilai odds ratio yang lebih kecil dari 1 menunjukkan hubungan negatif. Nilai odds ratio dilihat pada kolom Exp (B) yang menjelaskan bentuk pengaruh dari variable independen terhadap variable dependen. Jika tujuan penelitian untuk melihat besar risiko masing-masing variabel bebas, maka kriteria nilai Exp (B) ditentukan berdasarkan :

OR > 1, variabel bebas merupakan faktor risiko,
OR < 1, variabel bebas merupakan faktor protektif,
OR = 1, variabel bebas bukan merupakan faktor risiko.

a. Variabel umur memiliki nilai odds ratio 0.655 dengan nilai beta (B) -0.423 dan sig < 0.05, artinya terdapat korelasi negatif antara usia ibu dengan kejadian stunting. Semakin muda usia ibu maka semakin tinggi risiko anaknya mengalami stunting, besar risiko 0.655 yang berarti ibu usia muda lebih berisiko 0.655 kali lebih besar anaknya mengalami stunting dibanding ibu yang berusia lebih dewasa.
b. Variabel jarak kelahiran memiliki nilai odds ratio 15.445 dan sig < 0.05, jarak kelahiran yang berisiko tinggi memiliki risiko 15.445 kali lebih besar mengalami stunting dibanding jarak kelahiran risiko rendah.
c. Variabel Pemberian Makanan Tambahan (PMT) memiliki nilai odds ratio 6.960 dan sig < 0.05, artinya tidak mengkonsumsi PMT memiliki risiko 6.960 kali lebih besar mengalami stunting dibanding yang mengkonsumsi PMT.
d. Variabel Pemberian Makanan Tambahan (PMT) memiliki nilai odds ratio 5.474 dan sig < 0.05, artinya pengetahuan ibu yang rendah memiliki risiko 5.474 kali lebih besar anaknya mengalami stunting dibanding ibu yang memiliki pengetahuan cukup.

8. Wald: Bagian ini menunjukkan nilai uji Wald untuk setiap variabel dalam model. Uji Wald mengukur apakah koefisien regresi secara signifikan berbeda dari nol. Semakin besar nilai uji Wald, semakin signifikan pengaruh variabel terhadap variabel respon.
9. Confidence Interval: Bagian ini menunjukkan interval kepercayaan untuk koefisien regresi logistik binomial pada

tingkat kepercayaan tertentu (biasanya 95%). Interval kepercayaan yang lebih kecil menunjukkan bahwa kita lebih yakin bahwa koefisien regresi sebenarnya berada di dalam rentang tersebut.

Dari hasil analisis, kita dapat mengevaluasi pengaruh variabel independen pada variabel dependen. Jika nilai p-value kurang dari 0,05, maka hubungan antara variabel independen dan variabel dependen signifikan. Koefisien regresi positif menunjukkan bahwa semakin besar nilai variabel independen, semakin besar pula kemungkinan variabel dependen memiliki nilai "1" (Yes). Sedangkan koefisien regresi negatif menunjukkan hubungan sebaliknya. Kita juga dapat menggunakan nilai odds ratio untuk mengevaluasi pengaruh variabel independen pada variabel dependen. Semakin besar nilai odds ratio, semakin besar pengaruh variabel independen pada variabel dependen. Selain itu, kita juga dapat mengevaluasi seberapa baik model dalam membedakan antara nilai "1" dan "0" melalui grafik ROC.

Kesimpulan dari tujuan penelitian adalah adalah sebagai berikut :

1. Terdapat pengaruh umur terhadap kejadian stunting pada balita di Korta X
2. Tidak terdapat pengaruh pendapatan terhadap kejadian stunting pada balita di Kota X
3. Terdapat pengaruh jarak kelahiran terhadap kejadian stunting pada balita di Kota X
4. Terdapat pengaruh PMT terhadap kejadian stunting pada balita di Kota X
5. Terdapat pengaruh pengetahuan ibu terhadap kejadian stunting pada balita di Kota X

Untuk tujuan penelitian 6 (membentuk model)

Pada analisis pembentukan model menggunakan metode Forward dan Backward. Sebelum analisis multivariate dilakukan, maka terlebih dahulu variabel yang akan dimasukkan dalam model diseleksi. Variabel yang dimasukkan dalam analisis model adalah variabel yang pada analisis bivariate mempunyai nilai $\rho < 0.25$. pada contoh kasus ini untuk variabel pendapatan, jarak kelahiran, PMT, dan pengetahuan bias menggunakan uji chi square, sedangkan untuk variabel umur digunakan uji regresi logistic sederhana, dengan hasil sebagai berikut :

Variabel Umur

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) | 95.0% C.I.for EXP(B) | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Lower | Upper |
| Step 1[a] | Umur | -.439 | .064 | 46.388 | 1 | .000 | .645 | .568 | .732 |
| | Constant | 13.446 | 2.033 | 43.734 | 1 | .000 | 6.909E5 | | |

a. Variable(s) entered on step 1: Umur.

Variabel Pendapatan

**Chi-Square Tests**

| | Value | df | Asymp. Sig. (2-sided) | Exact Sig. (2-sided) | Exact Sig. (1-sided) |
|---|---|---|---|---|---|
| Pearson Chi-Square | 13.302[a] | 1 | .000 | | |
| Continuity Correction[b] | **12.128** | **1** | **.000** | | |
| Likelihood Ratio | 13.533 | 1 | .000 | | |
| Fisher's Exact Test | | | | .000 | .000 |
| Linear-by-Linear Association | 13.214 | 1 | .000 | | |
| N of Valid Cases[b] | 150 | | | | |

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 32.07.

b. Computed only for a 2x2 table

## Variabel Jarak Kelahiran

**Chi-Square Tests**

| | Value | df | Asymp. Sig. (2-sided) | Exact Sig. (2-sided) | Exact Sig. (1-sided) |
|---|---|---|---|---|---|
| Pearson Chi-Square | 31.851[a] | 1 | .000 | | |
| Continuity Correction[b] | **29.991** | **1** | **.000** | | |
| Likelihood Ratio | 34.051 | 1 | .000 | | |
| Fisher's Exact Test | | | | .000 | .000 |
| Linear-by-Linear Association | 31.639 | 1 | .000 | | |
| N of Valid Cases[b] | 150 | | | | |

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 26.87.

b. Computed only for a 2x2 table

## Variabel PMT

**Chi-Square Tests**

| | Value | df | Asymp. Sig. (2-sided) | Exact Sig. (2-sided) | Exact Sig. (1-sided) |
|---|---|---|---|---|---|
| Pearson Chi-Square | 52.705[a] | 1 | .000 | | |
| Continuity Correction[b] | **50.318** | **1** | **.000** | | |
| Likelihood Ratio | 55.823 | 1 | .000 | | |
| Fisher's Exact Test | | | | .000 | .000 |
| Linear-by-Linear Association | 52.353 | 1 | .000 | | |
| N of Valid Cases[b] | 150 | | | | |

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 28.17.

b. Computed only for a 2x2 table

Variabel Pengetahuan

**Chi-Square Tests**

| | Value | df | Asymp. Sig. (2-sided) | Exact Sig. (2-sided) | Exact Sig. (1-sided) |
|---|---|---|---|---|---|
| Pearson Chi-Square | .452[a] | 1 | .501 | | |
| Continuity Correction[b] | .255 | 1 | .614 | | |
| Likelihood Ratio | .452 | 1 | .501 | | |
| Fisher's Exact Test | | | | .507 | .307 |
| Linear-by-Linear Association | .449 | 1 | .503 | | |
| N of Valid Cases[b] | 150 | | | | |

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 26.00.

b. Computed only for a 2x2 table

Dari hasil analisis bivariate dapat disimpulkan seperti table berikut :

**Table 13.2.** Hasil Uji Bivariat

| **Variabel** | **Uji (Analisis Bivariat)** | **Nilai ρ** |
|---|---|---|
| Umur | Regresi Logistik Sederhana | 0.000 |
| Pendapatan | Chi Square | 0.000 |
| Jarak Kelahiran | Chi Square | 0.000 |
| PMT | Chi Square | 0.000 |
| Pengetahuan | Chi Square | 0.614 |

Variabel Pengetahuan dikeluarkan dari model karena nilai ρ = 0.614 > 0.25.

## Metode Forward

Metode forward adalah salah satu metode yang digunakan dalam analisis regresi logistik binomial untuk memilih variabel

yang paling berpengaruh terhadap variabel respon. Dalam metode ini, variabel-variabel independen dimasukkan ke dalam model satu per satu berdasarkan signifikansi statistik dan nilai kontribusinya terhadap model.

Proses metode forward dimulai dengan memasukkan satu variabel independen yang dianggap paling signifikan berdasarkan uji statistik, seperti uji chi-square atau uji likelihood ratio. Kemudian, variabel independen berikutnya dimasukkan ke dalam model secara berurutan satu per satu, dan setiap kali variabel baru dimasukkan, model diuji untuk melihat apakah variabel tersebut signifikan secara statistik dan meningkatkan nilai goodness of fit model.

Proses ini dilanjutkan hingga tidak ada lagi variabel yang signifikan secara statistik dan dapat meningkatkan nilai goodness of fit model. Hasil akhir dari metode forward adalah model yang terdiri dari variabel-variabel independen yang paling berpengaruh terhadap variabel respon, serta tingkat signifikansi dan koefisien regresi masing-masing variabel.

Berikut adalah langkah-langkah metode forward regresi logistik pada SPSS:

1. Variabel umur, pendapatan, jarak kelahiran, dan PMT dimasukkan dalam model
2. Buka data pada SPSS dan pilih Analyze > Regression > Binary Logistic.
3. Pilih variabel dependen dan variabel independen yang akan dimasukkan ke dalam model.
4. Pilih metode forward pada tab Selection dan tentukan kriteria signifikansi yang diinginkan, misalnya pilih significance level 0.05.
5. Klik tombol Options untuk memilih opsi output yang diinginkan. Anda dapat memilih untuk menampilkan statistik kesalahan standar, konfidensi interval, dan berbagai pengukuran kinerja model.

6. Klik tombol Run untuk menjalankan analisis regresi logistik dengan metode forward.
7. SPSS akan mengeluarkan hasil output yang mencakup tabel coefficient, tabel pengujian signifikansi, tabel odds ratio, dan berbagai pengukuran kinerja model.
8. Evaluasi model dengan melihat signifikansi variabel independen dan nilai goodness of fit model, seperti deviance atau likelihood ratio.
9. Validasi model dengan menggunakan metode-metode lain seperti validasi silang atau bootstrap.

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) | 95.0% C.I.for EXP(B) | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Lower | Upper |
| Step 1[a] | Umur | -.439 | .064 | 46.388 | 1 | .000 | .645 | .568 | .732 |
| | Constant | 13.446 | 2.033 | 43.734 | 1 | .000 | 6.909E5 | | |
| Step 2[b] | Umur | -.425 | .064 | 43.665 | 1 | .000 | .654 | .577 | .742 |
| | Jarak_Kelahiran(1) | 2.236 | .596 | 14.087 | 1 | .000 | 9.353 | 2.910 | 30.061 |
| | Constant | 11.985 | 1.908 | 39.456 | 1 | .000 | 1.603E5 | | |
| Step 3[c] | Umur | **-.389** | .068 | 32.531 | 1 | .000 | .678 | .593 | .775 |
| | Jarak_Kelahiran(1) | **2.266** | .658 | 11.867 | 1 | .001 | 9.637 | 2.655 | 34.973 |
| | PMT(1) | **-2.041** | .627 | 10.601 | 1 | .001 | .130 | .038 | .444 |
| | Constant | **11.921** | 2.090 | 32.526 | 1 | .000 | 1.504E5 | | |

a. Variable(s) entered on step 1: Umur.

b. Variable(s) entered on step 2: Jarak_Kelahiran.

c. Variable(s) entered on step 3: PMT.

Tabel *Variables in the Equation* adalah hasil proses metode *forward*, pada tahap pertama variabel yang pertama kali dipilih adalah variabel umur, karena variabel ini memiliki korelasi

terbesar terhadap kejadian stunting. Kemudian tahap kedua menyusul variabel jarak kelahiran karena variabel ini mempunyai korelasi parsial yang paling besar. Kemudian tahap ketiga dipilih variabel PMT karena variabel ini memiliki korelasi parsial yang paling besar. Proses tersebut dihentikan karena variabel-variabel lain tidak bias meningkatkan sumbangan efektif secara signifikan. Dalam hal ini model yang diperoleh adalah sebagai berikut :

$$f(z) = \frac{1}{1 + e^{-(11.921 - 0.389\,Umur + 2.266 Jarak\ Kelahiran - 2.041 PMT)}}$$

**Metode Backward**

Metode Backward dalam analisis regresi logistik binomial adalah salah satu teknik pemodelan yang digunakan untuk memilih variabel yang paling signifikan dalam model regresi logistik. Metode ini sering digunakan ketika terdapat banyak variabel independen dalam model, dan kita ingin mengevaluasi mana yang paling berpengaruh terhadap variabel dependen.

Dalam metode Backward, kita memulai dengan memasukkan semua variabel independen ke dalam model regresi logistik. Kemudian, kita menghitung koefisien regresi untuk setiap variabel independen dan menguji apakah koefisien tersebut signifikan secara statistik. Variabel yang memiliki koefisien yang tidak signifikan kemudian dihapus dari model.

Setelah variabel yang tidak signifikan dihapus, kita kembali menghitung koefisien regresi untuk setiap variabel independen dalam model yang telah diperbarui dan mengulangi proses penghapusan variabel yang tidak signifikan sampai hanya tinggal variabel yang signifikan secara statistik.

Metode Backward dapat membantu kita membangun model regresi logistik yang lebih sederhana dan lebih mudah dipahami dengan mempertahankan hanya variabel yang paling signifikan. Namun, penting untuk diingat bahwa penghapusan

variabel yang tidak signifikan dapat mengurangi jumlah informasi dalam model dan mengurangi kemampuan model untuk memprediksi dengan akurat. Oleh karena itu, metode ini harus digunakan dengan hati-hati dan selalu dilakukan dengan mempertimbangkan tujuan analisis dan konteks data yang digunakan.

Langkah-langkah untuk melakukan metode Backward dalam analisis regresi logistik

1. Buka program SPSS 16 dan buka file data yang ingin dianalisis.
2. Klik menu "Analyze" dan pilih "Regression" kemudian pilih "Binary Logistic".
3. Pilih variabel dependen (y) dan masukkan ke kotak "Dependent" dan pilih variabel independen (x) dan masukkan ke kotak "Covariates".
4. Klik tombol "Method" dan pilih "Backward: Wald".
5. Pilih level signifikansi untuk menghapus variabel yang tidak signifikan. Bila tidak yakin, nilai baku yang digunakan adalah 0,05.
6. Klik tombol "Options" untuk menentukan metode iterasi. Bila tidak yakin, nilai baku yang digunakan adalah "Enter" atau "Stepwise".
7. Klik "OK" untuk memulai analisis regresi logistik binomial.
8. Hasil analisis akan ditampilkan pada jendela output SPSS. Perhatikan tabel "Omnibus Tests of Model Coefficients" untuk mengevaluasi signifikansi model dan koefisien regresi masing-masing variabel independen pada model.
9. Perhatikan tabel "Variables in the Equation" untuk melihat variabel independen mana yang terhapus karena tidak signifikan secara statistik.
10. Evaluasi model yang dihasilkan menggunakan kriteria evaluasi seperti AIC atau BIC untuk memastikan bahwa

model yang dipilih adalah model yang terbaik secara statistik.

11. Validasi model dengan menggunakan data uji atau data yang belum pernah dilihat sebelumnya untuk memastikan bahwa model dapat menghasilkan prediksi yang akurat dan dapat diandalkan.

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) | 95.0% C.I.for EXP(B) | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Lower | Upper |
| Step 1[a] | Umur | -.368 | .070 | 27.876 | 1 | .000 | .692 | .604 | .794 |
| | Pendapatan | .582 | .697 | .698 | 1 | .404 | 1.790 | .457 | 7.015 |
| | Jarak_Kelahiran | 2.093 | .686 | 9.310 | 1 | .002 | 8.112 | 2.114 | 31.127 |
| | PMT | 2.088 | .637 | 10.737 | 1 | .001 | 8.068 | 2.314 | 28.129 |
| | Constant | 8.916 | 2.218 | 16.155 | 1 | .000 | 7.447E3 | | |
| Step 2[a] | Umur | **-.389** | .068 | 32.531 | 1 | .000 | .678 | .593 | .775 |
| | Jarak_Kelahiran | **2.266** | .658 | 11.867 | 1 | .001 | 9.637 | 2.655 | 34.973 |
| | PMT | **2.041** | .627 | 10.601 | 1 | .001 | 7.697 | 2.253 | 26.295 |
| | Constant | **9.880** | 2.023 | 23.858 | 1 | .000 | 1.954E4 | | |

a. Variable(s) entered on step 1: Umur, Pendapatan, Jarak_Kelahiran, PMT.

Tabel *Variables in the Equation* adalah proses metode *Backward Wald*, pada tahap pertama semua variabel dimasukkan kemudian menyingkirkan variabel satu per satu. Kemudian menyingkirkan variabel yang memiliki nilai ρ yang paling besar yaitu variabel pendapatan dengan nilai ρ = 0.404. tahap kedua tidak ada lagi variabel disingkirkan karena semua variabel

mempunyai nilai ρ < 0.05, sehingga model yang diperoleh sebagai berikut :

$$f(z) = \frac{1}{1 + e^{-(9.880 - 0.389\ Umur + 2.266 Jarak\ Kelahiran + 2.041 PMT)}}$$

## 13.9 Uji Konfounding

Konfounding adalah suatu distorsi (gangguan) dalam menaksirkan pengaruhpaparan terhadap kejadian penyakit/*outcome* sebagai akibat tercampurnya pengaruh sebuah atau beberapa variabel luar. Masalah ini terjadi karena pada dasarnya suadah ada perbedaan risiko terjadinya penyakit pada kelompok paparan dengan kelompok yang tidak terpapar, yang berarti risiko terjadinya penyakit pada kedua kelompok itu berbeda meskipun pajanan dihilangkan pada kedua kelompok tersebut. Secara umum, satu variabel digolongkan sebagai konfounder jika variabel tersebut merupakan faktor risiko terjadinya penyakit dan memiliki asosiasi dengan factor risiko. Jadi penilaian apakah satu variabel merupakan konfounder dilakukan dengan :

a. Pengetahuan yang sudah ada tentang hubungan variabel tersebut dengan penyakit dan pajanan pada populasi asal (*base population*)
b. Pertimbangan statistik atas adanya hubungan variabel tersebut dengan penyakit dan faktor risiko pada data penelitian.

Bila variabel yang dikeluarkantersebut mengakibatkan perubahan OR variabel-variabel yang masih ada (berubah > 10%), maka variabel tersebuta adalah confounding dan harus dimasukkan kembali ke dalam pemodelan multivariat.

$$Indeks\ Konfounder = \frac{Exp(B)Crude - Exp(B)\ Adjusted}{Exp(B)\ Adjusted} X100\%$$

Konfounder jika nilai indeks > 10%

Sesuai dengan kasus diatas sebagaicontoh, apakah variabel pendapatan merupakan konfounder?

Jawab :

1. Lakukan analisis regresi logistik dengan memasukkan variabel pendapatan dalam model dengan hasil sebagai berikut :

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | Umur | -.368 | .070 | 27.876 | 1 | .000 | .692 |
| | Pendapatan(1) | -.582 | .697 | .698 | 1 | .404 | .559 |
| | Jarak_Kelahiran(1) | 2.093 | .686 | 9.310 | 1 | .002 | 8.112 |
| | PMT(1) | -2.088 | .637 | 10.737 | 1 | .001 | .124 |
| | Constant | 11.586 | 2.048 | 32.006 | 1 | .000 | 1.075E5 |

a. Variable(s) entered on step 1: Umur, Pendapatan, Jarak_Kelahiran, PMT.

2. Lakukan analisis regresi logistik dengan mengeluarkan variabel pendapatan dalam model dengan hasil sebagai berikut :

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | Umur | -.389 | .068 | 32.531 | 1 | .000 | .678 |
| | Jarak_Kelahiran (1) | 2.266 | .658 | 11.867 | 1 | .001 | 9.637 |
| | PMT(1) | -2.041 | .627 | 10.601 | 1 | .001 | .130 |
| | Constant | 11.921 | 2.090 | 32.526 | 1 | .000 | 1.504E5 |

a. Variable(s) entered on step 1: Umur, Jarak_Kelahiran, PMT.

3. Bandingkan perubahan nilai OR masing-masing variabel pada saat memasukkan variabel pendapatan dengan ketika mengeluarkan variabel pendapatan, dengan menggunakan rumus indeks konfounder, dengan hasil sebagai berikut :

$$Indeks\ Konfounder = \frac{Exp(B)Crude - Exp(B)\ Adjusted}{Exp(B)\ Adjusted} X100\%$$

**Tabel 13.3.** Perubahan Nilai OR

| Variabel | OR Crude | OR adjusted | Perubahan OR (%) |
|---|---|---|---|
| Pendapatan | 0.559 | - | - |
| Umur | 0.692 | 0.678 | 2.065 |
| Jarak Kelahiran | 8.112 | 9.637 | 15.824 |
| PMT | 0.124 | 0.130 | 4.62 |

Karena ada nilai OR yang berubah > 10% yaitu variabel Jarak Kelahiran, maka variabel pendapatan adalah confounding dan tidak bias dikeluarkan dari pemodelan multivariat. Sehingga hasil akhir persamaan pemodelan multivariat ini adalah sebagai berikut :

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | Umur | -.368 | .070 | 27.876 | 1 | .000 | .692 |
| | Pendapatan(1) | -.582 | .697 | .698 | 1 | .404 | .559 |
| | Jarak_Kelahiran (1) | 2.093 | .686 | 9.310 | 1 | .002 | 8.112 |
| | PMT(1) | -2.088 | .637 | 10.737 | 1 | .001 | .124 |
| | Constant | 11.586 | 2.048 | 32.006 | 1 | .000 | 1.075E5 |

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | Umur | -.368 | .070 | 27.876 | 1 | .000 | .692 |
| | Pendapatan(1) | -.582 | .697 | .698 | 1 | .404 | .559 |
| | Jarak_Kelahiran (1) | 2.093 | .686 | 9.310 | 1 | .002 | 8.112 |
| | PMT(1) | -2.088 | .637 | 10.737 | 1 | .001 | .124 |
| | Constant | 11.586 | 2.048 | 32.006 | 1 | .000 | 1.075E5 |

a. Variable(s) entered on step 1: Umur, Pendapatan, Jarak_Kelahiran, PMT.

Jadi dapat disimpulkan model factor risiko terhadap kejadian stunting pada balita di Kabupaten X adalah :

$$f(z) = \frac{1}{1 + e^{-(11.921 - 0.368\,Umur - 0.582\,Pendapatan + 2.093\,Jarak\ Kelahiran - 2.088 PMT)}}$$

## 13.10 Uji Variabel Interaksi

Variabel interaksi adalah variabel yang memperhitungkan pengaruh dua variabel independen terhadap variabel dependen. Dalam analisis regresi logistik, variabel interaksi dapat digunakan untuk memahami bagaimana pengaruh dua variabel independen bekerja bersama-sama pada variabel dependen. Ada beberapa cara untuk menguji variabel interaksi dalam analisis regresi logistik, di antaranya adalah:

1. Uji signifikansi koefisien interaksi: Anda dapat menambahkan variabel interaksi ke model regresi logistik dan memeriksa apakah koefisien interaksi signifikan secara statistik. Jika koefisien interaksi signifikan, maka ini menunjukkan bahwa variabel interaksi memiliki pengaruh yang signifikan pada variabel dependen.
2. Analisis plot interaksi: Anda dapat membuat plot yang menunjukkan hubungan antara dua variabel independen

pada variabel dependen, dengan mempertimbangkan nilai variabel interaksi. Jika plot menunjukkan bahwa pengaruh kedua variabel independen berubah-ubah tergantung pada nilai variabel interaksi, maka ini menunjukkan adanya interaksi antara variabel independen tersebut.

3. Uji kesesuaian model: Anda dapat membandingkan model regresi logistik dengan variabel interaksi dengan model tanpa variabel interaksi, dan memeriksa apakah model dengan variabel interaksi lebih baik dalam menjelaskan variabel dependen. Jika model dengan variabel interaksi lebih baik, maka ini menunjukkan adanya interaksi antara variabel independen dalam mempengaruhi variabel dependen.
4. Uji multi-kolinearitas: Anda juga dapat memeriksa apakah terdapat multi-kolinearitas antara variabel independen dan variabel interaksi dalam model regresi logistik. Multi-kolinearitas dapat mempengaruhi hasil uji signifikansi koefisien dan dapat mengurangi keakuratan model regresi logistik.

Dalam menguji variabel interaksi dalam analisis regresi logistik, penting untuk mempertimbangkan asumsi-asumsi yang ada dalam analisis regresi, seperti normalitas residual, linearitas, dan homoskedastisitas. Jika asumsi-asumsi ini tidak terpenuhi, maka hasil analisis regresi logistik dapat menjadi tidak akurat.

Berikut adalah langkah-langkah uji variabel interaksi dalam analisis regresi logistik menggunakan SPSS :

1. Buka program SPSS dan buka data yang ingin dianalisis.
2. Pilih menu "Analyze" di atas layar dan pilih "Regression" dan kemudian "Binary Logistic..."
3. Pada jendela Binary Logistic Regression, pilih variabel dependen yang akan diprediksi.

4. Pilih variabel independen yang ingin dimasukkan ke dalam model regresi logistik. Pastikan variabel independen yang dipilih memiliki hubungan yang signifikan dengan variabel dependen.
5. Untuk menambahkan variabel interaksi, pilih "Interactions" di bagian bawah jendela Binary Logistic Regression.
6. Pilih variabel independen yang akan menjadi variabel interaksi dan kemudian pilih variabel independen kedua yang ingin Anda interaksikan.
7. Klik "Add" untuk menambahkan variabel interaksi ke dalam model regresi logistik.
8. Klik "OK" untuk menjalankan analisis regresi logistik.
9. SPSS akan menampilkan hasil output analisis regresi logistik termasuk koefisien interaksi dan nilai signifikansi untuk koefisien interaksi. Jika nilai signifikansi kurang dari 0,05, maka variabel interaksi signifikan secara statistik.
10. Untuk membuat plot interaksi menggunakan "Graphs" di menu utama SPSS untuk memvisualisasikan interaksi antara variabel independen dan variabel interaksi.
11. Untuk membandingkan model regresi logistik dengan variabel interaksi dan tanpa variabel interaksi, dapat menggunakan analisis kesesuaian model dengan mengklik "Model" pada jendela Binary Logistic Regression.
12. Untuk memeriksa multi-kolinearitas antara variabel independen dan variabel interaksi dalam model regresi logistik, Anda dapat menggunakan fitur "Collinearity Diagnostics" yang dapat diakses melalui jendela Binary Logistic Regression.

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | Umur | -.269 | .080 | 11.388 | 1 | .001 | .764 |
| | Pendapatan(1) | 21.470 | 16.588 | 1.675 | 1 | .196 | 2.109E9 |
| | Jarak_Kelahiran(1) | 1.890 | .743 | 6.469 | 1 | .011 | 6.620 |
| | PMT(1) | -1.496 | .694 | 4.642 | 1 | .031 | .224 |
| | Pendapatan(1) by Umur | -.654 | .493 | 1.762 | 1 | .184 | .520 |
| | Jarak_Kelahiran(1) by Pendapatan(1) | .151 | 4.786 | .001 | 1 | .975 | 1.163 |
| | PMT(1) by Pendapatan(1) | -4.164 | 4.656 | .800 | 1 | .371 | .016 |
| | Constant | 8.490 | 2.275 | 13.922 | 1 | .000 | 4.864E3 |

a. Variable(s) entered on step 1: Umur, Pendapatan, Jarak_Kelahiran, PMT, Pendapatan * Umur , Jarak_Kelahiran * Pendapatan , PMT * Pendapatan .

Tabel *Variables in the Equation* menjelaskan bahwa variabel pendapatan bukan merupakan interaksi terhadap variabel umur, Jarak Kelahiran, dan PMT, karena semua interaksi mempunyai nilai $\rho > 0.05$.

# DAFTAR PUSTAKA

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. 2013. Applied logistic regression. John Wiley & Sons.

Agresti, A. 2015. Foundations of linear and generalized linear models. John Wiley & Sons.

Kleinbaum, D. G., & Klein, M. 2010. Logistic regression: A self-learning text (3rd ed.). Springer.

Agresti, A. 2019. An introduction to categorical data analysis. John Wiley & Sons.

# BAB 14
# ANALISIS MULTIVARIAT

*Oleh Syamsu Rijal*

## 14.1 Pendahuluan

Pada tahun 1889 Galton membuat distribusi normal, metode statistik ini digunakan secara tradisional, menetapkan koefisien korelasi dan regresi linier. Fisher mengusulkan analisis varians dan analisis diskriminan, SS Wilkes mengembangkan analisis varians multivariat, dan H. Hotelling menentukan analisis komponen utama dan korelasi kanonik. Umumnya, pada paruh pertama abad ke-20 sebagian besar teori analisis multivariat telah terbentuk. Dengan perkembangan ilmu komputer, psikologi, dan metode analisis multivariat dalam banyak studi disiplin ilmu lainnya telah digunakan secara lebih luas.

Program-program seperti SAS dan SPSS, yang dulu terbatas pada penggunaan mainframe, sekarang sudah tersedia dalam berbasis Windows. Analis riset pemasaran sekarang memiliki akses ke rangkaian teknik canggih yang jauh lebih luas untuk mengeksplorasi data. Tantangannya adalah bagaimana mengetahui teknik mana yang harus dipilih, dan memahami dengan jelas kekuatan dan kelemahannya.

Analisis statistik multivariat adalah penggunaan metode statistik matematika untuk mempelajari dan memecahkan masalah teori dan metode multi-indeks. Sejak 20 tahun terakhir, teknologi aplikasi komputer dan kebutuhan mendesak untuk penelitian, teknik analisis statistik multivariat banyak digunakan dalam bidang geologi, meteorologi, hidrologi, kedokteran, industri, pertanian dan ekonomi dan banyak bidang lainnya, telah menjadi solusi praktis masalah dengan cara yang efektif. Arsitektur sistem

yang disederhanakan untuk menjelajahi kernel sistem, dapat menggunakan analisis komponen utama, analisis faktor, analisis korespondensi dan metode lainnya, sejumlah faktor di setiap variabel untuk menemukan subset informasi terbaik dari subset deskripsi yang terkandung dalam hasil multivariabel sistem dan dampak dari berbagai faktor pada sistem.

Dalam analisis multivariat, pengendalian prediksi model memiliki dua kategori. Salah satunya adalah model prediksi menggunakan regresi linier berganda, analisis regresi bertahap, analisis diskriminan atau analisis regresi bertahap dari pemodelan skrining ganda. Yang lainnya adalah model deskriptif, teknik pemodelan analisis cluster yang umum digunakan. Dalam sistem analisis multivariat, membutuhkan sifat yang sama dari hal-hal atau fenomena yang dikelompokkan bersama, untuk mengidentifikasi hubungan antara mereka dan keteraturan yang melekat, banyak penelitian sebelumnya sebagian besar menggunakan kualitatif oleh satu faktor, sehingga hasilnya tidak mencerminkan karakteristik umum. dari sistem. Misalnya klasifikasi numerik, model klasifikasi umum dibangun menggunakan analisis cluster dan teknik analisis diskriminan.

***Di bidang apa kita bisa menggunakan analisis multivariat?***

Teknik multivariat digunakan untuk mempelajari kumpulan data dalam riset konsumen dan pasar, kontrol kualitas dan jaminan kualitas, optimalisasi proses dan kontrol proses, serta penelitian dan pengembangan. Teknik ini sangat penting dalam penelitian ilmu sosial karena peneliti sosial umumnya tidak dapat menggunakan eksperimen laboratorium acak, seperti yang digunakan dalam kedokteran dan ilmu alam. Di sini teknik multivariat dapat secara statistik memperkirakan hubungan antara variabel yang berbeda, dan mengkorelasikan seberapa penting masing-masing variabel untuk hasil akhir dan di mana ada ketergantungan di antara mereka.

***Mengapa Kita menggunakan teknik multivariat?***

Karena sebagian besar analisis data mencoba menjawab pertanyaan kompleks yang melibatkan lebih dari dua variabel, pertanyaan ini sebaiknya dijawab dengan teknik statistik multivariat. Ada beberapa teknik multivariat berbeda yang dapat dipilih, berdasarkan asumsi tentang sifat data dan jenis asosiasi yang dianalisis.

Setiap teknik menguji model teoritis dari pertanyaan penelitian tentang asosiasi terhadap data yang diamati. Model teoretis didasarkan pada fakta ditambah hipotesis baru tentang hubungan yang masuk akal antara variabel. Teknik multivariat memungkinkan peneliti untuk melihat hubungan antar variabel dengan cara menyeluruh dan untuk mengukur hubungan antar variabel. Mereka dapat mengontrol hubungan antar variabel dengan menggunakan tabulasi silang, korelasi parsial dan regresi berganda, dan perkenalkan variabel lain untuk menentukan hubungan antara variabel independen dan dependen atau untuk menentukan kondisi di mana asosiasi tersebut terjadi. Ini memberikan gambaran yang jauh lebih kaya dan realistis daripada melihat satu variabel dan memberikan uji signifikansi yang kuat dibandingkan dengan teknik univariat.

***Apa kesulitan menggunakan teknik multivariat?***

Teknik multivariat rumit dan melibatkan matematika tingkat tinggi yang membutuhkan program statistik untuk menganalisis data. Program statistik ini umumnya mahal. Hasil analisis multivariat tidak selalu mudah untuk ditafsirkan dan cenderung didasarkan pada asumsi yang mungkin sulit untuk dinilai. Agar teknik multivariat memberikan hasil yang berarti, mereka membutuhkan sampel data yang besar; jika tidak, hasilnya tidak berarti karena kesalahan standar yang tinggi. Kesalahan standar menentukan seberapa yakin Anda pada hasil, dan Anda bisa lebih yakin pada hasil dari sampel besar daripada sampel kecil. Menjalankan program statistik cukup

mudah tetapi membutuhkan ahli statistik untuk memahami hasilnya. *Bagaimana memilih metode yang tepat untuk memecahkan masalah praktis?*

1. Masalah perlu diperhitungkan. Masalah dapat diintegrasikan pada berbagai metode statistik yang digunakan untuk analisis. Misalnya, model prediksi dapat didasarkan pada prinsip-prinsip biologis, ekologis, untuk menentukan model teoritis dan desain eksperimen; berdasarkan hasil pengujian, pengumpulan data pengujian; ekstraksi awal data; dan kemudian menerapkan metode analisis statistik (seperti analisis korelasi, dan analisis regresi bertahap, analisis komponen utama) untuk mempelajari korelasi antar variabel, memilih subset variabel terbaik; atas dasar ini membangun model peramalan, diagnosis akhir model dan optimasi, dan diterapkan pada produksi aktual. Analisis multivariat, dengan mempertimbangkan beberapa variabel respon dari metode analisis statistik. Isi utamanya meliputi dua vektor rata-rata pengujian hipotesis, analisis varians multivariat, analisis komponen utama, analisis faktor, analisis kluster, dan analisis terkait model.
2. Jadi, Dari beberapa metode analisis multivariat mereka memiliki kelebihan dan keterbatasan, masing-masing metode memiliki asumsi, kondisi dan persyaratan data tertentu, seperti normalitas, linieritas, dan varians yang sama dan seterusnya. Oleh karena itu, dalam penerapan metode analisis multivariat, harus dikaji pada tahap perencanaan untuk menentukan kerangka teori untuk menentukan data apa yang dikumpulkan, bagaimana mengumpulkan, dan bagaimana menganalisis data.

***Pendekatan Model Linier***

Metode analisis multivariat yang umum digunakan meliputi tiga kategori:

1. Analisis varians multivariat, analisis regresi berganda dan analisis kovarians, dikenal dengan pendekatan model linier pada penelitian untuk menentukan variabel independen dan hubungan antara variabel dependen.
2. Analisis fungsi diskriminan dan analisis politipe untuk mempelajari klasifikasi benda.
3. Analisis komponen utama, korelasi kanonik, dan analisis faktor untuk mempelajari bagaimana kombinasi faktor dengan jumlah variabel yang lebih sedikit dan bukan lebih banyak.

   Analisis varians multivariat dalam varians total sesuai dengan sumbernya (atau desain eksperimen) dibagi menjadi beberapa bagian, yaitu menguji berbagai faktor terhadap variabel dependen dan interaksi antar faktor dengan metode statistik. Misalnya, dalam analisis data desain faktorial 2 × 2, total varians dapat dibagi menjadi dua faktor milik dua kelompok variasi, interaksi antara dua faktor, dan kesalahan (yaitu, variasi dalam kelompok) dan empat bagian, maka interaksi antara variasi kelompok dan signifikansi uji F.

Keuntungan dari analisis varians multivariat dapat diuji secara bersamaan dalam satu penelitian, sejumlah faktor dengan multi level dari masing-masing variabel dependen dan interaksi antar berbagai faktor. Batasan penerapannya adalah sampel dari setiap level dari setiap faktor harus merupakan sampel acak yang independen, data pengamatan yang berulang mengikuti distribusi normal, dan varians populasinya sama.

## 14.2 Beberapa Jenis Teknik Analisis Multivariat

### 1. Analisis regresi berganda

Analisis regresi berganda digunakan untuk menilai dan menganalisis sejumlah variabel independen dengan fungsi linier variabel dependen hubungan antara metode statistik. Sebuah variabel dependen y dan variabel independen$X_1$, $X_2$, ... $X_M$adalah hubungan regresi linier:

y----1X1-... --MXM--.....

Di manaα , $\beta_1$, ... , $\beta_M$ adalah parameter yang akan diestimasi, ε adalah eksperimen variabel X acak yang error. Diperoleh melalui

1, $X_2$, ... , $X_M$ dari beberapa set data dan yang sesuai nilai y

Analisis regresi berganda memiliki kelebihan yaitu suatu fenomena dapat digambarkan secara kuantitatif antara faktorfaktor tertentu dan suatu fungsi linier. Nilai-nilai yang diketahui dari masing-masing variabel ke dalam persamaan regresi dapat diperoleh perkiraan variabel dependen (prediktor) yang secara efektif dapat memprediksi terjadinya dan perkembangan suatu fenomena. Ini dapat digunakan untuk variabel kontinu, juga dapat digunakan untuk variabel dikotomis.

Dalam teknik ini kita mempertimbangkan hubungan linear antara satu atau lebih y (variabel dependen atau respons) dan satu atau lebih x (variabel independen atau prediktor). Kita akan menggunakan model linier untuk menghubungkan y dengan x dan akan diperhatikan dengan estimasi dan pengujian parameter dalam model. Salah satu aspek yang menarik adalah memilih variabel mana yang akan dimasukkan ke dalam model jika hal ini belum diketahui. Kita dapat membedakan tiga kasus menurut jumlah variabel:

a. Regresi linier sederhana: satu y dan satu x. Misalnya, kita ingin memprediksi nilai rata-rata nilai perguruan tinggi (IPK) berdasarkan IPK SMA pelamar.

b. Regresi linier berganda: satu y dan beberapa x. Sebagai contoh, kita dapat mencoba untuk meningkatkan prediksi IPK perguruan tinggi kita dengan menggunakan lebih dari satu variabel independen, seperti IPK sekolah menengah atas, nilai ujian standar (seperti ACT atau SAT), atau peringkat sekolah menengah atas.
c. Regresi linier berganda multivariat: beberapa y dan beberapa x. Dalam ilustrasi sebelumnya, kita mungkin ingin memprediksi beberapa y (seperti jumlah tahun kuliah seseorang akan menyelesaikan IPK dalam sains, seni, dan humaniora).

**2. Analisis Regresi Logistik**

Kadang-kadang disebut sebagai "model pilihan", teknik ini merupakan variasi dari regresi berganda yang memungkinkan prediksi suatu peristiwa. Penggunaan variabel dependen nonmetrik (biasanya biner) diperbolehkan, karena tujuannya adalah untuk sampai pada penilaian probabilistik dari pilihan biner. Variabel independen dapat berupa diskrit atau kontinu. Sebuah tabel kontingensi dihasilkan, yang menunjukkan klasifikasi pengamatan, apakah peristiwa yang diamati dan diprediksi cocok. Jumlah peristiwa yang diprediksi terjadi yang benar-benar terjadi dan peristiwa yang diprediksi tidak terjadi yang sebenarnya tidak terjadi, dibagi dengan jumlah kejadian, adalah ukuran keefektifan model. Alat ini membantu memprediksi pilihan yang mungkin diambil konsumen saat diberikan alternatif.

**3. *Multivariate Analysis of Variance* ( MANOVA)**

Teknik ini menguji hubungan antara beberapa variabel independen kategori dan dua atau lebih variabel dependen metrik. Sedangkan analisis varians (ANOVA) menilai perbedaan

antar kelompok (dengan menggunakan uji T untuk 2 rata-rata dan uji F antara 3 atau lebih rata-rata), MANOVA meneliti hubungan ketergantungan antara satu set tindakan tergantung di satu set kelompok. Biasanya analisis ini digunakan dalam desain eksperimental, dan biasanya hubungan yang dihipotesiskan antara ukuran-ukuran dependen digunakan. Teknik ini sedikit berbeda karena independen variabel bersifat kategoris dan variabel dependen adalah metrik. Ukuran sampel menjadi masalah, dengan 15-20 pengamatan diperlukan per sel. Namun, terlalu banyak pengamatan per sel (lebih dari 30) dan teknik ini kehilangan signifikansi praktisnya. Ukuran sel harus kira-kira sama, dengan sel terbesar memiliki kurang dari 1,5 kali pengamatan sel terkecil. Itu karena, dalam teknik ini, normalitas variabel dependen itu penting. Kecocokan model ditentukan dengan memeriksa persamaan vektor rata-rata antar kelompok. Jika terdapat perbedaan ratarata yang signifikan, maka hipotesis nol dapat ditolak dan perbedaan perlakuan dapat ditentukan.

**4. Analisis Faktor**

Ketika ada banyak variabel dalam desain penelitian, sering membantu untuk mengurangi variabel menjadi seperangkat faktor yang lebih kecil. Ini adalah teknik independensi, di mana tidak ada variabel dependen. Sebaliknya, peneliti sedang mencari struktur yang mendasari matriks data. Idealnya, variabel independen normal dan kontinu, dengan setidaknya 3 sampai 5 variabel dimuat ke dalam sebuah faktor. Ukuran sampel harus lebih dari 50 pengamatan, dengan lebih dari 5 pengamatan per variabel.

Ada dua metode analisis faktor utama: analisis faktor umum, yang mengekstraksi faktor berdasarkan varians yang dibagi oleh faktor, dan analisis komponen utama, yang mengekstraksi faktor berdasarkan varians total faktor. Analisis faktor umum digunakan untuk mencari laten (mendasari)

faktor, sedangkan analisis komponen utama digunakan untuk menemukan jumlah variabel paling sedikit yang menjelaskan varian terbanyak. Faktor pertama yang diekstraksi menjelaskan varian paling banyak. Biasanya, faktor diekstraksi selama nilai eigen lebih besar dari 1,0 atau uji sekrup secara visual menunjukkan berapa banyak faktor yang diekstraksi. Dalam analisis faktor kami mewakili variabely1, y2, . . . , yPsebagai kombinasi linier dari beberapa variabel acak F1, F2, . . . , FM (m < p) disebut faktor. Faktor-faktor tersebut mendasari konstruksi atau variabel laten yang "menghasilkan" y. Seperti variabel aslinya, faktornya bervariasi dari individu ke individu; tetapi tidak seperti variabel, faktor tidak dapat diukur atau diamati.

Tujuan dari analisis faktor adalah untuk mengurangi redundansi antar variabel dengan menggunakan jumlah faktor yang lebih sedikit.

**5. Analisis fungsi diskriminan (deskripsi pemisahan kelompok)**

Analisis fungsi diskriminan digunakan untuk menentukan klasifikasi metode statistik individual. Prinsip dasarnya adalah: Menurut dua atau lebih sampel dari kelas data pengamatan yang diketahui untuk mengidentifikasi satu atau beberapa fungsi diskriminan linier dan indeks diskriminan, kemudian menentukan fungsi diskriminan berdasarkan indikator lain untuk menentukan termasuk dalam kategori mana seseorang. Ada dua tujuan utama dalam pemisahan kelompok:

a. Deskripsi pemisahan kelompok, di mana fungsi linier variabel (fungsi diskriminan) digunakan untuk menggambarkan atau menjelaskan perbedaan antara dua kelompok atau lebih. Tujuan analisis diskriminan deskriptif termasuk mengidentifikasi kontribusi relatif variabel p terhadap pemisahan kelompok dan

menemukan bidang optimal di mana titik-titik dapat diproyeksikan untuk menggambarkan konfigurasi terbaik dari kelompok.

b. Prediksi atau alokasi observasi ke grup, di mana fungsi linear atau kuadrat dari variabel (fungsi klasifikasi) digunakan untuk menetapkan unit sampling individu ke salah satu grup. Nilai terukur dalam vektor observasi untuk individu atau objek dievaluasi oleh fungsi klasifikasi untuk menemukan grup tempat individu tersebut paling mungkin berada.

Analisis diskriminan tidak hanya digunakan untuk variabel kontinu dan dengan teori bilangan juga bisa digunakan untuk data kualitatif, Ini membantu menentukan kriteria klasifikasi secara objektif. Namun, analisis diskriminan hanya dapat dilakukan untuk kategori yang telah diidentifikasi. Ketika kelas itu sendiri tidak pasti, kami menggunakan pemisahan awal dari kategori pertama dengan analisis diskriminan atau dengan analisis kluster.

**6. Analisis klaster :**

Dalam analisis klaster kami mencari pola dalam kumpulan data dengan mengelompokkan pengamatan (multivariat) ke dalam klaster. Tujuannya adalah untuk menemukan pengelompokan optimal yang pengamatan atau objek dalam setiap cluster serupa, tetapi cluster tidak mirip satu sama lain. Kami berharap dapat menemukan pengelompokan alami dalam data, pengelompokan yang masuk akal bagi peneliti.

Jadi, analisis cluster digunakan untuk memecahkan masalah metode klasifikasi statistik. Jika pengamatan diberikan dari n objek, setiap objek memiliki p karakteristik (variabel) yang diamati, bagaimana mereka dikelompokkan menjadi beberapa kelas yang didefinisikan? Jika objek pada

pengelompokan yang diamati, dikenal sebagai analisis Q-type; jika variabel bersama-sama kelas, yang disebut analisis tipe-R, prinsip dasar pengelompokan adalah membuat perbedaan internal yang serupa kecil, tetapi perbedaan besar antar kategori. Salah satu analisis cluster adalah metode hirarkis clustering, misalnya untuk n objek menjadi k kelas, n objek pertama menjadi kelas mereka sendiri, total n kelas, kemudian menghitung "jarak" berpasangan untuk menemukan dua kelas terdekat, ke kelas baru, lalu ulangi proses ini langkah demi langkah, hingga tanggal untuk k kelas.

Analisis klaster berbeda secara mendasar dari analisis klasifikasi. Dalam analisis klasifikasi, kami mengalokasikan pengamatan ke sejumlah kelompok atau populasi yang telah ditentukan sebelumnya. Dalam analisis klaster, baik jumlah kelompok maupun kelompok itu sendiri tidak diketahui sebelumnya. Untuk mengelompokkan pengamatan ke dalam cluster, banyak teknik yang dimulai dengan kesamaan antara semua pasangan pengamatan. Dalam banyak kasus kesamaan didasarkan pada beberapa ukuran jarak. Metode klaster lainnya menggunakan pilihan awal untuk pusat klaster atau perbandingan variabilitas dalam dan antar klaster. Dimungkinkan juga untuk mengelompokkan variabel, dalam hal kesamaan bisa menjadi korelasi.

Kita dapat mencari cluster secara grafis dengan memplot pengamatan. Jika hanya ada dua variabel ($p = 2$), kita bisa melakukan ini di sebar plot. Untuk $p > 2$, kita dapat memplot data dalam dua dimensi menggunakan komponen utama atau biplot . Analisis cluster juga telah disebut sebagai klasifikasi, pengenalan pola (khususnya, pembelajaran tanpa pengawasan), dan taksonomi numerik. Teknik analisis klaster telah banyak diterapkan pada data di berbagai bidang, seperti kedokteran, psikiatri, sosiologi, kriminologi, antropologi,

arkeologi, geologi, geografi, penginderaan jauh, riset pasar, ekonomi, dan teknik.

Ketika ukuran sampel besar, n sampel pertama dapat dibagi menjadi k kelas, dan kemudian dimodifikasi secara bertahap sesuai dengan prinsip-prinsip terbaik hingga klasifikasi sejauh ini masuk akal. Analisis cluster didasarkan pada hubungan antara individu atau jumlah variabel untuk mengklasifikasikan, objektivitas yang kuat, tetapi berbagai metode clustering hanya dapat dicapai dalam kondisi tertentu, optimum lokal, hasil akhir clustering ditetapkan, para ahli masih perlu identifikasi. Diperlukan untuk membandingkan beberapa metode yang berbeda untuk memilih yang lebih sesuai dengan persyaratan hasil klasifikasi profesional.

**7. Penskalaan Multidimensi ( MDS )**

MDS mengubah penilaian konsumen tentang kesamaan menjadi jarak yang diwakili dalam ruang multidimensi. Ini adalah pendekatan dekomposisi yang menggunakan pemetaan perseptual untuk menyajikan dimensi. Sebagai teknik eksplorasi, berguna dalam memeriksa dimensi yang tidak dikenal tentang produk dan mengungkap evaluasi komparatif produk ketika dasar perbandingan tidak diketahui.

**8. Analisis komponen utama :**

Dalam analisis komponen utama, kami berusaha untuk memaksimalkan varian dari kombinasi linier variabel. Misalnya, kita mungkin ingin mengurutkan siswa berdasarkan skor mereka pada tes prestasi dalam bahasa Inggris, matematika, membaca, dan sebagainya. Skor rata-rata akan memberikan satu skala untuk membandingkan siswa, tetapi dengan bobot mata pelajaran yang tidak sama kita dapat menyebarkan siswa lebih jauh pada skala dan mendapatkan peringkat yang lebih baik.

Pada dasarnya, analisis komponen utama adalah teknik satu sampel yang diterapkan pada data tanpa pengelompokan di antara pengamatan dan tanpa pembagian variabel ke dalam himpunan bagian y dan x . Semua kombinasi linier yang telah kita pertimbangkan sebelumnya terkait dengan variabel lain atau struktur data. Dalam regresi, kami memiliki kombinasi linear dari variabel independen yang paling baik memprediksi variabel dependen, dalam korelasi kanonik, kami memiliki kombinasi linear dari subset variabel yang secara maksimal berkorelasi dengan kombinasi linear dari subset variabel lain, dan analisis diskriminan melibatkan kombinasi linier yang secara maksimal memisahkan kelompok pengamatan. Komponen utama, di sisi lain, hanya berkaitan dengan struktur inti dari satu sampel pengamatan pada variabel p.

Analisis komponen utama dapat membantu mengidentifikasi faktor utama yang mempengaruhi variabel dependen, dan juga dapat diterapkan pada metode analisis multivariat lainnya dalam penyelesaian komponen utama tersebut setelah analisis regresi. Analisis komponen utama juga dapat berfungsi sebagai langkah pertama dalam analisis faktor.

Kelemahannya adalah hanya melibatkan sekumpulan interdependensi antar variabel, untuk membahas hubungan antar dua variabel diharuskan menggunakan korelasi kanonik.

**9. Analisis Korespondensi:**

Teknik ini memberikan pengurangan dimensi peringkat objek pada sekumpulan atribut, menghasilkan peta perseptual peringkat. Namun, seperti MDS, baik variabel independen maupun variabel dependen diperiksa pada saat yang bersamaan. Teknik ini lebih mirip dengan analisis faktor. Ini adalah teknik komposisi, dan berguna bila ada banyak atribut dan banyak perusahaan. Ini paling sering digunakan

dalam menilai efektivitas kampanye iklan. Ini juga digunakan ketika atribut terlalu mirip untuk analisis faktor menjadi bermakna. Pendekatan struktural utama adalah pengembangan tabel kontingensi. Ini berarti bahwa model dapat dinilai dengan memeriksa nilai Chi-kuadrat untuk model tersebut. Analisis korespondensi sulit untuk ditafsirkan, karena dimensinya merupakan kombinasi dari variabel independen dan dependen.

Ini adalah teknik grafis untuk merepresentasikan informasi dalam tabel kontingensi dua arah, yang berisi jumlah (frekuensi) item untuk klasifikasi silang dari dua variabel kategori. Dengan analisis korespondensi, kami membuat plot yang menunjukkan interaksi dua variabel kategori bersama dengan hubungan baris satu sama lain dan kolom satu sama lain.

**10. Analisis korelasi kanonik :**

Korelasi kanonik adalah perluasan dari korelasi berganda, yang merupakan korelasi antara satu y dan beberapa x, seringkali merupakan pelengkap yang berguna untuk analisis regresi multivariat. Ini bekerja melalui deskripsi komprehensif intervensi dari koefisien korelasi tipikal antara dua set metode statistik variabel acak multivariat.

Biarkan x menjadi variabel acak, dan biarkan y menjadi variabel acak, bagaimana menggambarkan tingkat korelasi antara mereka? Membosankan ini tidak mencerminkan sifat hal-hal. Jika kita menggunakan analisis korelasi kanonik, prosedur dasarnya adalah, dari fungsi linier dua variabel yang masing-masing mengambil salah satu dari setiap pasangan, mereka harus menjadi koefisien korelasi maksimum dari suatu pasangan, yang dikenal sebagai pasangan pertama dari variabel kanonik, demikian pula kita juga dapat menemukan dua pasangan pertama, 3 pasangan, ... , di antara pasangan

variabel yang tidak terkait ini, koefisien korelasi untuk variabel tipikal disebut koefisien korelasi kanonik. Koefisien korelasi kanonik yang dihasilkan lebih dari kumpulan variabel asli dalam nomor kumpulan apa pun variabel.

Analisis korelasi kanonik dari dua set variabel berkontribusi pada deskripsi komprehensif tentang hubungan tipikal di antara mereka. Syaratnya, kedua variabel tersebut merupakan variabel kontinu, informasinya harus mengikuti distribusi normal multivariat.

**11. Analisis gabungan:**

Analisis konjoin sering disebut "analisis trade-off", karena memungkinkan evaluasi objek dan berbagai tingkat atribut yang akan diperiksa. Ini adalah teknik komposisi dan teknik ketergantungan, di mana tingkat preferensi untuk kombinasi atribut dan level dikembangkan. Nilai bagian, atau utilitas, dihitung untuk setiap level dari setiap atribut, dan kombinasi atribut pada level tertentu dijumlahkan untuk mengembangkan preferensi keseluruhan untuk atribut di setiap level. Model dapat dibangun yang mengidentifikasi tingkat ideal dan kombinasi atribut untuk produk dan layanan.

## 14.3 Pemodelan Persamaan Struktural ( SEM) :

Berbeda dengan teknik multivariat lain yang dibahas, model persamaan struktural (SEM) memeriksa banyak hubungan antara set variabel secara bersamaan. Ini mewakili sekelompok teknik. SEM dapat menggabungkan variabel laten, yang tidak atau tidak dapat diukur secara langsung ke dalam analisis. Misalnya, tingkat kecerdasan hanya dapat disimpulkan, dengan pengukuran langsung variabel seperti nilai ujian, tingkat pendidikan, nilai rata-rata, dan ukuran terkait lainnya. Alat-alat ini sering digunakan untuk mengevaluasi banyak atribut yang diskalakan atau membangun skala yang dijumlahkan.

**Log – model linier**

Teknik ini berkaitan dengan klasifikasi data. Dimana ada variabel dependen dan satu atau lebih dari satu variabel independen. Data yang diamati diklasifikasikan dalam tabel kontingensi , Kemudian kita membuat model log linier tergantung pada data ini , dan kita menemukan nilai yang diharapkan untuk masing - masing model ini , dimana setiap model memiliki rumus untuk mencari data yang diharapkan ini setelah itu kita hitung (Pearson -2) dan (Kemungkinan – Statistik Ransum G2), untuk membandingkannya dengan tabel -2untuk mengetahui apakah model yang dipilih merepresentasikan data dengan baik atau tidak, yaitu jika dihitung nilai (-2dan G2) kurang dari tabel -2, itu berarti nilai yang dihitung adalah moral, dan model membuat representasi yang baik terhadap data, dan jika tidak (itu adalah nilai yang dihitung dari (-2dan G2) tidak kurang dari tabel -2) maka kita harus memilih model lain.

## 14.4 Kesimpulan

Setiap teknik di atas memiliki titik kekuatan dan titik kelemahan, artinya analis harus berhati-hati dalam menggunakan teknik tersebut, dimana dia harus memahami kekuatan dan kelemahan dari masing-masing teknik tersebut. Masing-masing teknik multivariat mendeskripsikan beberapa jenis data yang berbeda dengan teknik lainnya. Program statistik seperti ( spss , sas, dan lain - lain ) memudahkan untuk menjalankan suatu prosedur .

# DAFTAR PUSTAKA


Alvin C. Rencher.2002: "Multivariate Analysis Methods", A. John Wiley & son, inc. publication, Second Edition.

Garson David . 2009 :"Factor Analysis". from Statnotes, Topics in Multivariate Analysis. Retrieved April 13, from http://www2.chass.ncsu.edu/garson/pa765/statnote.htm.

Raubenheimer, JE 2004: "The procedure for selecting items to maximize the reliability and validity of the scale". South African Journal of Industrial Psychology, 30(4), pages 59–64.

Satish Chandra & Dennis Menezes. 2001: "Applications of Multivariate Analysis in International Tourism Research: Perspectives on the Marketing Strategy of NTOs" Journal of Economic and Social Research 3(1), pages 77-98 .

Yoshimasa Tsuruoka, Junichi Tsujii and Sophia Ananiadou tochastic. 2009 : "Gradient Descent Training for 1-regularized Log-linear Models with Cumulative Penalty Process 47th ACL and 4th IJCNLP AFNLP Annual Meeting", pages 477–485 , Suntek, Singapore

file:///C:/Users/Ny%20%2007/Downloads/MultivariateStatisticalAnalysis.pdf

# BAB 15
# ANALISIS DISKRIMINAN

*Oleh A Fahira Nur*

## 15.1 Pengertian Analisis Diskriminan

Analisis Diskriminan adalah sebuah teknik statistik multivariat yang digunakan untuk membedakan atau mengklasifikasikan objek atau individu ke dalam kelompok atau kategori yang berbeda berdasarkan pada beberapa variabel independen yang diberikan. Artinya analisis diskriminan merupakan teknik statistik yang digunakan untuk membedakan antara dua atau lebih kelompok dengan cara menentukan variabel mana yang paling penting dalam membedakan kelompok-kelompok tersebut. Dalam konteks statistik multivariat, Analisis Diskriminan adalah salah satu teknik yang sering digunakan untuk analisis klasifikasi.

Menurut Hair et al. (2017), analisis diskriminan merupakan suatu teknik statistik multivariat yang memungkinkan untuk membedakan antara dua atau lebih kelompok yang saling berbeda berdasarkan pada beberapa variabel prediktor yang diberikan. Johnson dan Wichern (2007) mendefinisikan Analisis Diskriminan sebagai suatu teknik multivariat yang digunakan untuk membedakan antara dua kelompok atau lebih berdasarkan pada beberapa variabel prediktor. Sedangkan menurut Agresti (2018), analisis diskriminan adalah teknik statistik yang digunakan untuk membedakan antara dua atau lebih kelompok atau kategori dengan memanfaatkan sejumlah variabel yang relevan. Tujuan dari analisis ini adalah untuk menemukan model atau fungsi matematis yang dapat membedakan antara kelompok-kelompok tersebut.

## 15.2 Tujuan Analisis Diskriminan

Tujuan utama dari analisis diskriminan adalah untuk membedakan atau mengklasifikasikan objek atau individu ke dalam kelompok atau kategori yang berbeda berdasarkan serangkaian variabel prediktor yang terukur. Dalam konteks ini, analisis diskriminan dapat membantu dalam memahami karakteristik yang membedakan antara kelompok-kelompok tersebut dan menemukan variabel-variabel yang paling penting dalam membedakan antara kelompok-kelompok tersebut. Artinya analisis diskriminan mengidentifikasi variabel-variabel yang paling berpengaruh dalam membedakan atau mengklasifikasikan objek atau individu ke dalam kelompok atau kategori yang berbeda.

Secara khusus, tujuan analisis diskriminan dapat dinyatakan sebagai berikut:

1. Mempelajari karakteristik dan perbedaan antara kelompok-kelompok yang telah ditentukan sebelumnya.
2. Membuat model matematis yang dapat membedakan antara kelompok-kelompok tersebut.
3. Mengklasifikasikan objek atau individu baru ke dalam kelompok atau kategori yang tepat berdasarkan nilai-nilai variabel prediktor yang dimilikinya.
4. Menentukan kombinasi linear dari variabel prediktor yang paling baik membedakan antara kelompok-kelompok tersebut.
5. Mengetahui pengaruh relatif masing-masing variabel prediktor dalam membedakan antara kelompok-kelompok.
6. Mengidentifikasi variabel prediktor yang paling penting dalam membedakan antara kelompok-kelompok tersebut dan dapat digunakan sebagai basis untuk pengambilan keputusan.
7. Meningkatkan pemahaman tentang perbedaan antara kelompok-kelompok yang telah ditentukan sebelumnya dan faktor-faktor yang mempengaruhinya.

Dengan demikian, analisis diskriminan dapat membantu dalam memahami hubungan antara variabel prediktor dan kelompok-kelompok yang diinginkan serta memberikan dasar bagi pengambilan keputusan dalam membedakan antara kelompok-kelompok tersebut. Dalam aplikasi praktis, analisis diskriminan dapat digunakan dalam berbagai bidang seperti bisnis, pemasaran, keuangan, kesehatan, dan sebagainya untuk mengklasifikasikan pelanggan, produk, pasar, pasien, atau objek lainnya ke dalam kelompok-kelompok yang relevan dan memahami karakteristik yang membedakan antara kelompok-kelompok tersebut.

## 15.3 Metode Pembentukan Diskriminan

### 1. Simultaneous Estimation

Metode pembentukan diskriminan pada analisis diskriminan dengan simultaneous estimation merupakan metode yang memperhitungkan korelasi antara variabel prediktor. Metode ini mencari diskriminan yang meminimalkan total kesalahan klasifikasi. Metode simultaneous estimation dapat diaplikasikan dalam metode linear maupun kuadratik.

Dalam metode pembentukan diskriminan dengan simultaneous estimation, nilai koefisien diskriminan dihitung secara bersamaan untuk seluruh variabel prediktor yang terlibat. Metode ini memperhitungkan adanya korelasi antara variabel prediktor, sehingga memperbaiki estimasi koefisien diskriminan dan membuat hasil yang lebih akurat.

Namun, metode simultaneous estimation memiliki kelemahan dalam interpretasi koefisien diskriminan. Koefisien diskriminan yang dihasilkan cenderung lebih kompleks dan sulit diinterpretasikan. Selain itu, metode ini

juga membutuhkan ukuran sampel yang besar untuk mendapatkan hasil yang akurat.

Pada dasarnya, metode simultaneous estimation dapat meningkatkan akurasi dalam membedakan atau mengklasifikasikan objek atau individu ke dalam kelompok atau kategori yang berbeda, terutama pada kasus-kasus di mana variabel prediktor saling berkorelasi. Metode Simultaneous Estimation ini memiliki keuntungan dalam mengatasi masalah multikolinearitas yang terjadi ketika terdapat korelasi tinggi antara variabel prediktor. Selain itu, metode ini juga dapat memperhitungkan semua variabel prediktor secara bersamaan, sehingga menghasilkan diskriminan yang lebih akurat dalam membedakan antara kelompok-kelompok yang berbeda. Namun, metode ini juga memiliki kelemahan, yaitu memerlukan jumlah sampel yang besar untuk menghasilkan diskriminan yang akurat.

Prosedur Simultaneous Estimation terdiri dari tiga tahap, yaitu:

a. Menghitung matriks kovarians
   Matriks kovarians digunakan untuk mengukur hubungan antara setiap pasang variabel prediktor dan variabel respons. Matriks ini digunakan untuk menghitung diskriminan.
b. Menghitung diskriminan
   Diskriminan digunakan untuk membedakan antara kelompok-kelompok yang berbeda. Diskriminan dihitung dengan mengkombinasikan variabel-variabel prediktor dengan bobot tertentu. Bobot ini ditentukan oleh analisis faktor atau analisis komponen utama.
c. Memvalidasi diskriminan
   Diskriminan yang dihasilkan kemudian divalidasi dengan menggunakan sampel yang berbeda. Validasi dapat dilakukan dengan menggunakan teknik cross-

validation atau dengan membagi sampel menjadi dua bagian, yaitu data pelatihan dan data uji.

*Simultaneous Estimation* = Semua variabel dimasukkan secara bersama-sama kemudian dilakukan proses diskriminan.

### *2. Step-Wises Estimation*

Metode pembentukan diskriminan pada analisis diskriminan juga dapat dilakukan dengan menggunakan metode *Step-Wise Estimation*. Metode ini merupakan metode yang paling umum digunakan dalam analisis diskriminan karena dapat memilih variabel-variabel prediktor yang paling signifikan secara statistik.

Prosedur *Step-Wise Estimation* terdiri dari beberapa tahap, yaitu:

a. Seleksi variabel prediktor
   Variabel prediktor yang akan digunakan dalam analisis diskriminan dipilih berdasarkan asumsi yang telah ditentukan, seperti signifikansi statistik dan relevansi teoritis.
b. Pembentukan model awal
   Model awal dibentuk dengan menggunakan satu atau beberapa variabel prediktor. Model awal ini digunakan sebagai dasar untuk membangun model diskriminan yang lebih kompleks.
c. Seleksi variabel secara bertahap
   Variabel prediktor yang memiliki kontribusi yang signifikan dalam membedakan antara kelompok-kelompok yang berbeda ditambahkan ke dalam model diskriminan secara bertahap. Variabel yang tidak signifikan secara statistik kemudian dihapus dari model.

d. Pengujian model
   Model diskriminan yang telah dibentuk kemudian diuji dengan menggunakan sampel yang berbeda untuk mengukur kemampuannya dalam membedakan antara kelompok-kelompok yang berbeda.

Metode *Step-Wise Estimation* ini memiliki keuntungan dalam memilih variabel-variabel prediktor yang paling signifikan secara statistik, sehingga menghasilkan model diskriminan yang lebih sederhana dan mudah diinterpretasikan. Namun, metode ini juga memiliki kelemahan, yaitu tidak memperhitungkan variabel prediktor secara bersamaan, sehingga dapat mengabaikan interaksi antara variabel prediktor. Selain itu, metode ini juga dapat menghasilkan model diskriminan yang tidak stabil karena dapat dipengaruhi oleh sampel yang digunakan.

*Step-Wise Estimation* = variabel yang signifikan membentuk fungsi diskriminan dimasukkan satu per satu ke dalam model, yang tidak signifikan tidak masuk dalam model.

## 15.4 Model Analisis Diskriminan

Model persamaan matematika Analisis Diskriminan terdiri dari beberapa komponen utama, yaitu:

1. Skor diskriminan (D) : Skor diskriminan adalah ukuran seberapa dekat suatu observasi dengan kelompok tertentu. Skor ini diperoleh dari gabungan nilai-nilai variabel independen yang dimiliki oleh suatu observasi dan koefisien regresi dari masing-masing variabel independen.
2. Konstanta (a): Konstanta adalah nilai rata-rata dari skor diskriminan untuk kelompok referensi. Konstanta ini digunakan untuk menormalkan skor diskriminan yang

diperoleh, sehingga dapat dibandingkan dengan skor diskriminan dari kelompok lain.

3. Koefisien regresi (b) : Koefisien regresi adalah bobot yang diberikan pada setiap variabel independen. Koefisien ini menunjukkan seberapa besar pengaruh variabel independen terhadap perbedaan antara kelompok. Koefisien regresi ini diperoleh dari proses estimasi model dengan menggunakan teknik Analisis Diskriminan.
4. Variabel independen (X) : Variabel independen adalah variabel yang digunakan untuk membedakan antara kelompok-kelompok. Variabel ini dapat berupa variabel numerik maupun kategorik.

Dengan menggabungkan keempat komponen tersebut, maka model persamaan matematika Analisis Diskriminan dapat dituliskan sebagai berikut:

$$D = a + b1X1 + b2X2 + ... + bkXk$$

Di mana:
D adalah skor diskriminan
a adalah konstanta
b1, b2, ..., bk adalah koefisien regresi
X1, X2, ..., Xk adalah nilai-nilai variabel independen untuk suatu observasi.

Yang diestimasi adalah koefisien 'b' sehingga nilai 'D' setiap grup sedapat mungkin berbeda. Ini terjadi pada saat rasio jumlahkuadrat antar grup (*between group sum of squares*) terhadap jumlah kuadrat dalam grup (*within-group sum of square*) untuk skor disriminan memncapai maksimum. Berdasarkan nilai D keanggotaan seseorang dapat diprediksi.

Model persamaan matematika Analisis Diskriminan digunakan untuk memprediksi keanggotaan kelas suatu observasi berdasarkan nilai-nilai variabel independen yang dimilikinya. Untuk melakukan klasifikasi, skor diskriminan yang diperoleh dari model persamaan dibandingkan dengan nilai ambang batas yang ditentukan sebelumnya. Jika skor diskriminan lebih besar dari nilai ambang batas, observasi tersebut diklasifikasikan ke dalam kelompok tertentu, dan sebaliknya.

## 15.5 Asumsi Analisis Diskriminan

Analisis diskriminan memiliki beberapa asumsi dan syarat yang harus dipenuhi agar hasil analisis dapat diandalkan. Berikut adalah beberapa asumsi dan syarat analisis diskriminan menurut para ahli (Tabachnick, B. G., & Fidell, L. S, 2013; dan Johnson, R. A., & Wichern, D. W, 2007):

1. Normalitas: Data yang digunakan dalam analisis diskriminan harus memiliki distribusi normal. Asumsi normalitas ini penting untuk menghasilkan nilai yang akurat pada uji statistik yang digunakan dalam analisis.
2. Homogenitas kovarians: Kovarians dari variabel prediktor harus sama antara kelompok-kelompok yang dibandingkan. Asumsi homogenitas kovarians ini penting untuk memastikan bahwa variabel prediktor memberikan pengaruh yang sama untuk setiap kelompok. Artinya bersifat Homoskedastisitas, yaitu variasi dari variabel prediktor dalam setiap kelompok harus sama.
3. Independensi: Data yang digunakan dalam analisis diskriminan harus bersifat independen. Artinya, setiap subjek atau unit pengamatan tidak mempengaruhi satu sama lain.
4. Tidak ada multikolinieritas: Multikolinieritas dapat terjadi jika variabel prediktor saling berkorelasi dengan tingkat

yang tinggi. Hal ini dapat diuji dengan menggunakan uji korelasi.

5. Ketergantungan linear antara variabel prediktor dan variabel respon. Ini berarti bahwa hubungan antara variabel prediktor dan variabel respon harus bersifat linier.
6. Jumlah sampel: Jumlah sampel dalam setiap kelompok yang dibandingkan harus memadai. Jumlah sampel yang kurang dari 20 dapat mempengaruhi akurasi hasil analisis. Jumlah subjek dalam setiap kelompok harus seimbang. Jika jumlah subjek dalam setiap kelompok tidak seimbang, maka hasil analisis diskriminan dapat dipengaruhi oleh kelompok dengan jumlah subjek yang lebih besar.
7. Jumlah variabel prediktor: Jumlah variabel prediktor harus lebih sedikit dibandingkan jumlah unit pengamatan. Jika variabel prediktor terlalu banyak dibandingkan jumlah unit pengamatan, maka analisis diskriminan dapat mengalami masalah overfitting atau kelebihan penyesuaian model.

## 15.6 Berbagai Istilah dalam Analisis Diskriminan

Dalam analisis diskriminan, ada beberapa istilah yang perlu dipahami, di antaranya:

1. Kelompok referensi (*reference group*): Kelompok yang digunakan sebagai patokan dalam mengklasifikasikan data ke dalam kelompok yang berbeda.
2. Kelompok perbandingan (*comparison group*): Kelompok yang dibandingkan dengan kelompok referensi untuk melihat perbedaan dalam karakteristik atau variabel yang diukur.
3. Fungsi diskriminan (*discriminant function*): Suatu persamaan matematis yang digunakan untuk membedakan kelompok-kelompok yang berbeda berdasarkan pada variabel yang diukur.

4. Analisis diskriminan linear (*linear discriminant analysis*): Metode analisis diskriminan yang menggunakan fungsi linier untuk membedakan kelompok-kelompok yang berbeda.
5. Analisis diskriminan non-linear (*nonlinear discriminant analysis*): Metode analisis diskriminan yang menggunakan fungsi non-linier untuk membedakan kelompok-kelompok yang berbeda.
6. Variabel diskriminan (*discriminant variable*): Variabel yang digunakan dalam fungsi diskriminan untuk membedakan kelompok-kelompok yang berbeda.
7. Variabel atau prediktor: Merupakan variabel-variabel yang digunakan untuk membedakan atau memisahkan antara kelompok atau kelas dalam analisis diskriminan.
8. Koefisien diskriminan (*discriminant coefficient*): Koefisien yang muncul dalam fungsi diskriminan dan digunakan untuk memberikan bobot pada variabel diskriminan. Fungsi diskriminan Merupakan fungsi matematis yang digunakan untuk membedakan antara kelompok atau kelas dalam analisis diskriminan.
9. Analisis diskriminan parsial (*partial discriminant analysis*): Metode analisis diskriminan yang digunakan untuk menguji pengaruh setiap variabel terhadap pemisahan kelompok.
10. Analisis diskriminan kanonik (*canonical discriminant analysis*): Metode analisis diskriminan yang digunakan untuk membedakan kelompok-kelompok yang lebih dari dua dan menggunakan beberapa variabel diskriminan untuk membedakan kelompok tersebut.
11. Analisis diskriminan berganda (*multiple discriminant analysis*): Metode analisis diskriminan yang digunakan untuk membedakan kelompok-kelompok yang lebih dari dua dan menggunakan beberapa variabel diskriminan

untuk membedakan kelompok tersebut. Namun, metode ini memperhitungkan korelasi antara variabel diskriminan.

12. Error rate atau tingkat kesalahan: Merupakan persentase dari observasi yang salah diklasifikasikan dalam analisis diskriminan.
13. Validitas diskriminan: Merupakan tingkat keberhasilan analisis diskriminan dalam membedakan antara kelompok atau kelas. Artinya Ukuran sejauh mana fungsi diskriminan dapat membedakan antara kelompok diskriminan. Validitas diskriminan dapat diukur dengan menggunakan uji statistik yang sesuai. Contoh: Uji F dapat digunakan untuk menguji validitas diskriminan dalam analisis kinerja karyawan.
14. Variabel Non-Diskriminan: Variabel yang tidak digunakan untuk membedakan antara kelompok diskriminan. Variabel ini dapat digunakan untuk memperjelas analisis atau untuk menjelaskan hasil analisis. Contoh: Variabel non-diskriminan dalam analisis kinerja karyawan dapat berupa jenis kelamin, status pernikahan, dan lain sebagainya.
15. Matriks Konfusi: Tabel yang digunakan untuk memeriksa keakuratan fungsi diskriminan. Matriks konfusi menunjukkan jumlah pengamatan yang diklasifikasikan dengan benar dan yang salah oleh fungsi diskriminan. Contoh: Matriks konfusi dapat digunakan dalam analisis kinerja karyawan untuk memeriksa seberapa baik fungsi diskriminan membedakan antara karyawan efektif dan tidak efektif.
16. Kategori Referensi: Kelompok atau kategori yang digunakan sebagai dasar perbandingan dengan kelompok atau kategori lain dalam analisis diskriminan. Contoh: Dalam analisis diskriminan kinerja karyawan, kategori referensi dapat berupa karyawan yang efektif.

## 15.7 Langkah-Langkah Analisis Diskriminan

Langkah-langkah dalam analisis diskriminan terdiri dari :

1. Membuat fungsi diskriminan atau kombinasi linier dari variabel independen yang dapat mendiskriminasi (membedakan) kategori variabel dependen (kriteria kelompok) sehingga suatu objek masuk kategori/kelompok mana.
2. Menguji apakah ada perbedaan yang signifikan antar kategori/kelompok dikaitkan dengan variabel independen.
3. Menentukan variabel independen yang memberikan sumbangan terbesar dalam membedakan antar kelompok.
4. Mengklasifikasi atau mengelompokkan suatu objek ke dalam kelompok/kategori berdasarkan nilai variabel independen.
5. Mengevaluasi keakuratan klasifikasi.

## 15.8 Jenis Data Analisis diskriminan

Ada dua jenis data dalam analisis diskriminan, yaitu:

Dependen variabel menggunakan data kategori: data kategori adalah data yang memuat variabel dengan nilai yang terbatas pada kategori atau kelas tertentu. Contohnya, jenis kelamin, agama, pekerjaan, atau jenis pendidikan. Data kategori dapat dianalisis menggunakan Analisis Diskriminan Multinomial.

Independen variabel menggunakan data numerik: data numerik adalah data yang memuat variabel dengan nilai yang terukur atau kontinu, seperti tinggi badan, berat badan, atau usia. Data numerik dapat dianalisis menggunakan Analisis Diskriminan Linier atau Analisis Diskriminan Kuadratik, tergantung pada kecocokan model yang dipilih.

## 15.9 Contoh Analisis Diskriminan

Seorang peneliti ingin mengetahui faktor yang menyebabkan terjadinya stunting di Desa A Kabupaten X. Kelompok diskriminan dibagi menjadi dua: kelompok stunting dan kelompok tidak stunting. Variabel yang dianggap menyebabkan terjadinya stunting adalah : usia ibu, pendapatan keluarga, jumlah anggota keluarga, status gizi ibu, berat badan lahir, dan panjang badan lahir. Sampel diambil secara random 30 anak yang mengalami stunting dan 30 anak yang tidak stunting, dengan data sebagai berikut :

**Tabel 15.1.** Faktor risiko stunting di Desa A Kabupaten X.

| No | Usia Ibu | Pendapatan Keluarga | Jumlah Anggota Keluarga | Status Gizi Ibu (BB/U) | Berat Badan Lahir | Panjang Badan Lahir | Kejadian Stunting |
|---|---|---|---|---|---|---|---|
| 1 | 20 | 1400 | 3 | 20 | 2400 | 43 | 1 |
| 2 | 19 | 1700 | 3 | 21 | 2300 | 45 | 1 |
| 3 | 23 | 2300 | 4 | 23 | 2500 | 46 | 1 |
| 4 | 23 | 1800 | 4 | 23 | 2300 | 44 | 1 |
| 5 | 22 | 1700 | 3 | 20 | 2600 | 43 | 1 |
| 6 | 24 | 2300 | 3 | 21 | 2500 | 43 | 1 |
| 7 | 25 | 2100 | 4 | 20 | 2300 | 45 | 1 |
| 8 | 24 | 2000 | 3 | 23 | 2400 | 45 | 1 |
| 9 | 23 | 2500 | 3 | 21 | 2500 | 42 | 1 |
| 10 | 26 | 2500 | 5 | 23 | 2300 | 43 | 1 |
| 11 | 20 | 1400 | 3 | 20 | 2400 | 43 | 1 |
| 12 | 19 | 1700 | 3 | 21 | 2300 | 45 | 1 |
| 13 | 23 | 2300 | 4 | 22 | 2500 | 46 | 1 |
| 14 | 23 | 1800 | 4 | 21 | 2300 | 44 | 1 |
| 15 | 22 | 1700 | 3 | 20 | 2600 | 43 | 1 |
| 16 | 20 | 1400 | 3 | 20 | 2400 | 43 | 1 |
| 17 | 19 | 1700 | 3 | 21 | 2300 | 45 | 1 |
| 18 | 23 | 2300 | 4 | 23 | 2500 | 46 | 1 |
| .. | .. | .. | .. | .. | .. | .. | .. |
| .. | .. | .. | .. | .. | .. | .. | .. |
| .. | .. | .. | .. | .. | .. | .. | .. |
| 53 | 27 | 3400 | 3 | 23 | 2800 | 48 | 2 |
| 54 | 30 | 3000 | 3 | 24 | 2900 | 47 | 2 |

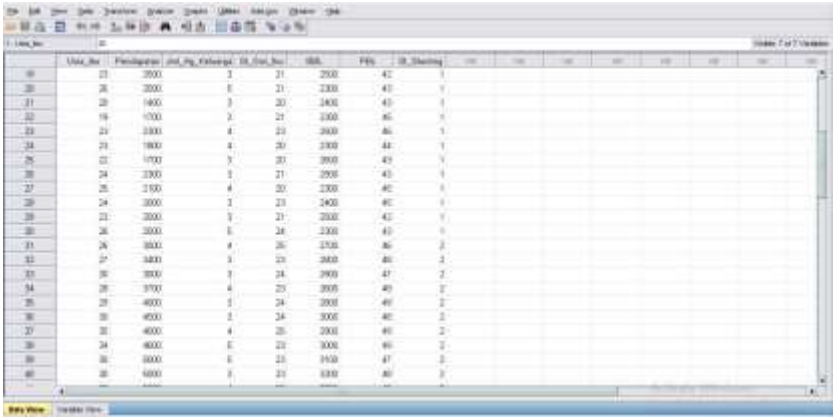
| No | Usia Ibu | Pendapatan Keluarga | Jumlah Anggota Keluarga | Status Gizi Ibu (BB/U) | Berat Badan Lahir | Panjang Badan Lahir | Kejadian Stunting |
|---|---|---|---|---|---|---|---|
| 55 | 28 | 3700 | 4 | 23 | 2600 | 48 | 2 |
| 56 | 29 | 4000 | 3 | 24 | 2800 | 49 | 2 |
| 57 | 30 | 4500 | 3 | 24 | 3000 | 48 | 2 |
| 58 | 35 | 4000 | 4 | 25 | 2900 | 49 | 2 |
| 59 | 34 | 4600 | 5 | 23 | 3000 | 49 | 2 |
| 60 | 30 | 5000 | 5 | 23 | 3100 | 47 | 2 |

Keterangan : Kejadian Stunting = 1 : Stunting, 2 : Tidak Stunting

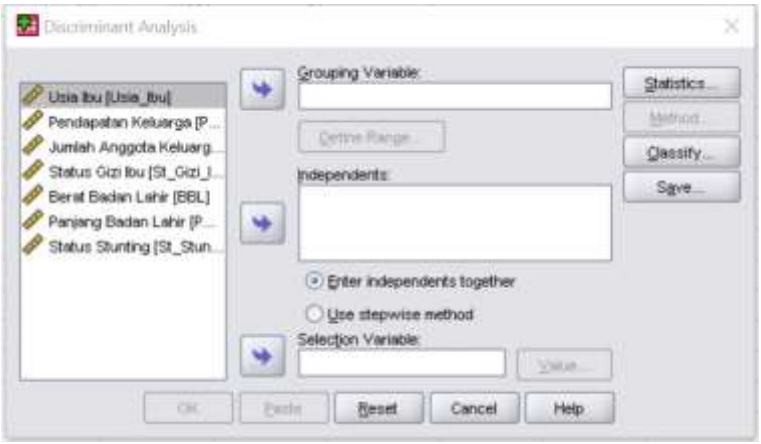
**Jawab :**

Langkah-langkah menyelesaikan contoh soal diatas dengan program SPSS

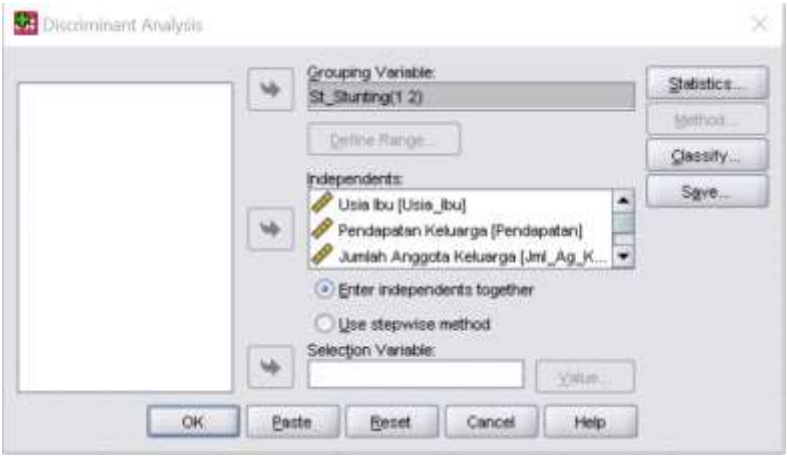
1. Buka File data\Diskriminan.sav akan muncul seperti berikut :



2. Klik ***Analyze*** kemudian klik ***Classify***
3. Klik ***Diskriminant***... akan muncul seperti berikut :

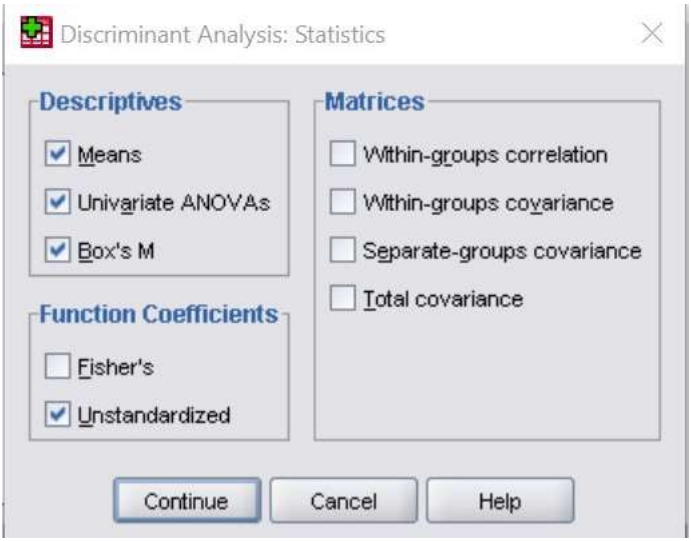



4. Masukkan variabel usia ibu, pendapatan keluarga, jumlah anggota keluarga, status gizi ibu, berat badan lahir, panjang badan lahir ke kotak ***Independents*** dan variabel status stunting ke kotak ***Grouping Variable*** akan muncul seperti berikut :

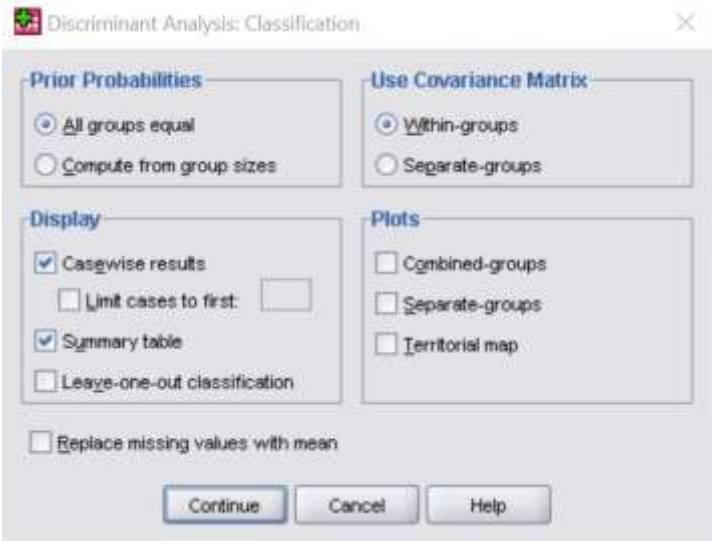


5. Klik ***Define Range***, isikan angka 1 dan 2, Klik ***Continue***

6. Klik ***Statistics***, klik ***Means, Univariate Anova, Box's M, Unstandardized***, kemudian klik ***Continue*** seperti berikut :




7. Klik ***Classify***, klik ***Casewise results***, ***summary table***




8. Klik ***Continue*** dan klik ***OK*** akan muncul hasil sebagai berikut :

**Tests of Equality of Group Means**

| | Wilks' Lambda | F | df1 | df2 | Sig. |
|---|---|---|---|---|---|
| Usia Ibu | .313 | 127.102 | 1 | 58 | .000 |
| Pendapatan Keluarga | .242 | 181.800 | 1 | 58 | .000 |
| Jumlah Anggota Keluarga | .981 | 1.094 | 1 | 58 | .300 |
| Status Gizi Ibu | .419 | 80.281 | 1 | 58 | .000 |
| Berat Badan Lahir | .254 | 170.242 | 1 | 58 | .000 |
| Panjang Badan Lahir | .229 | 194.958 | 1 | 58 | .000 |

Tabel *Tests of Equality of Group Means* mengidentifikasikan variabel apa yang signifikan membedakan antara kedua kelompok yaitu stunting dan tidak stunting. Dalam hal ini menggunakan uji f, hasil analisis diperoleh dari 6 variabel independen hanya 5 variabel

yang signifikan yaitu : usia ibu, pendapatan keluarga, status gizi ibu, berat badan lahir, dan panjang badan lahir karena masing-masing mempunyai nilai ρ < 0.05. sedangkan variabel jumlah anggota keluarga tidak signifikan karena karena mempunyai nilai ρ > 0.05. Maka dari hasil ini dilakukan analisis kedua dengan mengeluarkan variabel jumlah anggota keluarga dengan cara yang sama pada analisis pertama dengan hasil sebagai berikut :

**Tests of Equality of Group Means**

| | Wilks' Lambda | F | df1 | df2 | Sig. |
|---|---|---|---|---|---|
| Usia Ibu | .313 | 127.102 | 1 | 58 | .000 |
| Pendapatan Keluarga | .242 | 181.800 | 1 | 58 | .000 |
| Status Gizi Ibu | .419 | 80.281 | 1 | 58 | .000 |
| Berat Badan Lahir | .254 | 170.242 | 1 | 58 | .000 |
| Panjang Badan Lahir | .229 | 194.958 | 1 | 58 | .000 |

Tabel *Tests of Equality of Group Means* mengidentifikasikan bahwa dari 5 variabel independen sudah signifikan yaitu : usia ibu, pendapatan keluarga, status gizi ibu, berat badan lahir dan panjang badan lahir karena karena masing-masing mempunyai nilai ρ < 0.05. jadi ada 5 variabel yang dapat dipakai dalam model diskriminan untuk menentukan objek kedalam kelompok stunting dan tidak stunting.

**Test Results**

| | | |
|---|---|---|
| Box's M | | 82.396 |
| F | Approx. | 4.982 |
| | df1 | 15 |
| | df2 | 1.354E4 |
| | Sig. | .080 |

Tests null hypothesis of equal population covariance matrices.

Tabel *Tes Results* menunjukkan homogenitas varian antara kedua kelompok (stunting dan tidak stunting), dengan menggunakan uji statistic *Box's M.*
Ho : Varian antara dua kelompok sama
Ha : Varian antara dua kelompok tidak sama

Dari hasil analisis diperoleh nilai ρ = 0.080 > 0.05, maka dapat disimpulkan bahwa varians antara kedua kelompok (stunting dan tidak stunting) homogen.

**Structure Matrix**

| | Function |
|---|---|
| | 1 |
| Panjang Badan Lahir | .633 |
| Pendapatan Keluarga | .612 |
| Berat Badan Lahir | .592 |
| Usia Ibu | .511 |
| Status Gizi Ibu | .406 |

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions
Variables ordered by absolute size of correlation within function.

Tabel *Structure Matrix* menunjukkan urutan variabel yang paling membedakan antara kedua kelompok kejadian stunting. Variabel panjang badan lahir adalah variabel yang paling membedakan antara kedua kelompok kejadian stunting.

**Canonical Discriminant Function Coefficients**

| | Function |
|---|---|
| | 1 |
| Usia Ibu | -.011 |
| Pendapatan Keluarga | .000 |
| Status Gizi Ibu | .423 |
| Berat Badan Lahir | .004 |
| Panjang Badan Lahir | .520 |
| (Constant) | -45.225 |

Unstandardized coefficients

Tabel *Canonical Discriminant Function Coefficients* menjelaskan model diskriminan yang terbentuk, yaitu :

D = a + b1X1 + b2X2 + ... + bkXk
D = -45.225 – 0.011 Usia Ibu + 0.000 Pendapatan Keluarga + 0.423 Status Gizi Ibu + 0.004 Berat Badan Lahir + 0.520 Panjang Badan Lahir.

Model ini digunakan untuk mengklasifikasi suatu objek kedalam (kelompok stunting atau tidak stunting).

**Functions at Group Centroids**

| | Function |
|---|---|
| Status Stunting | 1 |
| Stunting | -2.846 |
| Tidak Stunting | 2.846 |

Unstandardized canonical discriminant functions evaluated at group means

Tabel *Functions at Group Centroids* menentukan *critical cutting score* untuk mengklasifikasi setiap responden dengan rumus sebagai berikut :

$$CU = \frac{N1C1 + N2C2}{N1 + N2}$$

Dimana :
CU = Critical Cutting Score
N1 = Besar sampel pada kelompok 1
N2 = Besar sampel pada kelompok 2
C1 = Nilai centroid kelompok 1
C2 = Nilai centroid kelompok 2

Jadi nilai CU contoh 1 dapat diperoleh sebagai berikut :

$$CU = \frac{30\ x\ (-2.846) + 30\ x\ 2.846}{30 + 30} = 0$$

Angka dibawah CU akan dimasukkan ke dalam kelompok stunting, sedangkan angka diatas sama dengan CU dimasukkan dalam kelompok tidak stunting.

**Classification Results[a]**

| | | | Predicted Group Membership | | |
|---|---|---|---|---|---|
| | | Status Stunting | Stunting | Tidak Stunting | Total |
| Original | Count | Stunting | 30 | 0 | 30 |
| | | Tidak Stunting | 0 | 30 | 30 |
| | % | Stunting | 100.0 | .0 | 100.0 |
| | | Tidak Stunting | .0 | 100.0 | 100.0 |

a. 100.0% of original grouped cases correctly classified.

Tabel *Classification Results* menggambarkan krostabulasi antara model awal (original) dengan pengklasifikasian hasil model diskriminan (*Predicted Group Membership*). Terlihat tidak ada responden yang salah klasifikasi. Secara keseluruhan model diskriminan yang terbentuk mempunyai tingkat validasi yang sangat tinggi, yaitu 100%.

# DAFTAR PUSTAKA

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. 2017. Multivariate Data Analysis, 8th Edition. Cengage Learning.

Agresti, A. 2018. An introduction to categorical data analysis (3rd ed.). Wiley.

Tabachnick, B. G., & Fidell, L. S. 2013. Using multivariate statistics (6th ed.). Boston: Pearson Education.

Johnson, R. A., & Wichern, D. W. 2007. Applied multivariate statistical analysis (6th ed.). Upper Saddle River, NJ: Pearson Education.

## BIODATA PENULIS

**DR. Hj. KMT Lasmiatun. SE.,MSi.**
**CRA.,C.RP.,CMA.,C.NNLP.,CM.NNLP.,C.LMA.,CHA.,CFHA.,CBHCM.,**
**CHCBP.,CBHRM.,CPSP.,CBPA.,CRSP.,CSEM.,C.MPdl**
Dosen Program Studi Manajemen
Fakultas Ekonomi Universitas Muhammadiyah Semarang

Penulis lahir di Blora tanggal 24 April 1976. Penulis adalah dosen tetap pada Program Studi Studi Manajemen Fakultas Ekonomi Universitas Muhammadiyah Semarang. Menyelesaikan pendidikan Diploma di UNNES Tahun 1997, Sarjana Ekonomi di Undip Tahun 2002, menyelesaikan Pra MM di Undip Tahun 2005 dan Magister Sains di Undip Tahun 2007. Penulis menyelesaikan program doktor pada Program Studi Pembangunan di UKSW Tahun 2015. Bekerja di dunia pendidikan sejak 1997, lalu pindah di perbankan tahun 2004 hingga 2007, dan akhirnya kembali lagi ke dunia pendidikan. Pernah menjadi Dosen di Akademi Widya Buana, UNISFAT, STIE Dharma Putra Semarang, UPBJJ-UT Semarang, Stebank Mr. Sjafruddin Prawiranegara Jakarta, Universitas Mercu Buana Jakarta, Universitas Wahid Hasyim Semarang, saat ini sebagai dosen di Universitas Muhammadiyah Semarang dan Direktur LPSDM RA Kartini serta Ketua LPP UKM.

## BIODATA PENULIS

**Ns. Solehudin, S.Kep, M.Kes., M.Kep**
Dosen Program Studi Keperawatan
Universitas Indonesia Maju Jakarta

Lahir 7 Maret 1975 di Kota Majalengka, Jawa Barat. Penulis menempuh pendidikan di Sekolah Perawat Kesehatan Depkes Cirebon, Diploma III Keperawatan di Poltekkes Bogor, memperoleh gelar Sarjana Keperawatan dari Program Studi Keperawatan Sekolah Tinggi Ilmu Kesehatan Indonesia Maju Jakarta pada tahun 2013, gelar Magister Kesehatan dari Program Studi Kesehatan Sekolah Tinggi Ilmu Kesehatan Indonesia Maju Jakarta pada tahun 2017, gelar Magister Keperawatan dari Program Studi Keperawatan Universitas Muhammadiyah Jakarta pada tahun 2021. Penulis pernah bekerja sebagai perawat di RSUD Cideres Majalengka tahun 1994–1995, di RS PMI Bogor tahun 1995-2020, dosen di Program Studi Keperawatan Sekolah Tinggi Ilmu Kesehatan Wijaya Husada Bogor sejak tahun 2017–2020. Sejak Oktober tahun 2020 bekerja sebagai dosen di Program Studi Keperawatan Universitas Indonesia Maju Jakarta.

## BIODATA PENULIS

**Dr. Maitri Anindita, S.ked., Sp.M.**
Dosen Program Studi Pendidikan Profesi Dokter Fakultas Kedokteran Universitas Airlangga

Penulis adalah dosen tetap pada Program Studi Pendidikan Profesi Dokter Fakultas Kedokteran, Universitas Airlangga Surabaya. Penulis menyelesaikan pendidikan S1 pada Program Studi Pendidikan Dokter Universitas Airlangga, Program Pendidikan Profesi Dokter Universitas Airlangga, dan Pendidikan Dokter Spesialis pada Program Studi Ilmu Kesehatan Mata Universitas Airlangga. Selain sebagai dosen, penulis juga aktif berprofesi sebagai dokter di Rumah Sakit Universitas Airlangga.

## BIODATA PENULIS

**Rusydi Fauzan, SE, MM**
Dosen Prodi Manajemen Bisnis Syariah
Fakultas Ekonomi dan Bisnis Islam
UIN Sjech M. Djamil Djambek Bukittinggi

Penulis lahir di Lubuk Aur tanggal 28 Mei 1986. Penulis merupakan dosen tetap Prodi Manajemen Bisnis Syariah UIN SMDD Bukittinggi. Penulis sudah menulis sejak tahun 2010. Penulis menyukai kegiatan membaca, menulis, dan *traveling*. Seputar kegiatan penulis dapat di follow pada akun instagram @rusydifauzan.

## BIODATA PENULIS

**Suwito Pomalingo, S.Kom., M.Kom.**
Dosen Program Studi Informatika
Fakultas Teknik dan Informatika
Universitas Multimedia Nusantara

Penulis adalah dosen tetap pada Program Studi Informatika Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Menyelesaikan pendidikan S1 pada Jurusan Teknik Informatika dan melanjutkan S2 pada Program Magister Informatika Universitas Islam Indonesia.

Saat ini sementara menyelesaikan studi Doktoral dengan topik penelitian di bidang Security Science dengan pendekatan Deep Learning. Penulis menekuni bidang keilmuan meliputi Cyber Security, Malware Analytics, Data Science, Machine Learning, dan Deep Learning, selain itu, penulis juga memiliki beberapa serifikasi internasional di bidang Cyber Security, Digital Forensics, Data Science dan Machine Learning.

## BIODATA PENULIS

**apt. Herda Ariyani, M.Farm.**
Dosen Program Studi S1 Farmasi
Fakultas Farmasi Universitas Muhammmadiyah Banjarmasin

Penulis lahir di Banjarmasin tanggal 29 Oktober 1990. Sejak tahun 2016, penulis adalah dosen tetap pada Program Studi S1 Farmasi Fakultas Farmasi, Universitas Muhammadiyah Banjarmasin. Penulis juga merupakan Kabid Pengabdian kepada Masyarakat dan Kuliah kerja Nyata LPPM Universitas Muhammadiyah Banjarmasin. Penulis menyelesaikan pendidikan SMK Farmasi di SMF-ISFI Banjarmasin Angkatan 41 pada tahun 2008, kemudian melanjutkan S1 pada Jurusan Farmasi Universitas Lambung Mangkurat dan melanjutkan profesi apoteker dan S2 pada Jurusan Farmasi Universitas Ahmad Dahlan Yogyakarta. Penulis menekuni bidang Menulis Farmasi Klinis dan Komunitas. Hingga saat ini penulis aktif dalam forum Perkumpulan Relawan Jurnal Indonesia Korwil Kalimantan, Forum Jurnal Asosiasi Perguruan Tinggi Muhammadiyah Aisyiyah, Asosiasi Penerbit Perguruan Tinggi (APPTI) Kalimantan dan Pengurus Daerah Ikatan Apoteker Indonesia (IAI) Kalimantan Selatan serta komunitas PERIISAI. Penulis telah mempublikasikan lebih dari 50 artikel baik di jurnal dan prosiding nasional maupun terindeks Scopus seperti

Taylor and Francis dengan *h-index* = 6. Saat ini penulis tercatat sebagai Managing Editor *Journal of Current Pharmaceutical Sciences, Editor Borneo Community Development Journal* dan *Editor Healthy-Mu Journal.* Reviewer Jurnal *Health Media,* reviewer Jurnal Kajian Ilmiah Kesehatan dan Teknologi serta reviewer *Pharmaceutical and Traditional Medicine Journal.*

## BIODATA PENULIS

**Ryryn Suryaman Prana Putra, SKM, M.Kes.**
Dosen Tetap Program Studi S1 Adminsitrasi Rumah Sakit
Institut Ilmu Kesehatan Pelamonia Makassar

Penulis adalah anak pertama dari bapak Jamaluddin dan Mastang. Pada tahun 2014 menyelesaikan Pendidikan Strata 2 di Prodi Kesehatan Masyarakat Departemen Administrasi dan Kebijakan Kesehatan Universitas Hasanuddin Makassar. Pada tahun 2011 sampai 2015 penulis menjadi asisten dosen, staf administrasi, dan pengelola jurnal artikel ilmiah di Fakultas Kesehatan Masyarakat Departemen Administrasi dan Kebijakan Kesehatan Universitas Hasanuddin Makassar. Sejak tahun 2015 sampai sekarang penulis bekerja sebagai Dosen Tetap pada Program Studi S-1 Administrasi Rumah Sakit Institut Ilmu Kesehatan Pelamonia Makassar dan sejak tahun 2021 sampai sekarang menjabat sebagai Ketua Bidang Penelitian pada Lembaga Penelitian dan Pengabdian kepada Masyarakat Institut Ilmu Kesehatan Pelamonia Kesdam XIV Hasanuddin Makassar. Buku yang telah ditulis oleh penulis adalah Buku Monograf "Pengaruh Penerapan Pelayanan Prima terhadap Kepuasan Pasien Rawat Jalan di RSUD Labuang Baji Makassar", Buku Monograf "Analisis GAP Harapan dan Kenyataan Pasien JKN terhadap Kualitas

Pelayanan Rawat Jalan dengan Metode SERVQUAL dan *Importance Performance Analysis* di RSUD H. Padjonga Daeng Ngalle Kabupaten Takalar", Buku Manajemen Pemasaran, Buku Manajemen Rumah Sakit (Teori Dan Aplikasi), Buku Metodologi Penelitian Kuantitatif Dan Kualitatif, Buku Hukum Kesehatan dan Sumber Daya Manusia Kesehatan, Buku Hukum Kesehatan, Buku Ekonomi Kesehatan, Buku Gizi Kronis pada Anak Stunting, Buku Penelitian Ilmu Kesehatan, Buku Kebijakan Kesehatan, serta Buku Administrasi dan Kebijakan Kesehatan.

## BIODATA PENULIS

**Putri Rahmadani, SKM, MKM.**
Dosen Program Studi Kesehatan Masyarakat
Fakultas Kesehatan Universitas Fort De Kock Bukittinggi

Penulis lahir di Bukittinggi tanggal 7 Agustus 1997. Penulis adalah dosen tetap pada Program Studi Kesehatan Masyarakat Fakultas Kesehatan, Universitas Fort De Kock Bukittinggi. Menyelesaikan pendidikan S1 pada Jurusan Ilmu Kesehatan Masyarakat Peminatan Epidemiologi dan Biostatistika Universitas Andalas dan melanjutkan S2 pada Jurusan Kesehatan Masyarakat Peminatan Biostatistika Universitas Indonesia.

## BIODATA PENULIS

**Dr. Joni Wilson Sitopu, M.Pd.**
Dosen FKIP Universitas Simalungun

Dr. Joni Wilson Sitopu, M.Pd, lahir di Medan, Pendidikan S1, S2 dan S3 Prodi Pendidikan Matematika dan Matematika FMIPA USU Medan. Dosen di FKIP dan SPS Program S2 serta Kepala Pusat Komputer Universitas Simalungun. Penulis aktif mengikuti sebagai pemakalah pada Seminar Nasional dan internasional, aktif mengikuti Seminar Nasional dan internasional dan aktif menulis Buku, Jurnal OJS, Sinta, dan Scopus.

## BIODATA PENULIS

**Sediyanto,ST,MM**
Dosen di Universitas Mercu Buana

Sediyanto,ST,MM adalah salah satu dosen di Universitas Mercu Buana. Ia menyelesaikan Pendidikan sarjana program studi teknik kimia di Institut Teknologi Sepuluh Nopember tahun 1995. Program Magister Manajemen diselesaikannya di Universitas Budi Luhur tahun 2008. Penulis sudah menulis beberapa artikel ilmiah yang telah diterbitkan dalam jurnal nasional.

## BIODATA PENULIS

**Nova Suryani, S.P., M.P.**
Dosen Program Studi Agribisnis
Universitas Adzkia

Penulis lahir di Kabupaten Lima Puluh Kota, Sumatera Barat, tanggal 19 November 1996. Penulis adalah dosen pada Program Studi Agribisnis, Universitas Adzkia, Padang. Menyelesaikan pendidikan S1 pada Jurusan Sosial Ekonomi Pertanian (Agribisnis) di Universitas Andalas dan melanjutkan S2 pada Jurusan Ekonomi Pertanian di Universitas Andalas.

## BIODATA PENULIS

**Adhar Arifuddin**
Dosen Program Studi Ilmu Kesehatan Masyarakat
Universitas Tadulako

Penulis lahir di Ujung Pandang tanggal 09 November 1982. Penulis adalah dosen tetap pada Program Studi Ilmu Kesehatan Masyarakat, Universitas Tadulako. Menyelesaikan pendidikan S1 Kesehatan Masyarakat Jurusan Epidemiologi pada tahun 2005. Pada tahun 2009 menyelesaikan pendidikan S2 pada Program Studi Ilmu Kesehatan Masyarakat Jurusan Epidemiologi, dan Tahun 2021 melanjutkan studi S3 di Jurusan Penelitian dan Evaluasi Pendidikan, Konsentrasi Metode Penelitian.

## BIODATA PENULIS

**Syamsu Rijal, S.E., M.Si., Ph.D,**
Dosen tetap pada Jurusan llmu Ekonomi, Fakultas Ekonorni dan Bisnis, Universitas Negeri Makassar (UNM) di Makassar Provinsi Sulawesi Selatan, Indonesia

Syamsu Rijal, S.E., M.Si., Ph.D, Penulis lahir di Ujung Pandang, 26 Desember 1973.di Provinsi Sulawesi Selatan, Indonesia. Dia Merupakan lulus sarjana Pada Program Studi Manajemen, Univenitas Hasanuddin (UNHAS) di Makassar Provinsi Sulawesi Selatan,Indonesia Tahun 1999, kemudian melanjutkan studi Magister pada Program Studi Ekonomi Sumber daya Program Pascasarjana, Universitas Hasanuddin Di Makassar Provinsi Sulawesi Selatan, Indonesia, dan lulus pada Tahun 2005. Pada Tahun 2017 beliau mendapatkan gelar Ph.D sebagai lulusan pada bidang Ilmu Ekonomi Publik, Central China Normal University (CCNU) di Wuhan Tiongkok. Sekarang îni Dia tergabung dalam dosen tetap pada Jurusan llmu Ekonomi, Fakultas Ekonorni dan Bisnis, Universitas Negeri Makassar (UNM) di Makassar Provinsi Sulawesi Selatan, Indonesia. Dia banyak menulis buku dan karya ilmiah yang berhubungan dengan Ilmu ekonomi khususnya Ekonomi Publik, Ekonomi Sumber Daya dan Manajemen Publik.

*Email : syamsurijalasnur@unm.ac.id*

## BIODATA PENULIS

**A Fahira Nur**
Dosen Program Studi DIII Kebidanan
Universitas Widya Nusantara

Penulis lahir di Sinjai tanggal 22 November 1988. Penulis adalah dosen tetap pada Program Studi DIII Kebidanan, Universitas Widya Nusantara. Menyelesaikan pendidikan Diploma III Kebidanan pada tahun 2010 dan melanjutkan pada Program DIV Bidan Pendidik tahun 2012. Tahun 2017 menyelesaikan pendidikan S2 pada Program Ilmu Kesehatan Masyarakat Jurusan Kesehatan Reproduksi dan tahun 2022 sampai sekarang melanjutkan Program S3 Ilmu Sosial Konsentrasi Perencanaan Kesehatan Sosial.