Editor : Paryono, S.Kep, Ns, M.Kes

Dr. Ir. Hj. Marhawati, M.Si. - Dr. Ramlan Mahmud, M.Pd - Nurdiana, S.P., M.Si
Dr. Sri Astuty SE, M.Si - Dodiet Aditya Setyawan, SKM.,MPH. - dr Prasaja STrKes., M.Kes
Nova Fahradina, M. Pd - Dr. La One ST, MT. - Rahma Faelasofi, S.Si., M.Sc.
Tri Widyasari, M.Pd - Risy Mawardati, M.Pd - Dr. Lian G. Otaya, M.Pd - Siti Rahmatina, M.Pd

STATISTIKA TERAPAN

STATISTIKA TERAPAN

Dr. Ir. Hj. Marhawati, M.Si.
Dr. Ramlan Mahmud, M.Pd
Nurdiana, S.P., M.Si
Dr. Sri Astuty SE, M.Si
Dodiet Aditya Setyawan, SKM.,MPH.
dr Prasaja STrKes., M.Kes
Nova Fahradina, M. Pd
Dr. La One ST, MT.
Rahma Faelasofi, S.Si., M.Sc.
Tri Widyasari, M.Pd
Risy Mawardati, M.Pd
Dr. Lian G. Otaya, M.Pd
Siti Rahmatina, M.Pd

Tahta Media Group

UU No 28 tahun 2014 tentang Hak Cipta

Fungsi dan sifat hak cipta Pasal 4

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

Pembatasan Pelindungan Pasal 26

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

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

Sanksi Pelanggaran Pasal 113

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

STATISTIKA TERAPAN

Penulis:
Dr. Ir. Hj. Marhawati, M.Si. | Dr. Ramlan Mahmud, M.Pd
Nurdiana, S.P., M.Si | Dr. Sri Astuty SE, M.Si
Dodiet Aditya Setyawan, SKM.,MPH.
dr Prasaja STrKes., M.Kes | Nova Fahradina, M. Pd
Dr. La One ST, MT. | Rahma Faelasofi, S.Si., M.Sc.
Tri Widyasari, M.Pd | Risy Mawardati, M.Pd
Dr. Lian G. Otaya, M.Pd | Siti Rahmatina, M.Pd

Desain Cover:
Tahta Media

Editor:
Paryono, S.Kep, Ns, M.Kes

Proofreader:
Tahta Media

Ukuran:
viii, 237, Uk: 15,5 x 23 cm

ISBN: 978-623-5981-77-2

Cetakan Pertama:
Mei 2022

PENERBIT TAHTA MEDIA GROUP
(Grup Penerbitan CV TAHTA MEDIA GROUP)
Anggota IKAPI (216/JTE/2021)

KATA PENGANTAR

Puji syukur kami panjatkan kehadirat Allah SWT, karena berkat rahmat dan karuniaNya Buku Kolaborasi dalam bentuk *Book Chapter* ini dapat dipublikasikan diharapkan sampai ke hadapan pembaca. *Book Chapter* ini ditulis oleh sejumlah Dosen dan Praktisi dari berbagai Institusi sesuai dengan kepakarannya serta dari berbagai wilayah di Indonesia.

Terbitnya buku ini diharapkan dapat memberi kontribusi yang positif dalam ilmu pengetahuan dan tentunya memberikan nuansa yang berbeda dengan buku lain yang sejenis serta saling menyempurnakan pada setiap pembahasannya yaitu dari segi Konsep yang tertuang sehingga mudah untuk dipahami. Sistematika buku yang berjudul "Statistika Terapan" terdiri dari 13 Bab yang dijelaskan secara terperinci sebagai berikut:

Bab 1 Pengertian Statistik dan Statistika
Bab 2 Teknik Pengumpulan Data
Bab 3 Ukuran Pemusatan
Bab 4 Ukuran Penyebaran Data
Bab 5 Distribusi Frekuensi
Bab 6 Pengertian Populasi dan Sampel
Bab 7 Distribusi Proporsi Sampling
Bab 8 Kesalahan Sampling dan Non Sampling
Bab 9 Uji Normalitas
Bab 10 Uji Homogenitas
Bab 11 Uji Rata – Rata dan Proporsi
Bab 12 Analisis Regresi Linear Sederhana
Bab 13 Korelasi dan Analisis Varians Satu Arah

Akhirnya kami mengucapkan terima kasih kepada semua pihak yang mendukung penyusunan dan penerbitan buku ini. Semoga buku ini dapat bermanfaat bagi pembaca sekalian.

Direktur Tahta Media
Dr. Uswatun Khasanah, M.Pd.I., CPHCEP

DAFTAR ISI

BAB 1
PENGERTIAN STATISTIK DAN STATISTIKA

Dr. Ir. Hj. Marhawati, M.Si.
Universitas Negeri Makassar

A. SEJARAH STATISTIKA

Ilmu statistika memiliki sejarah yang sangat panjang bersamaan peradaban manusia. Pada saat sebelum Masehi, bangsa-bangsa di Mesopotamia (Babilonia), Mesir, serta Tiongkok telah mengumpulkan informasi statistic untuk memperoleh data tentang besarnya pajak yang wajib dibayar oleh tiap warga, banyaknya produksi pertanian yang dapat dihasilkan, serta lain sebagainya. Pemakaian kata statistika berakar dari sebutan dalam bahasa latin moderen statisticum collegium (dewan negeri) serta bahasa Italia statista (negarawan ataupun politikus). Gottfried Achenwall (1749) memakai statistik dalam bahasa Jerman pada awalnya untuk nama untuk aktivitas analisis data kenegaraan, dengan mengartikannya sebagai ilmu tentang negeri (state).

Pada saat abad ke- 19 sudah terjadi perubahan arti menjadi " ilmu tentang pengumpulan serta klasifikasi informasi". Sir John Sinclair memperkenalkan nama (Statistics) serta penafsiran ini ke dalam bahasa Inggris. Jadi, statistika secara prinsip awalnya hanya mengurus data yang digunakan lembaga- lembaga administratif serta pemerintahan. Pengumpulan informasi terus berlangsung, terutama lewat sensus yang dikerjakan dengan tertib agar dapat memberikan data kependudukan yang berganti setiap waktu.

Penggunan Statistik telah ada sewaktu sebelum abad ke- 18, pada saat itu negara Babilon, Mesir, serta Roma menghasilkan catatan tentang nama, umur, tipe kelamin, pekerjaan, serta jumlah anggota keluarga. Setelah itu pada tahun 1500, pemerintahan Inggris menghasilkan catatan setiap minggu tentang kematian serta tahun 1662 dikembangkan catatan tentang kelahiran serta kematian. Pada awalnya Statistika di temukan oleh Aristoteles dalam bukunya yang bertajuk "politea", dalam bukunya Aristoteles menjelaskan informasi tentang kondisi 158 negara yang disebut sebagai statistika.

Pada abad ke- 17 di Inggris, statistika di sebut sebagai political aritmatic. Tahun 1772- 1791 G. Achenwall memakai sebutan statistik sebagai kumpulan informasi tentang Negara. Tahun 1791- 1799, Dokter. E. A. W Zimmesman mengenalkan kata statistika dalam bukunya Statistical Account of Scotland. Pada abad ke- 18, sebutan statistika dipopulerkan oleh Sir John Sinclair dalam bukunya bertajuk "statistical account of Scotland (1791-1799)", setelah terlebih dulu dikemukakan oleh seseorang pakar hitung asal Jerman yang

bernama Gottfried Achenwell (1719- 1772). Tahun 1880, F. Galton awal kali memakai korelasi dalam riset ilmu hayat. Pada abad 19 Karl Pearson mempelopori pemakaian metoda statistik dalam berbagai bidang riset biologi ataupun pemecahan masalah tentang sosio ekonomi. Tahun 1918- 1935, R. Fisher mengenalkan analisa varians dalam literatur statistiknya.

Pada abad ke-19 dan awal abad ke-20 statistika mulai banyak menggunakan bidang-bidang dalam matematika, terutama peluang. Cabang statistika yang pada saat ini sangat luas digunakan untuk mendukung metode ilmiah, statistika inferensi, dikembangkan pada paruh kedua abad ke-19 dan awal abad ke-20 oleh Ronald Fisher (peletak dasar statistika inferensi), Karl Pearson (metode regresi linear), dan William Sealey Gosset (meneliti problem sampel berukuran kecil). Penggunaan statistika pada masa sekarang dapat dikatakan telah menyentuh semua bidang ilmu pengetahuan, mulai dari astronomi hingga linguistika. Bidang-bidang ekonomi, biologi dan cabang-cabang terapannya, serta psikologi banyak dipengaruhi oleh statistika dalam metodologinya. Akibatnya lahirlah ilmu-ilmu gabungan seperti ekonometrika, biometrika (atau biostatistika), dan psikometrika. Meskipun ada pihak yang menganggap statistika sebagai cabang dari matematika, tetapi sebagian pihak lainnya menganggap statistika sebagai bidang yang banyak terkait dengan matematika melihat dari sejarah dan aplikasinya. Di Indonesia, kajian statistika sebagian besar masuk dalam fakultas matematika dan ilmu pengetahuan alam, baik di dalam departemen tersendiri maupun tergabung dengan matematika.

B. PENGERTIAN STATISTIK DAN STATISTIKA

Statistika berkembang sesuai dengan perkembangan ilmu pengetahuan dan kebutuhan penting masyarakat, karena memiliki hubungan dan fungsi berbagai hal dalam kehidupan manusia. Dalam kehidupan sehari-hari, kita sering dihadapkan pada berbagai masalah yang membutuhkan solusi yang cermat dan tepat. Penyelesaian dapat dilakukan dengan benar dan akurat jika memiliki informasi yang cukup tentang masalah yang akan dipecahkan dan analisisnya juga akurat. Dalam penulisan laporan penelitian atau hasil pengolahan data, istilah statistika dan statistika sering ditemukan dan digunakan disana. Kebanyakan orang tidak dapat membedakan antara statistik

dan statistik. Kedua kata ini terkait dan biasanya digunakan untuk menyelesaikan masalah terkait data.

1. **Statistik**

   Pada dasarnya antara statistik dan statistika itu sama yaitu berkaitan dengan permasalahan data, namun yang membedakan adalah pengertian dasar statistik dan statistika. Pengertian statistik menurut beberapa ahli antara lain:

   a. **Anderson dan Bancroft**

      Menurut Bancroft dan Anderson, statistik memiliki pengertian sebagai sebuah cabang ilmu sekaligus seni pengembangan dengan metode paling efektif lewat pengumpulan, penginterpretasian, dan pentabulasi data kuantitatif yang dibuat sedemikian rupa. Dengan data statistik, memungkinkan terjadinya kesalahan saat menyimpulkan dan mengestimasikan bisa diperkirakan melalui cara penalaran berdasarkan hitungan peluang.

   b. **Dr Sudjana**

      Prof. Dr. Sudjana, M.A, M.Sc juga berpendapat bahwa ilmu statistik memiliki arti yaitu pengetahuan yang berkaitan dengan pengumpulan, pengolahan analisa data, hingga penarikan kesimpulan menurut data-data yang dikumpulkan beserta analisa yang telah dilakukan.

   c. **Steel dan Torrie**

      Steel dan Torrie menjelaskan bahwa statistik ialah sebuah metode yang menjelaskan sebuah cara untuk melakukan penilaian ketidaktentuan terhadap penarikan kesimpulan bersifat induktif.

   d. **Supranto**

      Ahli Indonesia, J. Supranto pun berpendapat bahwa pengertian statistik ialah sebuah ilmu yang di dalamnya mempelajari tentang berbagai cara pengumpulan data, penyajian, lalu penganalisa dan terakhir ditariklah kesimpulan menurut hasil penelitian yang sudah dilakukan secara menyeluruh.

   e. **Croxton dan Cowden**

      Menurut Croxton dan Cowden statistik adalah metode untuk mengumpulkan, mengolah dan menyajikan serta menginter-pretasikan data yang berwujud angka.

f. Sutrisno Hadi

Sutrisno Hadi mengatakan statistik kegiatan ilmiah untuk mengumpulkan, menyusun, meringkas dan menyajikan data penyelidikan. Selanjutnya data diolah dan menarik kesimpulan secara teliti serta membuat keputusan yang logik dari hasil pengolahan data. (batasan umum). Statistik digunakan untuk menunjuk angka-angka pencatatan dari suatu kejadian atau kasus tertentu (batasan khusus).

Konsep statistika (statistika) adalah kumpulan informasi yang dikumpulkan atau disajikan dalam bentuk daftar atau gambar yang mewakili atau menggambarkan sesuatu. Sebahagian statistik sederhana dalam bentuk grafik atau diagram yang dapat digunakan sebagai dokumen analitis dengan membandingkan pola operasi pemrosesan yang berbeda, pelaporan, dll. Untuk tujuan pelaporan, grafik merupakan bentuk penyajian yang jauh lebih mudah dipahami/diterima oleh pemangku kepentingan.

Kata statistik berasal dari bahasa latin state yang berarti negara. Awalnya, statistik didefinisikan sebagai informasi yang diperlukan untuk negara dan berguna untuk negara itu sendiri. Jika ada hubungan antara statistik dan statistika, maka setidaknya ada perbedaan antara keduanya yang memanifestasikan dirinya dalam beberapa hal, termasuk dalam hal pemahaman, tujuan, dan metode penggunaan. Contohnya tentu berbeda, karena statistik lebih ditekankan pada hasil atau data, sedangkan statistik lebih ditekankan pada ilmu atau pengolahan. Contoh gambaran statistik pada kehidupan kita sehari-hari antara lain :

a. Data kependudukan dan perekonomian dari Badan Pusat Statistik (BPS)
b. Data kepemilikan kendaraan bermotor di suatu kawasan perkotaan
c. Data belanja daerah dan anggaran pemerintah daerah milik Kemenkeu (DPJK)
d. Data kependudukan suatu desa yang didapatkan lewat survei
e. Data guna lahan suatu kabupaten yang dikeluarkan oleh dinas pertanahan

Lalu apa sebenarnya pengertian statistika? Istilah Statistik berbeda dengan Statistik. Ketika berbicara tentang statistik, yang terlintas dalam pikiran adalah tumpukan data yang berisi deretan angka, gambar, bagan, dan objek terkait data lainnya, baik kuantitatif maupun kualitatif. Data merupakan dokumen dasar kegiatan penelitian, yang selanjutnya akan dianalisis dengan metode uji statistik tertentu sesuai dengan kebutuhan penelitian.

2. Statistika

Pengertian statistika (*statistics*) adalah metode atau ilmu yang mempelajari cara merencanakan, mengumpulkan, menganalisis, dan menyajikan data untuk mengambil keputusan yang tepat. Statistika adalah bagian dari matematika yang secara khusus berhubungan dengan cara mengumpulkan, menganalisis, dan menafsirkan data. Dengan kata lain, istilah statistik digunakan di sini untuk merujuk pada kumpulan pengetahuan tentang metode pengambilan sampel (pengumpulan data), serta analisis dan interpretasi data. (Furqon, 1999: 3).

Gasperz (1989: 20) juga menegaskan bahwa "statistika adalah ilmu yang berkaitan dengan cara mengumpulkan data, mengolah dan menganalisisnya, menarik kesimpulan dan membuat keputusan rasional berdasarkan fakta yang tersedia". Somantri (2006: 17) senada menyatakan bahwa “statistika dapat dipahami sebagai ilmu yang mempelajari bagaimana kita mengumpulkan, mengolah, menganalisis, dan menginterpretasikan data sehingga dapat disajikan dengan lebih baik”. Sedangkan istilah statistik menurut Dajan (1995) diartikan sebagai suatu metode pengumpulan, pengolahan, penyajian, analisis dan interpretasi data dalam bentuk numerik.

Dengan demikian, statistika adalah ilmu yang berkaitan dengan pengumpulan, struktur, presentasi, analisis, dan interpretasi data menjadi informasi untuk membantu membuat keputusan yang efektif. Statistika adalah bidang pengetahuan yang memungkinkan peneliti untuk menarik dan mengevaluasi kesimpulan tentang populasi dan jenis sampel tertentu. Dengan kata lain, statistika adalah generalisasi tentang kelompok besar berdasarkan apa yang kita temukan dalam kelompok yang lebih kecil. Padahal, bidang statistika berkaitan dengan pengumpulan, pemilihan, klasifikasi, interpretasi dan analisis data yang digunakan untuk mengumpulkan dan mengevaluasi validitas dan reliabilitas kesimpulan berdasarkan data. Data yang diolah dan dihasilkan dari proses statistika ini disebut data statistik.

Mengapa Anda perlu mempelajari statistika? Statistika memiliki banyak kegunaan untuk membuat keputusan yang tepat di berbagai bidang kehidupan. Ada dua alasan penting untuk mempelajari statistika. Pertama, statistik memberikan pengetahuan dan kemampuan seseorang untuk

mengevaluasi data. Dengan pengetahuan statistika yang kita miliki, kita dapat menerima, meragukan, dan bahkan menyanggah data (kebenaran, validitas). Dalam kehidupan sehari-hari, kita benar-benar berurusan dengan statistik. Berikut ini adalah contoh statistika yang biasa dimanfaatkan pada kehidupan sehari hari antara lain :

a. Pengolahan data kependudukan suatu wilayah untuk menentukan piramida penduduk dengan menggunakan analisis *cohort* demografi
b. Pengolahan data kependudukan suatu wilayah untuk menentukan transisi demografi
c. Prediksi penduduk di masa depan dengan memanfaatkan proyeksi penduduk aritmatik
d. Menemukan rata-rata dan standar deviasi dari nilai ujian mahasiswa di suatu perguruan tinggi
e. Menemukan median umur penduduk di suatu desa.

Contoh yang mudah kita temukan belakangan ini adalah hasil polling yang diberikan oleh sejumlah media cetak, baik koran maupun Capital Matters. Beberapa hasil polling ini membuat kesimpulan berdasarkan pola yang ditarik. Beberapa kesimpulan yang diambil dari hasil survei valid, tetapi ada juga yang tidak. Selain masalah validitas, kita juga perlu memperhatikan masalah sampel karena ada survei yang dilakukan dengan jumlah sampel yang tidak mencukupi (besar). Untuk dapat menilai otentisitas atau validitas hasil (data) suatu penelitian, diperlukan statistik. Namun, statistik dapat dengan mudah digunakan untuk menyampaikan hasil yang berbeda dari keadaan sebenarnya jika orang yang menggunakan hasil atau kesimpulan dari suatu penelitian tidak memahami statistic.

Ketika kita menerapkan statistik untuk masalah ilmiah, industri atau sosial, pertama-tama kita mulai dengan studi populasi. Yang dimaksud dengan populasi dalam statistika dapat berupa populasi makhluk hidup, benda mati, atau benda abstrak. Populasi juga dapat menjadi ukuran suatu proses pada titik waktu yang berbeda, yang disebut deret waktu. Pengumpulan data (collecting data) dari seluruh populasi disebut sensus. Sebuah sensus tentu memakan banyak waktu dan biaya. Untuk itu dalam statistika sampling sering dilakukan, yaitu sebagian kecil dari suatu populasi, yang dapat mewakili seluruh populasi. Jika sampel yang diambil cukup representatif, maka inferensi

(pengambilan keputusan) dan kesimpulan yang diambil dari sampel tersebut dapat dipakai dalam menggambarkan populasi secara menyeluruh. Metode statistika bagaimana mendapatkan sampel yang tepat disebut teknik sampling.

Analisis statistik terutama menggunakan probabilitas sebagai konsep intinya, dapat dilihat bahwa uji statistik banyak digunakan dan didasarkan pada distribusi peluang. Sedangkan Matematika Statistika merupakan cabang dari Matematika Terapan yang menggunakan teori probabilitas dan analisis matematis untuk mendapatkan dasar-dasar teori statistik.

C. PEMBAGIAN STATISTIKA

Statistika dapat dibagi atas beberapa macam seperti cara pengolahan data, ruang lingkup penggunaan atau disiplin ilmu yang menggunakannya, dan bentuk parameternya.

1. Berdasarkan Cara Pengolahan Datanya

   Berdasarkan cara pengolahan datanya, statistika dapat dibedakan menjadi dua, yaitu : statistika deskriptif dan statistika Inferensial.

   a. Statistika Deskriptif

      Statistika deskriptif atau statistika deduktif mempunyai tujuan untuk mendeskripsikan atau memberi gambaran objek yang diteliti sebagaimana adanya tanpa menarik kesimpulan atau generalisasi. Dalam statistika deskriptif ini dikemukakan cara-cara penyajian data dalam bentuk tabel maupun diagram, penentuan rata- rata (mean), modus, median, rentang serta simpangan baku.

      Berikut adalah ruang lingkup Statistika Deskriptif menurut beberapa ahli:

      (1) Somantri (2006:19) berpendapat bahwa statistika deskriptif membahas cara-cara pengumpulan data, penyederhanaan angka-angka pengamatan yang diperoleh (meringkas dan menyajikan), serta melakukan pengukuran pemusatan dan penyebaran data untuk memperoleh informasi yang lebih menarik, berguna dan mudah dipahami.

      (2) Furqon (1999:3) menyatakan bahwa statistika deskriptif bertugas hanya untuk memperoleh gambaran (*description*) atau ukuran-ukuran tentang data yang ada di tangan.

(3) Pasaribu (1975:19) mengemukakan bahwa statistika deskriptif ialah bagian dari statistik yang membicarakan mengenai penyusunan data ke dalam daftar- daftar atau jadwal, pembuatan grafik-grafik, dan lain-lain yang sama sekali tidak menyangkut penarikan kesimpulan.

Jadi, statistik deskriptif adalah statistik yang berkaitan dengan pengumpulan, pengolahan, penyajian, dan perhitungan nilai suatu item data yang digambarkan dalam tabel atau diagram dan tidak mementingkan pengambilan kesimpulan. Contoh soal statistik deskriptif: (a) Tabel data, (b) Grafik batang, (c) Grafik lingkaran, (d) Grafik perubahan harga dari tahun ke tahun. Ruang lingkup pembahasan statistik deskriptif meliputi:

(1) Distribusi frekuensi dan bagian-bagiannya, seperti: grafik distribusi (histogram, polygon frekuensi, dan ogif); ukuran nilai pusat (rata-rata, median, modus, kuartil, dan sebagainya); ukuran dispersi (jangkauan, simpangan rata-rata, varians, dan sebagainya); kemiringan atau kurtosis kurva
(2) Angka indeks
(3) *Time series* deret waktu atau data berkala
(4) Korelasi dan regresi sederhana

b. Statistika Inferensial

Statistik inferensial atau induktif adalah statistik yang memberikan aturan atau metode yang dapat digunakan untuk membuat prediksi atau perkiraan dan untuk menarik kesimpulan umum dari sekumpulan data yang dipilih secara acak dalam subjek kumpulan data (populasi). Tujuan statistika adalah untuk menarik kesimpulan. Sebelum menarik kesimpulan, hipotesis dirumuskan berdasarkan statistik deskriptif. Contoh soal statistik inferensial: (a) Estimasi statistik, (b) Pengujian hipotesis (c) Prediksi dengan regresi/korelasi. Somantri (2006: 19) menyatakan bahwa statistik inferensial berkaitan dengan bagaimana data dianalisis dan keputusan dibuat (berkaitan dengan estimasi parameter dan pengujian hipotesis). Menurut Sudijono (2008:5), statistika inferensial adalah statistika yang memberikan aturan atau metode yang dapat digunakan sebagai upaya untuk menarik kesimpulan umum dari suatu kumpulan data yang telah disusun dan diolah. Subana (2000:12) mengemukakan bahwa

statistika inferensi adalah statistika untuk menarik kesimpulan umum dari data yang telah dikumpulkan dan diproses.

Statistik inferensia sifatnya jauh lebih dalam dan ditindaklanjuti dengan statistik deskriptif. Oleh karena itu, untuk mempelajari dan memahami statistika inferensial, terlebih dahulu harus mempelajari statistik deskriptif. Oleh karena itu, statistik inferensial adalah statistik yang mempelajari bagaimana keputusan dibuat.

2. Berdasarkan Ruang Lingkup Dan Penggunaannya

Berdasarkan atas ruang lingkup penggunaannya atau disiplin ilmu yang menggunakannya, statistika dapat dibagi atas beberapa macam, yaitu sebagai berikut :

a. Statistika Sosial adalah statistika yang dipakai dalam ilmu-ilmu sosial.
b. Statistika Pendidikan adalah statistika yang dipakai dalam ilmu dan bidang pendidikan.
c. Statistika Ekonomi adalah statistika yang dipakai dalam ilmu-ilmu ekonomi.
d. Statistika Perusahaan adalah statitistika yang dipakai dalam bidang perusahaan.
e. Statistika Kesehatan adalah statitistika yang dipakai dalam bidang kesehatan.
f. Statistika Pertanian adalah statitistika yang dipakai dalam bidang pertanian.

Berdasarkan penggunaannya, statistika yang digunakan dalam pengambilan keputusan secara luas oleh berbagai bidang ilmu seperti pemasaran, akuntansi, produksi, dan lain-lain. Tetapi secara umum pemahaman dan penggunaan alat-alat statistika diperlukan untuk membantu (Lies Maria Hamzah, et al, 2016) :

a. Mendeskripsikan dan memahami suatu hubungan. Sebagai alat, peneliti akan menggunakan teknik statistik untuk mengumpulkan data, mengolah dan menyajikannya untuk mengambil keputusan untuk tujuan tertentu. Karena membutuhkan kemampuan untuk mengidentifikasi dan menggambarkan hubungan antar variabel, misalnya seorang peneliti pemasaran untuk menggambarkan

hubungan antara permintaan produk dan harga dan pendapatan. Berdasarkan data tersebut, kegiatan penyaluran diarahkan kepada kelompok masyarakat tertentu.

b. Membuat keputusan yang lebih baik. Sebagai alat pengambilan keputusan, operasi statistik dalam pengumpulan data presentasional di mana data disajikan menunjukkan karakteristik semua data. Data dengan karakteristik tersebut merupakan informasi yang dapat digunakan untuk menentukan suatu tindakan.

c. Mengukur tingkat perubahan. Pada dasarnya, seseorang atau organisasi perlu merencanakan tujuan untuk masa depan. Perencanaan dimulai berdasarkan kejadian saat ini yang dapat digunakan sebagai dasar untuk peramalan masa depan. Meskipun metode statistik tidak tepat digunakan untuk memprediksi masa depan, mereka dapat membantu mengukur perubahan yang terjadi hari ini dan digunakan untuk proses peramalan.

3. Berdasarkan Bentuk Parameternya

   Berdasarkan bentuk parameternya (data yang sebenarnya), statistika dapat dibagi dua, yaitu statistika parametrik dan statistika nonparametric.

   a. Statistika Parametrik

      Statistika parametrik adalah bagian statistika yang parameter dari populasinya mengikuti suatu distribusi tertentu, seperti distribusi normal, dan memiliki varians yang homogen.

   b. Statistika Nonparametrik

      Statistika nonparametrik adalah bagian statistika dimana parameter dari populasinya tidak mengikuti suatu distribusi tertentu atau memiliki distribusi yang bebas dari persyaratan, dan variansnya tidak perlu homogen. Dengan perkataan lain uji statistik non-*parametrik* adalah uji yang modelnya tidak menetapkan syarat-syarat mengenai parameter populasi yang merupakan induk sampel penelitiannya. Perlu diketahui bahwa uji nonparametrik selayaknya tidak digunakan apabila uji parametrik dapat diterapkan, karena tingkat keampuhan uji nonparametrik lebih rendah dari pada uji parametrik. Namun kita sebagai pengambil keputusan jangan salah menafsirkan bahwa derajat kegunaan metode statistik nonparametrik dibawah metode statistik

parametrik. Tentu saja tidak demikian, masing-masing metode dibuat dengan spesifikasi khusus sesuai dengan macam data yang digunakan. Peningkatan keampuhan uji nonparametrik harus dengan memperbesar sampel. Namun seperti kita ketahui memperbesar sampel berarti akan menambah biaya, waktu, dll.

4. Berdasarkan Skala Pengukurannya

   Terdapat 4 skala pengukuran dalam statistik dan statistika, yaitu: skala nominal, skala ordinal, skala interval dan skala ratio.

   a. Skala Nominal

      Skala nominal merupakan skala data yang hanya dapat digunakan untuk membedakan objek yang bersifat kualitatif atau kategoris saja. Seperti jenis kelamin, agama, warna kulit, dan lain sebagainya. Data nominal adalah data statistik dimana dalam menyusun angkanya berdasarkan beberapa kategori dengan tidak memakai urutan tertentu. Jadi kedudukan kategori satu dengan yang lain adalah sama.Label, kode, symbol yang diberikan pada kategori untuk membedakan antara kategori satu dengan yang lain. Dengan demikian operasi aritmatika tidak berlaku bagi data nominal. Misalnya jenis kelamin laki-laki diberi nilai kode 1, dan jenis kelamin perempuan diberi kode 0, begitupun sebaliknya (Nata Wirawan, 2014). Contoh : Data mengenai barang-barang yang dihasilkan oleh sebuah mesin dapat digolongkan dalam kategori cacat atau tidak cacat. Barang yang cacat bisa diberi angka 0 dan yang tidak cacat diberi angka 1. Data 1 tidaklah berarti mempunyai arti lebih besar dari 0. Data satu hanyalah menyatakan lambang untuk barang yang tidak cacat.

   b. Skala Ordinal

      Skala ordinal merupakan skala pengukuran kualitatif dan nantinya akan diklasifikasikan dalam suatu kelompok tertentu dan diberi kode berhierarki. Seperti pendidikan, tingkat kepuasan pengguna, dan lain sebagainya. Data ordinal adalah data statistic yang angkanya disusun berdasarkan beberapa kategori dengan memperhatikan kedudukan ranking. Label, kode yang diberikan kepada masing-masing kategori menunjukkan peringkat dan urutan penilaian. Contoh : Sistem kepangkatan dalam dunia militer adalah satu contoh dari data berskala

ordinal Pangkat dapat diurutkan atau dirangking dari Prajurit sampai Sersan berdasarkan jasa, dan lamanya pengabdian.

c. Skala Interval

Skala interval merupakan skala pengukuran kualitatif dengan nilai nol mutlak maupun tidak sehingga menyebabkan bergesernya pengukuran tersebut sesuai dengan keinginan orang yang mengukur. Seperti suhu yang dinyatakan dalan celcius. Data interval adalah data yang disusun berdasarkan jarak yang sama antar golongan satu dengan yang lainnya. Data interval memiliki sifat data ordinal dan nominal serta ditambah satu sifat yaitu jarak yang sama antara kategori satu dengan yang lain. Contoh : Data tentang suhu empat buah benda A, B, C , dan D yaitu masing-masing 20. 30, 60, dan 70 derajat Celcius, maka data tersebut adalah data dengan skala pengukuran interval karena selain dapat dirangking, peneliti juga akan tahu secara pasti perbedaan antara satu data dengan data lainnya. Perbedaan data suhu benda pertama dengan benda kedua misalnya, dapat dihitung sebesar 10 derajat, dst

d. Skala Ratio

Skala ratio merupakan salah satu jenis pengukuran yang memiliki nol alamiah atau nol absolute, sehingga memungkinkan kita membandingkan magnitude angka-angka absolute. Seperti tinggi dan berat adalah contoh dari skala ratio. Seseorang yang memiliki berat badan 80kg bisa dikatakan dua kali lipat dari berat badan yang 40 kg. Data ratio ini diperoleh dengan membandingkan nilai variable yang satu dengan nilai absolut lainnya (variable pendamping). Data ratio memiliki sifat tiga jenis data yaitu : nominal, interval, ordinal ditambah satu sifat yaitu memiliki nilai nol atinya tidak ada. Penghasilannya nol artinya tidak mempunyai penghasilan sepeser pun. Contoh : Data mengenai berat adalah data yang berskala rasio. Dengan skala ini kita dapat mengatakan bahwa data berat badan 80 kg adalah 10 kg lebih berat dari yang 70 kg, tetapi juga dapat mengatakan bahwa data 80 kg. adalah 2x lebih berat dari data 40 kg. Berbeda dengan interval, skala rasio mempunyai titik nol yang mutlak.

D. PERBEDAAN STATISTIK DAN STATISTIKA

Perbedaan antara statistik dan statistika dapat dilihat dari beberapa aspek, meskipun pada dasarnya keduanya saling berkaitan. Penjelasan, lengkapnya sebagai berikut;

No	Perbedaan	Statistik	Statistika
1	Berdasarkan pengertiannya	Statistik adalah hasil pengolahan data yang disajikan dalam bentuk tabel, diagram, grafik sebagainya	Statistika adalah adalah metode ilmiah mengenai cara untuk mengumpulkan, mengelola, menganalisa penyajian data, menginterpretasi, dan mempresentasikan data.
2	Berdasarkan Tujuannya	Tujuan dari statistik yaitu untuk memperoleh gambaran atas data-data yang telah dikumpulkan dan dikaji sebelumnya. Berdasarkan data yang telah disajikan tersebut, selanjutnya dapat ditarik kesimpulan atas permasalahan yang sedang dikaji	Data hasil pengolahan statistika inilah yang dinamakan data statistik. Oleh sebab itu, secara umum statistika berguna untuk mengubah data dan informasi acak menjadi sebuah data statistik yang dapat dimengerti.
3	Berdasarkan metode yang digunakan	Statistik menggunakan metode kajian untuk mendapatkan kumpulan data terlebih dahulu yang diperoleh dan diolah	Statistika menggunakan metode penelitian yang bisa berupa survei dan eksperimen. Kedua metode tersebut sama-sama mengkaji

		pada proses statistika untuk dikaji setelahnya. Atau bisa dikatakan bahwa, metode yang digunakan dalam statistik lebih bersifat interpretatif terhadap 'statistik' apa yang telah dikeluarkan dari pengolahan menggunakan metode statistika.	tentang perilaku respons yang disebabkan oleh perubahan penjelas maupun pengaruhnya.

Berbagai penjelasan yang telah dikemukakan, dapat dikatakan bahwa statistik memiliki arti yang berbeda bagi orang yang berbeda. Bagi penggemar baseball, statistik adalah informasi tentang rata-rata *run* yang diperoleh pelempar atau persentase *slugging* pemukul atau jumlah *home run*. Bagi manajer pabrik di perusahaan distribusi, statistik adalah laporan harian tentang tingkat persediaan, ketidakhadiran, efisiensi tenaga kerja, dan produksi. Bagi seorang peneliti medis yang menyelidiki efek obat baru, statistik adalah bukti keberhasilan upaya penelitian. Dan bagi seorang mahasiswa, statistik adalah nilai yang dibuat pada semua ujian dan kuis dalam suatu mata pelajaran selama satu semester. Saat ini, statistik dan analisis statistik digunakan di hampir setiap profesi, dan bagi manajer khususnya, statistik telah menjadi alat yang paling berharga

Dapat dikatakan bahwa statistik adalah hasil data yang dikumpulkan melalui proses atau metode ilmiah, yang kemudian diolah serta diinterpretasikan untuk ditampilkan dalam bentuk grafik atau diagram. Data yang sudah diolah dan ditampilkan itulah yang dinamakan sebagai statistik. Sedangkan statistika merupakan suatu ilmu yang berguna untuk mempermudah dalam pengolahan dan menginterpretasikan data yang sudah dikumpulkan. Melalui statistika, data dari sebuah permasalahan yang sebelumnya telah dikumpulkan, bisa diolah, diinterpretasikan, dan kemudian digunakan untuk tujuan tertentu. Perbedaannya hanya terletak pada

pelaksanaan dari proses kajian tersebut. Atau bisa dikatakan bahwa, metode yang digunakan dalam statistika lebih berfokus pada pengelolaan data yang dikumpulkan untuk menghasilkan data statistik.

DAFTAR PUSTAKA

Anderson, R.L. and Bancroft, T.A. (1952) *Statistical Theory in Research*. New York, McGraw-Hill,

Dajan, Anto. (1995). *Pengantar Metode Statistik Jilid I.* Jakarta LP3ES.

Furqon, (1999),.*Statistika Terapan untuk Penelitian,* Bandung, Alfabeta.

Gaspersz, Vincent. (1989). *Statistika*. Bandung Armico

Hadi, Sutrisno,. (1989). *Statistik 1*. Yogyakarta: Andi Offset.

Hanafiah Adang Sutedja, Iskandar Ahmaddien,. (2020). *Pengantar Statistik,* Bandung, Penerbit Widina Bhakti Persada.

Iqbal Hakim,.(2020). *Statistik dan Statistika : Pengertian, Perbedaan dan Contoh.* https://insanpelajar.com/statistik-dan-statistika/

Lies Maria Hamzah,. Imam Awaluddin,. Emi Maimunah,.(2016). *Pengantar Statistika Ekonomi*. Bandar Lampung, Penerbit CV. Anugrah Utama Raharja

Nuryadi, Tutut Dewi Astuti, Endang Sri Utami, Budiantara,.(2017). *Dasar Statistik Penelitian,* Yogyakarta. Penerbit Sibuku Media.

Pasaribu, Amudi. (1975). *Pengantar Statistik*. Gahlia,. Jakarta, Gahlia Indonesia

Subana dkk,.(2000),. Statistik Pendidikan, Bandung, Pustaka Setia

Sudijono, Anas. (2008). *Pengantar Statistik Pendidikan*. Jakarta, Raja Grafindo Persada

Sudjana, M.A. (2005). *Metode Statistika.* Bandung: Tarsito

Wirawan,N,.(2014). *Cara Mudah Memahami Statistika Ekonomi dan Bisnis (Statistika Deskriptif)*, Denpasar, Penerbit Kararas Emas.

PROFIL PENULIS

MARHAWATI, dilahirkan di Birengere, Kabupaten Pangkajene dan Kepulauan, Provinsi Sulawesi Selatan tanggal 21 Juli 1963. Pendidikan Sekolah Dasar diselesaikan di kampung halamannya SDN Tonasa tahun 1975. Melanjutkan Pendidikan SMP Muhammadiyah di kota Ujung Pandang tahun 1979. Setelah tamat SMA NEGERI I Ujung Pandang tahun 1982, penulis melanjutkan studi S1 di Fakultas Pertanian jurusan Sosial Ekonomi Pertanian Universitas Hasanuddin dan meraih gelar (Ir) tahun 1987. Menyelesaikan S2 Program Studi Agribisnis, Program Pascasarjana Universitas Hasanuddin dan meraih gelar (M.Si) tahun 1997. Selanjutnya penulis melanjutkan S3 Ilmu Pertanian Sekolah Pasca Sarjana Universitas Hasanuddin dan meraih gelar (Dr) tahun 2019.

Sejak tahun 1989 sampai 2010 penulis mengajar di Fakultas Pertanian Universitas Tadulako Provinsi Sulawesi Tengah, kemudian hijrah ke Universitas Negeri Makassar tahun 2011, sampai saat ini penulis menjadi dosen tetap di Jurusan Bisnis dan Kewirausahaan Fakultas Ekonomi Dan Bisnis Universitas Negeri Makassar. Beberapa buku telah dihasilkan baik berupa buku referensi maupun book chapter dan telah memiliki hak kekayaan intelektual berupa hak cipta. Sebagai peneliti, telah menghasilkan beberapa artikel yang terbit pada jurnal dan prosiding baik skala nasional maupun Internasional.

BAB 2
TEKNIK PENGUMPULAN DATA

Dr. Ramlan Mahmud, M.Pd
Universitas Negeri Makassar

Dalam menulis atau mengerjakan karya ilmiah, peneliti tentu harus memilih teknik pengumpulan data yang tepat. Teknik tersebut dinilai sangat krusial atau penting demi lancarnya penelitian yang dilakukan. Selain itu, teknik pengumpulan data juga harus dilakukan agar penelitian lebih terarah dan terkendali.

Dalam memilih teknik pengumpulan data, tentu ada beberapa teknik yang harus dilakukan untuk meminimalisasi adanya hambatan, kesalahan, atau masalah yang terjadi selama penelitian berlangsung. Sehingga teknik yang dipilih juga harus tepat dan berlangsung secara sistematis.

Untuk itu, Anda harus mengetahui berbagai hal mengenai teknik pengumpulan data, mulai dari pengertian, proses pengumpulan data, berbagai macam teknik pengumpulan data, dan juga jenis-jenis data yang akan dikumpulkan memiliki jenis atau klasifikasi seperti apa.

Di bawah ini, akan dijelaskan berbagai hal mengenai teknik pengumpulan data dan bisa menjadi bekal bagi Anda yang ingin melakukan penelitian.

A. PENGERTIAN TEKNIK PENGUMPULAN DATA

Teknik pengumpulan data merupakan teknik atau metode yang digunakan untuk mengumpulkan data yang akan diteliti. Artinya, teknik pengumpulan data memerlukan langkah yang strategis dan juga sistematis untuk mendapatkan data yang valid dan juga sesuai dengan kenyataannya.

Selain itu, teknik atau metode pengumpulan data ini biasanya digunakan untuk peneliti demi mengumpulkan data yang merujuk pada satu kata abstrak yang tidak diwujudkan dalam benda, tetapi hanya dapat dilihat penggunaannya. Misalnya adalah melalui angket, wawancara, pengamatan, uji atau tes, dokumentasi, dan lain sebagainya.

Dilakukannya pengumpulan data untuk penelitian agar data dan teori yang ada di dalamnya valid dan juga sesuai kenyataan, sehingga peneliti harus benar-benar terjun langsung dan mengetahui teknik pengumpulan data tersebut. Dengan demikian, peneliti akan mengetahui validitas atau kebenaran konsep penelitiannya.

Secara umum, teknik pengumpulan data ini digunakan peneliti untuk dapat mengumpulkan data atau informasi berdasarkan fakta pendukung yang

ada di lapangan demi keperluan penelitian dan teknik yang dilakukan sangat ditentukan oleh metodologi penelitian yang dipilih oleh peneliti itu sendiri.

Di dalam melakukan teknik pengumpulan data atau proses mengumpulkan data, keberadaan instrumen penelitian menjadi bagian yang sangat integral dan termasuk ke dalam komponen metodologi penelitian karena instrumen penelitiannya berupa alat yang digunakan untuk mengumpulkan, memeriksa, dan menyelidiki masalah yang diteliti.

Tentu saja, keberadaan instrumen tersebut akan membantu berbagai penelusuran terhadap gejala yang ada pada penelitian sehingga dapat digunakan untuk membuktikan kebenaran atau untuk menyanggah berbagai hipotesis. Oleh sebab itu, instrumen yang digunakan harus memiliki validitas dan reliabilitas yang baik.

Sebelum memulai melakukan teknik pengumpulan data, ada beberapa hal yang harus diperhatikan yaitu:

1. Data biasanya sudah ditentukan oleh beberapa variabel penelitian. Ketika semua data terkumpul, langkah berikutnya adalah mengolah data, sehingga data yang dikumpulkan memiliki arti karena diolah dengan sistematis.
2. Data yang sudah diolah tersebut dipakai dan dipilih berdasarkan data yang berhubungan atau relevan dengan konsep, kejadian, atau objek penelitian. Datanya bisa berbentuk huruf, angka, simbol, gambar, dan lainnya.
3. Setelah itu, pengumpulan data dilakukan untuk memperoleh informasi yang dibutuhkan dalam rangka mencapai tujuan penelitian yang diungkap dalam bentuk hipotesis yaitu jawaban sementara terhadap pertanyaan penelitian.
4. Data yang sudah dikumpulkan ditentukan oleh berbagai variabel yang ada di dalam hipotesis dan dikumpulkan dalam bentuk sampel yang sudah ditentukan sebelumnya dan sampelnya digunakan untuk menganalisis sasaran penelitian.

B. PROSES PENGUMPULAN DATA

Dalam teknik pengumpulan data, tentu saja ada proses yang harus dilakukan. Prosesnya harus terlaksana secara sistematis dan terarah agar data

yang dikumpulkan bisa dibuktikan kebenarannya. Karena pada dasarnya, proses pengumpulan data dalam teknik mengumpulkan data ini nanti harus bisa membuktikan hipotesis dari data yang hasilnya sudah dikumpulkan oleh peneliti.

Berikut ini, ada 8 tahap atau proses yang harus dilakukan sebagai tahapan pengumpulan data.

1. Tinjau literatur dan konsultasi dengan ahli.

   Proses atau tahap pertama yang harus dilakukan untuk mengumpulkan data yakni mengumpulkan berbagai informasi yang berhubungan dengan masalah penelitian. Informasi ini diperoleh melalui tinjauan literatur dan konsultasi dengan para ahli sehingga peneliti benar-benar mengerti isu, konsep, dan variabel yang ada di dalam penelitian.
2. Mempelajari dan melakukan pendekatan terhadap kelompok masyarakat di mana data akan dikumpulkan. Tahap kedua atau proses yang dilakukan setelah tinjauan literatur adalah peneliti harus mempelajari dan melakukan pendekatan terhadap kelompok masyarakat yang kemudian penelitiannya bisa diterima dan juga berkaitan dengan tokoh-tokoh yang bersangkutan.
3. Membina dan memanfaatkan hubungan yang baik dengan responden dan lingkungannya.Tahap selanjutnya adalah membina hubungan baik dengan responden dan lingkungannya. Ini termasuk pada mempelajari bagaimana kebiasaan yang dilakukan responden dan cara berpikir mereka, melakukan sesuatu, bahasa yang digunakan, dan lain sebagainya untuk mendukung berlangsungnya penelitian.
4. Uji coba atau *pilot study*

   Selanjutnya, tahapan yang harus dilakukan adalah melakukan uji coba instrumen penelitian pada kelompok masyarakat yang merupakan bagian dari populasi, bukan sampel. Maksudnya untuk mengetahui apakah instrumen yang digunakan cukup dipahami, bisa digunakan, komunikatif atau tidak, dan lain sebagainya.
5. Merumuskan dan menyusun pertanyaan

   Setelah itu, instrumen yang sudah didapatkan disusun dalam bentuk pertanyaan yang relevan dengan tujuan penelitian. Pertanyaan yang dirumuskan harus mengandung makna yang signifikan dan substantif.

6. Mencatat dan memberi kode (*recording and coding*)
   Setelah instrumen penelitian disiapkan, dilakukan pencatatan terhadap data yang dibutuhkan dari setiap responden. Berbagai informasi yang diperoleh ini perlu dicatat guna memudahkan proses analisis.
7. *Cross checking*, validitas, dan reliabilitas
   Setelah itu, dilakukan metode *cross checking* terhadap data yang didapatkan untuk menguji lagi kebenarannya dan memeriksa sehingga tidak ada keraguan terhadap validitas dan reliabilitasnya.
8. Pengorganisasian dan kode ulang data yang telah terkumpul supaya dapat dianalisis
   Terakhir, setelah data terkumpul, penulis harus melakukan koordinasi terhadap berbagai data yang sudah dikumpulkan, dan Anda bisa mulai menganalisis data tersebut sehingga tidak ada data yang kurang valid.

C. TEKNIK PENGUMPULAN DATA

Setelah memahami pengertian dan juga proses pengumpulan data, selanjutnya akan dijelaskan mengenai berbagai teknik pengumpulan data. Setidaknya ada empat teknik pengumpulan data. Teknik pengumpulan data adalah teknik atau metode yang digunakan untuk mengumpulkan data. Konsekuensi dari data yang dikumpulkan secara tidak benar meliputi:

1. Ketidakmampuan untuk menjawab pertanyaan penelitian secara akurat
2. Ketidakmampuan untuk mengulang dan memvalidasi penelitian
3. Temuan yang menyimpang menghasilkan sumber daya yang terbuang
4. Menyesatkan peneliti lain untuk menemukan jalan investigasi tanpa hasil
5. Keputusan kompromi untuk kebijakan publik
6. Menyebabkan kerusakan pada subjek penelitian (misalnya manusia dan hewan)

Teknik pengumpulan data dapat diartikan sebagai langkah strategis dalam penelitian. Baik itu bisnis, pemasaran, humaniora, ilmu fisika, ilmu sosial, atau bidang studi atau disiplin lainnya, data memainkan peran yang sangat penting, yang berfungsi sebagai titik awal masing-masing. Sebelum menuju ke pembahasan teknik pengumpulan data, kita kenali dulu apa itu data. Dalam suatu penelitian, data yang baik harus memenuhi persyaratan sebagai berikut:

1. Tujuan, Data diperoleh dari lapangan dan dilaporkan sebagaimana adanya
2. Relevan, data harus sesuai dengan masalah yang diteliti
3. Up to Date, data harus selalu menyesuaikan dengan perkembangan (tidak boleh ketinggalan jaman)
4. Representatif, data harus diperoleh dari sumber yang sesuai dan mewakili kondisi sebenarnya dari suatu kelompok atau populasi

Secara umum data terbagi menjadi 2 bagian yaitu data kuantitatif, data kuantitatif adalah data yang dapat dinyatakan dalam angka dan dapat diukur ukurannya. Contoh data kuantitatif adalah harga smartphone, berat dan tinggi badan, jumlah pembeli, dan lain sebagainya.data kualitatif, data kualitatif adalah data yang berkaitan dengan pengelompokan atau karakteristik yang tidak dapat diukur ukurannya. Dengan kata lain, data kualitatif diekspresikan dalam bentuk kata-kata yang memiliki makna.

Berikut ini merupakan teknik pengumpulan data menurut Sugiyono (2017).

1. Observasi (Pengamatan)
   Teknik observasi artinya melakukan pengamatan dan pencatatan secara sistematis mengenai gejala yang tampak pada objek penelitian. Observasi ini tergolong teknik pengumpulan data yang paling mudah dilakukan dan biasanya juga banyak digunakan untuk statistika survei, misalnya meneliti sikap dan perilaku suatu kelompok masyarakat. Dengan teknik observasi, peneliti biasanya terjun ke lokasi yang bersangkutan untuk memutuskan alat ukur yang tepat untuk digunakan.
2. *Kuestioner* (Kuesioner/Angket)
   Teknik yang kedua adalah kuestioner atau kuesioner yang artinya teknik pengumpulan data dengan cara memberikan seperangkat pertanyaan atau pernyataan kepada orang lain yang berperan sebagai responden agar dapat menjawab pertanyaan dari peneliti. Meski terlihat mudah, teknik ini cukup sulit dilakukan jika jumlah respondennya besar dan tersebar di berbagai wilayah. Ada beberapa prinsip yang harus diperhatikan saat memilih teknik pengumpulan data kuesioner, yaitu:
   a. Isi dan tujuan pertanyaannya ditujukan untuk mengukur mana yang harus ada dalam skala yang jelas dan dalam pilihan jawaban.

b. Bahasa yang digunakan harus sesuai dengan kemampuan responden, sehingga tidak mungkin menggunakan bahasa yang penuh dengan istilah asing atau bahasa asing yang tidak dimengerti responden.
c. Tipe dan bentuk pertanyaannya bisa terbuka atau tertutup. Terbuka artinya jawaban yang diberikan bebas, dan tertutup artinya responden hanya boleh memilih jawaban yang sudah disediakan.

3. *Interview* (Wawancara)
   Teknik wawancara atau *interview* ini dilakukan secara tatap muka melalui tanya jawab antara peneliti atau pengumpul data dengan responden atau narasumber atau sumber data. Teknik pengumpulan data dengan wawancara biasanya dilakukan sebagai studi pendahuluan, karena teknik ini tidak mungkin dilakukan jika respondennya dalam jumlah besar.
4. *Document* (Dokumen)
   Teknik pengumpulan data yang terakhir adalah dokumen yang mana peneliti mengambil sumber penelitian atau objek dari dokumen atau catatan dari peristiwa yang sudah berlalu, baik dalam bentuk tulisan, gambar, atau karya monumental dari seseorang. Bisa diambil dari catatan harian, sejarah kehidupan, biografi, peraturan, dan lain sebagainya.

D. JENIS-JENIS DATA

Dalam teknik pengumpulan data, tentu saja dibutuhkan data yang berupa fakta yang valid sebagai informasi. Sehingga ada beberapa jenis data yang bisa dipilih dan dikategorikan. Pembagian data ini dibagi menjadi tiga, yaitu: (a) berdasarkan tipe penelitian, (b) berdasarkan sumber, dan (c) berdasarkan cara memperoleh.

1. Berdasarkan Tipe Penelitian
   a. Data Kualitatif
      Data kualitatif merupakan data yang berbentuk narasi atau deskripsi yang bertujuan untuk menjelaskan suatu fenomena atau kualitas sebuah fenomena yang tidak bisa diukur secara numerik.
      Contoh
      1) deskripsi suatu daerah yang diteliti

2) biografi narasumber yang dijadikan referensi di dalam penelitian
3) sejarah berdirinya suatu tempat yang diteliti

Pengumpulan data dalam penelitian tentunya harus dilakukan secara ilmiah dan sistematis. Peneliti melakukan survey dengan cara menyebar kuesioner atau angket sebagai instrumen penelitian, kuesioner menjadi wadah yang efektif dan efesien untuk mengumpulkan data yang akan diukur secara numerik.

b. Data Kuantitatif

Sementara itu, data yang didapatkan pada teknik pengumpulan data berupa data kuantitatif maksudnya jenis data yang dapat diukur atau measurable dan bisa dihitung langsung sebagai variabel angka atau suatu bilangan. Variabel ini menjadi atribut atau karakteristik untuk mengukur dan mendeskripsikan suatu kasus atau objek pada penelitian tersebut.

Contoh

1) data jumlah karyawan setiap tahun pada suatu perusahaan yang akan dijadikan objek penelitian
2) data penjualan barang pada suatu toko tiap harinya
3) data mengenai usia siswa dalam suatu kelas

2. Berdasarkan Sumbernya

Berdasarkan sumbernya, data dalam teknik pengumpulan data dibagi menjadi dua yakni data primer dan data sekunder.

a. Data Primer

Data primer pada teknik pengumpulan data adalah data utama atau data pokok yang digunakan di dalam penelitian. Biasanya, data primer ini bisa dideskripsikan sebagai jenis data yang diperoleh langsung dari tangan pertama subjek penelitian atau responden atau narasumber, dan lain sebagainya, kecuali pada riset kuantitatif.

Contoh data primer misalnya:

Pada sensus ibu dan anak-anak di salah satu kelurahan, karyawan di kelurahan tersebut akan mengambil data dari ketua RT atau RW yang mana data tersebut didapatkan dari berapa jumlah ibu dan anak dalam suatu KK di wilayah tersebut. Kemudian data tersebut dijadikan data primer.

b. Data Sekunder

Data sekunder merupakan data dalam teknik pengumpulan data yang menjadi data pelengkap. Artinya data tersebut diperoleh tidak melalui tangan pertama responden atau narasumber, melainkan dari tangan kedua, tangan ketiga, dan seterusnya. Sama halnya dengan data primer, perkecualian ini berlaku pada riset kuantitatif. Biasanya, peneliti akan mencontohkan berbagai dokumen, misalnya seperti literatur atau naskah akademik, koran, majalah, pamflet, dan lain sebagainya sebagai media yang tepat mendapatkan data sekunder.

Contoh

Adanya catatan atau dokumentasi pada sebuah perusahaan berupa presensi, besaran gaji, laporan keuangan, publikasi perusahaan, laporan pajak perusahaan, dan lain sebagainya yang diperoleh melalui pembukuan atau majalah dan lain sebagainya.

3. Berdasarkan Cara Memperolehnya

Berdasarkan cara memperolehnya, teknik pengumpulan data dibagi menjadi tiga.

a. Observasi

Observasi merupakan teknik pengumpulan data yang dilakukan melalui pengamatan langsung. Peneliti biasanya terjun ke lapangan untuk melakukan pengamatan terhadap objek penelitian dan diamati menggunakan pancaindera. Dalam teknik ini, peneliti biasanya berperan sebagai orang asing yang mengamati secara langsung. Setelah itu, data atau hasil yang didapat dan dikumpulkan dicatat baik dalam bentuk tulisan, rekaman suara, foto, video, dan lain sebagainya. Sifat penelitian observasi ini partisipatoris yang mana peneliti langsung bergabung dan juga melakukan pengamatan objeknya bersama-sama. Cara pengambilan data pada teknik observasi ada dua, yaitu observasi partisipasi dan non partisipan.

b. Observasi Partisipasi

Observasi partisipasi ini dilakukan dengan cara peneliti turut langsung untuk berpartisipasi dalam kegiatan yang dilakukan kelompok yang diteliti. Peneliti kemudian melakukan aktivitas yang dilakukan oleh kelompok yang diteliti, sehingga meski hanya melakukan pengamatan, peneliti ikut membaur dalam kegiatan

tersebut. Metode ini sangat cocok untuk penelitian yang sifatnya memuat aspek psikis, misalnya kesan, pemaknaan, apa yang dirasakan, dan lain-lain. Akan tetapi, penelitian ini dirasa kurang objektif karena peneliti hanya mengetahui orang yang diteliti atau partisipan umumnya mengetahui bahwa mereka sedang diteliti.

Contoh observasi partisipasi adalah ketika peneliti ikut terjun bermain gobak sodor dengan anak-anak di kampung yang diteliti ketika ia meneliti mengenai permainan tradisional anak di wilayah tersebut.

c. Observasi Non Partisipan

Observasi ini dilakukan dengan cara tidak berpartisipasi atau mengikuti aktivitas yang dilakukan kelompok yang diteliti. Ia hanya menempatkan diri sebagai penonton. Teknik pengumpulan data ini biasanya dilakukan secara diam-diam, agar partisipan tidak menyadari bahwa mereka sedang diamati. Sehingga akurasi data bisa terjamin. Akan tetapi, peneliti harus memiliki pengetahuan yang lebih dan sudah lebih dulu membaca teori-teori penelitian yang dilakukan karena teknik pengumpulan data ini akan sulit jika dilakukan hanya dengan cara mengamati saja.

Contoh observasi non partisipasi adalah misalnya peneliti meneliti tradisi atau adat Pasang Kembar Mayang pada upacara pernikahan adat Jawa. Peneliti hanya melihat dan mengamati apa yang dilakukan saat berlangsungnya kegiatan tersebut.

4. Wawancara

Teknik pengumpulan data berupa wawancara ini dilakukan dengan cara tanya jawab antara peneliti dengan responden agar mendapat informasi atau persepsi subjektif dari informan terkait topik yang ingin diteliti. Peneliti sebelumnya harus menyiapkan pertanyaan-pertanyaan wawancara terlebih dahulu.

5. Eksperimental

Terakhir yakni teknik pengumpulan data dengan metode eksperimental. Artinya penelitian ini dilakukan dengan sengaja dan peneliti melakukan manipulasi terhadap satu atau lebih variabel tertentu yang kemudian akan berpengaruh pada variabel lain yang diukur. Teknik ini dilakukan dengan tujuan meneliti berbagai kemungkinan sebab akibat dengan menggunakan

satu atau lebih kondisi perlakuan pada satu atau lebih kelompok eksperimen dan kemudian membandingkan hasilnya sebagai kontrol.

DAFTAR PUSTAKA

Anselm Staruss, Juliet Corbin. 1995. Basic of Qualitative Research For Education ; An Introductio to Theory and Method : Allyn and Bacon; Boston London.

Bogdan, Robert C, Biklen, 1982. Qualitative Research For Education : An Introduction to Theory and Methods, Allyn and Bacon: Boston London

Borg R. Waltre, Gall Meredith. D.1989. Educational Research : In Introduction, Fifth Edition : Logman

Cook Thomas D.1979. Qualitative and Quantitative Method Instrument Evaluation Research, Sage Publication: Beverly Hills.

Hopkins, D., A 2021.Teacher's Guide to Classroom Research, Buckingham. Open University Press.

Kerlinger, Fred. N.1973. Foundation of Behavioral Research. Holt, Renehart.

Nana Sudjana, Ibrahim. 1989. Penelitian dan Penilaian Pendidikan, Sinar Baru: Bandung

Siregar, N.1998. Penelitian Kelas : Teori, Metodologi & Analisis. IKIP Bandung Press.

Sugiyono, 2011.Metode Penelitian Kuantitatif, Kualitatif dan R&D. Alfabeta: Bandung

PROFIL PENULIS

Dr. Ramlan Mahmud, M.Pd. lahir di Kabupaten Sinjai Provinsi Sulawesi Selatan. Pada tahun 1995 masuk Sekolah Dasar (SD) Negeri 1 Sinjai dan tamat pada tahun 1997. Pada tahun yang sama melanjutkan pendidikan pada Sekolah Menengah Pertama (SMP) Negeri 3 Sinjai dan tamat pada tahun 2000. Pada tahun itu juga melanjutkan pendidikan pada Sekolah Menengah Atas (SMA) Negeri 1 Sinjai dan tamat pada tahun 2003. Pada tahun yang sama melanjutkan pendidikan pada jurusan pendidikan Matematika FMIPA UNM dan memperoleh gelar sarjana pendidikan matematika pada tahun 2007. Pada tahun 2010 penulis menyelesaikan program pascasarjana pada Universitas Negeri Makassar (UNM) dan memperoleh gelar Magister Pendidikan Matematika. Tahun 2015 menyelesaikan Program Doktor Ilmu Pendidikan di UNM dan Sandwich like di Nothern Illionis University, USA. Beberapa organisasi antara lain Ketua Relawan Jurnal Indonesia (RJI) Sulawesi Selatan, Fasilitator RJI, Anggota Forum Dosen Indonesia (FDI), ADRI, IndoMS, HEBII. Tercatat sebagai Asessor BKD Dosen tahun 2021-sekarang. Reviewer dan editor pada beberapa jurnal Nasional dan International, diantaranya pada jurnal Daya Matematis (Sinta3) pada Program Pascasarjana UNM, jurnal IJOLE (Q1), EST (Sinta2). Sejak tahun 2010 mengajar pada PGSD FIP UNM, STKIP YPUP Makassar, FTK UIN Alauddin, STKIP Andi Matappa, STIM Nitro, dan UNISMUH Makassar. Beberapa buku antara lain: Penulisan Jurnal ilmiah berbasis OJS dan OJS 3 (2020), Filsafat Pendidikan Matematika (2022), Strategi Metakognitif Model Pembelajaran sosial dalam matematika (2016), Pengantar Ilmu Pendidikan (2019). Beliau juga sebagai Direktur Sains Global Intitut Indonesia, dan Pendiri Publikasi Ilmiah Ilin Institut.

BAB 3
UKURAN PEMUSATAN

Nurdiana, S.P., M.Si
Universitas Negeri Makassar

Pemusatan data adalah sebarang ukuran yang menunjukkan pusat segugus data,yang telah diurutkan dari yang terkecil sampai yang terbesar atau sebaliknya dari yang terbesar sampai yang terkecil. Salah satu kegunaan dari ukuran pemusatan data adalah untuk membandingkan dua (populasi) atau contoh, karena sangat sulit untuk membandingkan masing-masing anggota dari masing-masing anggota populasi atau masing-masing anggota data contoh. Nilai ukuran pemusatan ini dibuat sedemikian sehingga cukup mewakili seluruh nilai pada data yang bersangkutan

Mean, Median, dan Modus adalah tiga jenis metrik konsentrasi data yang digunakan dalam statistik deskriptif. Namun, masing-masing memiliki kelebihan dan kekurangannya sendiri dalam menjelaskan ukuran konsentrasi data. Untuk memahami untuk apa masing-masing dan kapan menerapkannya, terlebih dahulu memahami apa itu analisis statistik deskriptif dan pengukuran konsentrasi data. Analisis statistik deskriptif adalah strategi untuk menyajikan data dengan cara yang berharga bagi pengguna. Upaya penyajian ini bertujuan untuk menyaring informasi kunci dari data ke dalam format yang lebih ringkas dan lugas, sehingga memerlukan penjelasan dan interpretasi. Dalam bentuknya yang paling dasar, statistik adalah seperangkat ide dan prosedur untuk mengumpulkan, menganalisis, dan menggambarkan data sampel untuk menarik kesimpulan yang berguna. informasi terkait pengumpulan data yang penyelidikan dan kesimpulannya didasarkan pada bukti numerik.

A. PENGERTIAN UKURAN PEMUSATAN

Istilah pemusatan data atau tendensi sentral digunakan dalam statistik untuk menggambarkan ukuran statistik yang mewakili nilai tunggal dari distribusi atau kumpulan data. Tentu saja, tujuannya adalah untuk mendapatkan gambaran lengkap dari semua data yang disebarluaskan. Pemusatan data, dengan kata lain, adalah ringkasan deskriptif dari sekumpulan data.

Pusat distribusi data dapat dicerminkan oleh satu nilai dari kumpulan data. Selain itu, tidak termasuk informasi tentang data individu dari kumpulan data, yang umumnya dilakukan dalam ringkasan kumpulan data.

Ukuran pemusatan adalah angka tunggal yang mewakili kumpulan data dan menjelaskan fitur-fiturnya. Pusat nilai data ditunjukkan dengan ukuran

pemusatan. Tahap pertama dalam menentukan ukuran pusat data adalah mengidentifikasi sebaran suatu kumpulan data, apakah itu populasi atau sampel.

Selanjutnya, setiap metrik yang mewakili pusat kumpulan data yang telah disusun dari terkecil ke terbesar atau sebaliknya dari terbesar ke terkecil disebut sebagai ukuran pusat data. Karena sulitnya membandingkan setiap anggota dari setiap populasi atau setiap anggota data sampel, salah satu kegunaan ukuran konsentrasi data adalah dengan membandingkan dua (populasi) atau sampel. Nilai ukuran sentralisasi ini dipilih agar cukup mewakili semua nilai yang ditemukan dalam data yang relevan. Secara umum, berbagai jenis ukuran statistik, seperti mean (rata-rata), mode (nilai yang sering muncul), dan median, dapat digunakan untuk menilai tendensi sentral dari suatu kumpulan data (nilai tengah).

Adapun definisi ukuran pemusatan data menurut para ahli adalah sebagai berikut.

1. Walpole, Ronald E. (1993), Setiap ukuran yang menampilkan pusat kumpulan data yang telah diurutkan dari terkecil ke terbesar atau sebaliknya dari data terbesar ke terkecil dikenal sebagai ukuran pusat data.
2. Iqbal I (2001), Ukuran pemusatan data didefinisikan sebagai metrik yang dapat digunakan untuk mewakili data secara keseluruhan. Artinya, jika semua nilai data diurutkan berdasarkan besarnya, nilai rata-rata dicatat di posisi tengah atau tengah.

Ukuran konsentrasi adalah suatu nilai yang menggambarkan atau menggambarkan nilai-nilai data yang dikumpulkan dari suatu peristiwa yang telah diamati, menurut buku Matematika kelas XI (Kemendikbud, 2014: 3). Rata-rata, median, dan modus adalah contoh metrik konsentrasi data. Ukuran konsentrasi, menurut Boediono dan Koster (2014: 56), adalah nilai tunggal yang mewakili semua data dan menampilkan pusat data, yang dipisahkan menjadi tiga bagian: mean (rata-rata terhitung), median, dan modus. Reaburn (2013: 562) sependapat dengan Boediono dan Koster bahwa metrik konsentrasi yang paling banyak digunakan adalah mean, median, dan modus, yang masing-masing dapat merepresentasikan data.)

Mahasiswa harus memahami bagaimana mean, median, atau modus dapat merepresentasikan data dengan menggunakan pengertian ukuran

pemusatan data (Amiruzzaman et all, 2016:20). Jika ada satu nilai ekstrim dalam data, median atau modus dapat memberikan representasi data yang lebih baik (Batanero, 1994: 532). Menurut Reaburn (2013: 562), salah satu metrik konsentrasi data (rata-rata, median, atau mode) mungkin lebih mencerminkan data daripada yang lain. Jika data memiliki satu atau lebih nilai ekstrem, seperti nilai yang secara signifikan lebih tinggi atau lebih rendah daripada data lainnya, rata-ratanya mungkin lebih tinggi atau lebih rendah dari sebagian besar data, dan median akan lebih khas dari data tersebut.

B. JENIS-JENIS UKURAN PEMUSATAN DATA

1. Mean

Perkiraan rata-rata (mean) adalah hasil perkalian semua nilai data dengan jumlah titik data. Rata-rata terhitung adalah nilai yang menggambarkan pusat nilai dan dapat digunakan untuk menunjukkan pusat data.

Mean dihitung dengan membagi jumlah nilai individu dalam data dengan jumlah total nilai, menurut Reaburn (2013: 562). Mean adalah salah satu ukuran konsentrasi data dengan membandingkan seluruh jumlah data dengan sejumlah besar data, menurut buku Matematika untuk kelas XI (Kemendikbud, 2014: 7). Mean adalah nilai yang menggambarkan atau menggambarkan nilai data yang dikumpulkan dari suatu peristiwa dengan membagi total nilai individu dalam data dengan jumlah nilai, seperti dijelaskan di atas.

Namun, setiap ukuran konsentrasi data memiliki fitur tersendiri (Boediono dan Koster, 2014: 67). Berikut ciri-ciri mean:

a. Memperhitungkan semua nilai
b. Memiliki kemampuan untuk menggambarkan mean populasi
c. Variabel yang paling stabil adalah varians.
d. Sesuai untuk data yang homogen (tidak ada nilai ekstrim dalam data)

a. Rataan populasi

Nilai rata-rata statistik populasi adalah rata-rata jumlah populasi. Populasi terdiri dari semua individu dari suatu lingkungan atau kelompok. Rumus berikut digunakan untuk menghitung jumlah populasi rata-rata:

$$Rata - rata\ hitung\ populasi = \frac{Jumlah\ seluruh\ nilai\ dalam\ populasi}{jumlah\ data\ dalam\ populasi}$$

Rataan hitung populasi ini juga sering disebut dengan istilah parameter juga dapat disajikan dalam bentuk simbol yaitu :

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

Keterangan :

μ = rata-rata hitung populasi

$\sum X$ =jumlah dari nilai rata yang berada dalam populasi

N =jumlah total dalam populasi

b. Rata-rata sampel

Rata-rata sampel dihitung dengan menggunakan unsur sampel yang dimiliki. Untuk menghitung rata-rata sampel dapat digunakan dengan menggunakan rumus berikut.

$$Rata - rata\ hitung\ sampel = \frac{Jumlah\ seluruh\ nilai\ dalam\ sampel}{jumlah\ data\ dalam\ sampel}$$

Dalam statistik, rata-rata sampel juga disajikan dalam bentuk simbol berikut ini.

$$\bar{X} = \frac{\sum X}{n}$$

Mengapa disebut sebagai sampel mean? Ini karena, dalam statistik, sampel dan populasi memiliki arti yang sangat berbeda, dan perbedaan ini signifikan, bahkan jika dihitung dengan cara yang sama dalam hal mean. Model kumpulan data kami pada dasarnya adalah mean. Ini adalah nilai

yang paling banyak digunakan. Namun, kita akan melihat bahwa mean tidak selalu merupakan salah satu nilai aktual yang ditemukan dalam pengumpulan data kita. Namun, salah satu kualitas terpentingnya adalah mengurangi kesalahan dalam prediksi nilai kumpulan data mana pun. Artinya, itu adalah nilai dalam kumpulan data yang menciptakan kesalahan paling sedikit dari semua yang lain. Fakta bahwa setiap nilai dalam kumpulan data kami termasuk dalam perhitungan adalah properti penting dari mean. Selanjutnya, mean adalah satu-satunya ukuran tendensi sentral di mana total penyimpangan setiap nilai dari mean selalu nol.

c. Rata-rata data perkelompok

Data yang telah dikelompokkan dalam bentuk distribusi frekuensi disebut sebagai data yang dikelompokkan. Kualitas data yang telah diklasifikasikan ke dalam satu kelas akan sama, yang ditunjukkan oleh nilai tengah kelas tersebut.

$$\bar{X} = \frac{\sum fX}{n}$$

Dimana :

$\bar{X}$ = rata-rata hitung data berkelompok

$\sum fX$ =jumlah dari seluruh hasil perkalian antara frekuensi dan nilai tengah masing-masing kelas

n =jumlah total data

Contoh :

Berikut adalah data yang sudah dikelompokkan dari 20 perusahaan yang sahamnya menjadi pilihan pada bulan Maret 2003. Buatlah nilai rata-rata untuk harga saham pilihan tersebut!

Interval	**Nilai Tengah (X)**	**Frekuensi**
160 – 303	231,5	2
304 – 447	375,5	5
448 – 591	519,5	9
592 – 735	663,5	3
736 – 878	807	1

Penyelesaian :

$$\bar{X} = \frac{\sum fX}{n}$$

$$= \frac{2(231{,}5) + 5(375{,}5) + 9(519{,}5) + 3(663{,}5) + 1(807)}{20}$$

$$= \frac{9.813{,}5}{20}$$

$$= 490{,}7$$

2. Modus

Modus adalah nilai yang diamati secara teratur. Simbol untuk modus adalah Mo. Modus disebut unimodus jika hanya ada satu nilai yang muncul. Bimodus digunakan bila ada dua atau lebih jenis nilai.

Modus menurut buku Matematika kelas XI (Kemendikbud, 2014: 7) adalah data yang paling sering muncul atau nilainya paling banyak. Pengertian modus menurut Kosko dan Amirruzzaman (2016:627) adalah "nilai yang paling sering muncul dalam suatu kumpulan data". Definisi ini, bagaimanapun, tidak cukup. Tujuannya lebih penting daripada mode, yaitu nilai yang paling sering muncul dan menunjukkan atau melambangkan kumpulan data. Berikut ciri-ciri modus (Boediono dan Koster, 2014: 67):

a. Tidak relevan dengan nilai ekstrem
b. Dapat digunakan dengan data homogen (tidak ada nilai ekstrem) maupun heterogen (data dengan satu atau lebih nilai ekstrem)

Menurut Ahmad Sudijono (2011:105), modus adalah skor atau nilai yang memiliki frekuensi tertinggi atau frekuensi maksimum dalam distribusi data. Berdasarkan sudut pandang tersebut, modus dapat didefinisikan sebagai nilai yang paling sering muncul atau memiliki frekuensi tertinggi dalam suatu kumpulan data.

Modus biasanya digunakan dengan kategori, ordinal, dan data diskrit. Pada kenyataannya, dengan data kategoris, seperti rasa es krim paling populer, modus adalah satu-satunya ukuran tendensi sentral yang dapat digunakan. Namun, tidak ada nilai sentral untuk data kategorikal karena kami tidak dapat mengurutkan pengelompokan. Modus data ordinal dan

diskrit mungkin merupakan nilai yang tidak terpusat. Mode ini menunjukkan nilai yang paling umum sekali lagi.

Modus digunakan untuk mengekspresikan fenomena yang paling umum. Mo adalah singkatan umum untuk mode atau mode. Satu mode (unimodus), dua mode (bimodus), lebih dari dua mode (multimodus), atau tidak ada mode sama sekali dapat ditemukan dalam data. Jika jumlah frekuensi pada data pengamatan sama, maka data tersebut tidak memiliki modus. Ada dua jenis modus yaitu modus data tunggal dan modus data berkelompok

a. Cara menghitung modus

1) Modus data tunggal

Cara menentukan modus untuk data belum perkelompok terbilan cukup mudah. Langkah pertama yang dilakukan adalah mengurutkan data dimulai dari data yang terkecil sampai data yang terbesar sehingga data-data yang memiliki nilai yang sama akan berdekatan satu sama lain. Langkah selanjutnya adalah mencari frekuensi dari masing masing nilai data.

Contoh :

Misalnya kita mendapatkan data yang disiapkan dengan tunggal yang disusun sebagai berikut. 3, 4, 7, 4, 5, 4, 5, 4, 12, 3. Tentukan modusnya!

Langkah pertama kita susun data tersebut maka didapat :3, 3, 4, 4, 4, 4, 5, 5, 7, 12 .

Kemudian tentukan frekuensi nya :

Nilai 3 memiliki frekuensi 2

Nilai 4 memiliki frekuensi 4

Nilai 5 memiliki frekuensi 2

Nilai 7 memiliki frekuensi 1

Nilai 12 memiliki frekuensi 1

Dengan demikian maka modusnya dalah nilai 4, yaitu nilai yang paling banyak muncul atau yang memiliki frekuensi terbesar.

2) Data perkelompok

Jika data telah dikelompokkan, dalam arti telah disajikan dalam bentuk distribusi frekuensi. Kelas modus sering disebut sebagai

kelas dengan frekuensi tertinggi. Rumus berikut digunakan untuk menentukan nilai modus:

$$Mod = L_0 + C \left\{ \frac{(f_1)_0}{f_1)_0 + f_2)_0} \right\}$$

Dimana :

L_0 = batas bawah untuk kelas dimana modus berada

C = interval kelas

$(f_1)_0$ = selisih frekuensi yang memuat modus dengan frekuensi kelas sebelumnya

$(f_2)_0$ = selisih frekuensi yang memuat modus dengan frekuensi kelas sesudahnya

Contoh : dari data yang disajikan dalam tabel frekuensi berikut ini, carilah modusnya :

Nilai	f
60 – 62	4
63- 65	10
66 - 68	17
69 - 71	9
72 – 74	5
Total	45

Jawab :

Dari tabel ditemukan frekuensi terbesar adalah 17, maka :

$(f_1)_0$ = 17- 10 = 7

$(f_2)_0$ = 17 – 9 = 8

L_0 = 65,5

Sehingga hasilnya adalah :

$$Mod = L_0 + C \{ \frac{(f_1)_0}{f_1)_0 + f_2)_0} \}$$

$$Mod = 65{,}5 + 3 \left\{ \frac{7}{7 + 8} \right\} = 66{,}9$$

b. Pengujian hipotesis

Sebagai peneliti, kita dapat menggunakan berbagai modus, termasuk:

1) Untuk mendapatkan nilai yang menunjukkan aturan rata-rata dalam jumlah waktu terkecil.
2) Peneliti dapat menghapus faktor akurasi sambil mencari hasil yang mencerminkan ukuran rata-rata, yang menyiratkan bahwa ukuran rata-rata hanya diperlukan sebagai perkiraan.
3) Peneliti hanya ingin menentukan karakteristik berdasarkan data yang diselidiki (data yang sedang dicari modusnya).

c. Kelebihan dan kekurangan modus

Seperti yang dapat dilihat dari uraian sebelumnya, kelebihan modus adalah dapat membantu kita memperoleh rata-rata dalam waktu sesingkat mungkin, yang merupakan ciri dari data yang sedang kita hadapi. Dengan kondisi

Kekurangannya adalah tidak akurat karena terlalu sederhana atau mudah diperoleh (achieve). Selanjutnya, jika frekuensi tertinggi distribusi lebih besar dari satu dan frekuensi data yang kita lihat lebih besar dari satu, kita akan menerima mode yang lebih besar dari buah. Skenario lain adalah bahwa kita tidak akan dapat menemukan atau menentukan mode dalam distribusi frekuensi karena semua skor memiliki frekuensi yang sama. Pada akhir hari, sebagai sebuah ukuran rata-rata, sedangkan modus sifatnya tidak stabil.

d. Interpretasi mengenai penggunaan modus.

Kebanyakan orang yang belum pernah mempelajari metode statistik mungkin memiliki pengertian “rata-rata” yang mirip dengan modus sebagai rata-rata. Hal ini terutama benar jika menyangkut konsep sifat kualitatif. Istilah "rata-rata" dan "biasanya" atau "paling sering" sering dipertukarkan. Ini bukan blunder. Pengusaha sering menghitung biaya produksi rata-rata dengan mengambil biaya produksi yang paling sering dikeluarkan oleh pengusaha selama periode tertentu daripada mencari biaya produksi rata-rata untuk periode tersebut. Pengusaha sering menggunakan modus sebagai dasar perhitungan daripada perhitungan rata-rata ketika merumuskan aturan tentang penimbunan dan pemesanan kembali barang.

Dalam penelitian ilmiah, bagaimanapun, rata-rata aritmatika yang tepat lebih disukai daripada mode. Mayoritas studi di bidang sains memerlukan pengukuran yang sangat akurat. Variabel kontinu digunakan untuk mendapatkan akurasi.

Secara umum, menggunakan variabel kontinu untuk perhitungan Hati-hati jarang menghasilkan nilai dengan besaran yang sama, menghasilkan bentuk frekuensi tunggal.

3. Median

Median adalah nilai sentral dari suatu distribusi frekuensi, menurut Anto Dajan (2008: 130). Dalam hal posisi pusat yang didudukinya dalam suatu distribusi, nilai seperti itu adalah nilai pusat. Rata-rata posisi adalah nama lain untuk median. Median membagi jumlah total pengamatan atau ukuran menjadi dua bagian yang sama dalam teori. Banyaknya nilai observasi dengan frekuensi lebih kecil dari median akan sama dengan jumlah nilai observasi dengan frekuensi lebih besar dari median. Median adalah suatu nilai yang berada di tengah data setelah diurutkan dari data yang terkecil sampai yang terbesar atau sebaliknya. Dengan kata lain, data yang ada kemudian dibagi menjadi dua bagian yang sama oleh median. Dalam perhitungan statistik, median dibagi menjadi dua bagian yaitu median untuk data tunggal dan median untuk data berkelompok.

a. Cara perhitungan median

1) Median dta tunggal

Jika jumlah suatu data (n) berjumlah ganjil, maka nilai mediannya adalah sama dengan data yang memiliki nilai di urutan paling tengah yang memiliki nomor urt k. Dimana untuk menentukan nilai k, dapat dihitung menggunakan rumus berikut:

$$k = \frac{n+1}{2}$$

Contoh :

Berikut adalah nilai ujian matematika dari 9 mahasiswa di mana masing-masing nilai secara berurutan adalah sebagai berikut:

90, 70, 60, 75, 65, 80, 40, 45, 50. Carilah berapa nilai median ?

Penyelesaian :
40, 45, 50, 60, 65, 70, 75, 80, 90.
$k = \frac{n+1}{2} = \frac{9+1}{2} = 5; med = X_5 = 65$

Selanjutnya, apabila jumlah n dari suatu data berjumlah genap, maka untuk menghitung mediannya dengan menggunakan rumus :
$k = \frac{n}{2}$

Contoh :
Data upah dari 8 karyawan yang dinyatakan dalam rupiah adalah sebagai berikut:
20, 80, 75, 60, 50, 85, 45, 90. Carilah nilai mediannya !.
Penyelesaian :
20, 45, 50, 60, 75, 80, 85, 90

$k = \frac{n}{2} = \frac{8}{2} = 4$
Maka mediannya terletak dirutan ke 4, dengan nilai :
Me $= \frac{1}{2}(X_K + X_{K+1}) = \frac{1}{2}(X_4 + X_5) = \frac{1}{2}(60 + 75) = 67{,}5$

2) Median data perkelompok

Untuk mencari median dari data perkelompok, dapat dilakukan dengan cara menggunakan bantuan dari angka frekuensi kumulatif kurang dari. Adapun untuk menghitung median data perekelompok, dapat dilakukan dengan rumus dibawah ini.

$$Med = L_0 + C\left\{\frac{\frac{n}{2} - (\sum f_1)_0}{f_m}\right\}$$

Dimana :

L_0 = batas bawah untuk kelas dimana median berada
C = interval kelas
$(\sum f_1)_0$ = jumlah frekuensi dari semua kelas dibawah kelas yang mengandung median
f_m = frekuensi dari kelas yang mengandung median

Contoh :

Carilah median dari data tabel frekuensi yang telah disajikan dibawah ini.

Nilai	F
60 – 62	4
63- 65	10
66 - 68	17
69 - 71	9
72 – 74	5
Total	45

Penyelesaian :

Nilai	F	fk
60 – 62	4	4
63- 65	10	14
66 - 68	17	31
69 - 71	9	40
72 – 74	5	45
Total	45	

$L_0 = \frac{65+66}{2} = 65{,}5$; C = 3; $(\sum f_1)_0 = 14$; $f_m = 17$

$$Med = L_0 + C\{\frac{\frac{n}{2} - (\sum f_1)_0}{f_m}\}$$

$$Med = 65{,}5 + 3\{\frac{\frac{45}{2} - 14}{17}\}$$

$$= 67$$

b. Pengaplikasian nilai median

Ketika seorang peneliti menemukan kondisi data berikut, nilai rata-rata atau median dapat digunakan.

1) Ketika peneliti tidak memiliki cukup waktu untuk menghitung nilai rata-rata (mean) atau tidak mampu melakukannya.

2) Ketika seorang peneliti menginginkan nilai rata-rata akurasi tinggi tetapi hanya ingin mengetahui skor atau nilai yang merupakan nilai tengah dari data yang mereka lihat.
3) Data yang kita hadapi memiliki distribusi frekuensi yang simetris (tidak normal). 4. Ukuran statistik lainnya tidak akan digunakan untuk memeriksa lebih lanjut data yang sedang kita lihat.

c. Kelebihan dan kekuarangan median

Karena teknik perhitungannya sederhana dan tidak rumit, median sebagai ukuran rata-rata memiliki keunggulan karena dapat mengumpulkan media dalam waktu singkat. Median memiliki kelemahan karena kurang tepat sebagai metrik rata-rata.

DAFTAR PUSTAKA

Anto Dajan. (2008). Pengantar Metode Statistik Jilid I. Jakarta: LP3ES.

Anas Sudijono. (2011). Pengantar Statistik Pendidikan. Jakarta: Rajawali Press.

James Popham dan Kennetha Sirotnik. (1973). Educational Statistics Use and Interpretation. New York: Harper and Row Publisher.

PPS UNY. (2011). Statistika: Matrikulasi S2 Program Pascasarjana UNY. Yogyakarta: UNY.

Akker, Jan Van Den, Gravemeijer K., McKenney S., danNieveen N..(2006). Educational Design Research. New York: Taylor and Francis Group.

Alghandi, Ahmed Hassan dan Li Li. (2013). Adapting Design-Based Research Methodology in Educational Settings. International Journal of Education Research:1-9

Maryati, Iyam.(2017). Analisis Kesulitan dalaam Materi Statistika ditinjau dari Kemampuan Penalaran dan Komunikasi Statistik. Jurnal PKISMA Universitas Suryakencana 6(2): 174-179.

Mumu, Jeinne. (2018). Desain Pembelajaran Materi Operasi Pada Himpunan Menggunakan Permainan Lemon Nipis. Jurnal Of Honai Math, Vol. 1 No. 1. pp. 14-23.

Abadyo, dkk. 2004. *Metoda Statistika Praktis*. Malang: Universitas Negeri Malang.

Anton Dajan. 1981. *Pengantar Metode Statistik Jilid I halaman 100-146"*. Jakarta : Lembaga Penelitian, Pendidikan dan Penerangan Ekonomi dan Sosial.

Hadi, Sutrisno. 2015. Statistik. Yogyakarta: Pustaka Belajar.

Ronald E.Walpole. 1993. *Pengantar Statistika, halaman 22-27"*. Jakarta: PT Gramedia Pustaka Utama. ISBN 979-403-313-8.

Suharyadi, SK, dan Purwanto. *Statistika Untuk Ekonomi dan Keuangan Modern Buku I*. 2003. Jakarta: Penerbit Salemba Empat.

PROFIL PENULIS

Nurdiana, S.P., M.Si., Lahir di Kabupaten Bone, 24 Maret 1982. SD hingga SMA beliau selesaikan di kota kelahirannya yaitu Kabupaten Bone. Beliau lulus S1 di Program Studi Agribisnis, Fakultas Pertanian dan Kehutanan, Universitas Hasanuddin Tahun 2006, kemudian melanjutkan studi S2 pada Program Studi Agribisnis Program Pascasarjana, Universitas Hasanuddin dan lulus pada Tahun 2011. Sekarang ini penulis merupakan Dosen Tetap Program Studi Pendidikan Ekonomi, Jurusan Ilmu Ekonomi Fakultas Ekonomi, Universitas Negeri Makassar. Selain itu, juga menjadi Dosen Luar Biasa di Sekolah Tinggi Ilmu Ekonomi Pembangunan Indonesia (STIE PI). Sebagai peneliti, telah menghasilkan beberapa artikel penelitian, yang terbit pada jurnal dan prosiding, baik yang berskala nasional mapun internasional. Sebagai dosen, telah menghasilkan beberapa buku, baik yang berupa buku ajar maupun buku referensi. Selain itu telah memiliki hak kekayaan intelektual berupa hak cipta. Penulis aktif berorganisasi sehingga saat ini juga merupakan anggota dari beberapa organisasi profesi dan keilmuan.

Email Penulis: diana@unm.ac.id

BAB 4
UKURAN PENYEBARAN DATA

Dr. Sri Astuty SE, M.Si
Universitas Negeri Makassar

A. PENGERTIAN

Ukuran Penyebaran Data adalah suatu ukuran dari serangkaian atau sekelompok data yang menunjukan seberapa jauh nilai nilai sekolompok data yang menyimpang dari nilai rata rata tersebut. Pada ukuran penyebaran data, kita akan mempelajari materi {Daerah Jangkauan (Range), Simpangan Rata-Rata / Ragam (variansi), Koefiensi Varians (KV), Pengukuran Penyebaran pada Nilai Kuartil, Desil, dan Persentil}.

B. DAERAH JANGKAUAN (*RANGE*)

Jangkauan sering kita sebut Range atau Rentang. Jangkauan didefinisikan sebagai selisih antara nilai terbesar dengan data terkecil tersebut, jangkauan disimbolkan jangkauan dengan huruf R. jangkauan berbagai 2 macam; data jangkauan tunggal dan data jangkauan kelompok.

Ada beberapa pendapat pendapat tentang rumus yang digunakan dalam range data kelompok yaitu**:**

1. Range adalah hasil dari pengurangan nilai tengah kelas terakhir dengan nilai tengah pertama. Nilai tengah ini cara mencarinya adalah menjumlahkan batas atas dan batas bawah kembudian dibagi dua.**(Anto Dayan, 1986)**
2. Range adalah hasil dari pengurangan tepi atas kelas terakhir dikurangi dengan tepi bawa kelas terakhir. Untuk tepi atas ditambah dengan 0,5 sedangkan tepi bawah dikurangi 0,5. **(Husein Tampomas, 2007)**
3. Range adalah hasil dari pengurangan dari batas atas kelas tertinggi dengan batas bawah terendah. **(Nugroho Budoyuwono, 1990)**

Kelebihan Range ini bisa digunakan untuk mengukur gambaran mengenai penyebaran data dengan waktu yang cukup singkat. Namun range juga memiliki kekurangan yaitu range tidak tepat untuk menentukan suatu perhitungan apalagi untuk data yang kompleks. **(Alvin, 2021)**

Data Jangkauan Tunggal

1. Rumus jangkauan tunggal (range) :

$$\mathbf{J = D_{max} - D_{min}}$$

Dimana :

J : Daerah Jangkauan

$\mathbf{D_{max}}$: Nilai terbesar dari serangkaian data

$\mathbf{D_{min}}$ **:** Nilai terbesar dari serangkaian data

Kasus :

Ada 8 siswa dari Jurusan Akutansi SMKN 4 MAKASSAR mengikuti sebuah tes perguruan tinggi / SBMTN dengan nilai 85,75,95,80, 82,90,91,77. Tentukan jangkauan nilai diatas ini !

Penyelesaian :

Data yang kita peroleh

D_{max}= 95

D_{min} = 75

Menentukan jangkauannya :$\mathbf{J = D_{max} - D_{min}}$

= 95 – 75

= 20

Jadi, jangkauan data tersebut adalah 20.

2. Data Jangkauan Kelompok

Rumus ini digunakan saat datanya tergolong dari nilai tertinggi diambil nilai tengah kelas tertinggi dan nilai terendah dari nilai kelas terendah. Dibawah ini rumusnya data jangkauan kelompok ;

$$\mathbf{J_k = D_{max} - D_{min}}$$

Dimana :

$\mathbf{J_k}$ = Daerah Jangkauan

$\mathbf{D_{max}}$ = Batas atas kelas dari kelas tertinggi

$\mathbf{D_{min}}$ = Batas bawah kelas dari kelas terendah

Contoh kasus

Carilah range dari data berikut!Diketahuilah nilai ujian kimia dari kelas X MIPA 2, ruang Lab. Kimia di SMAN 8 Makassar yang diikuti 21 siswa adalah sebagai berikut:

Interval	Frekuensi
41-48	4
51-58	3
61-68	7
71-78	4
81-88	3

Penyelesaiannya :

$\mathbf{J_k = D_{max} - D_{min}}$

$\mathbf{J_k} = 88 - 41$

$\mathbf{J_k} = 47$

Kelebihan Range ini bisa digunakan untuk mengukur gambaran mengenai penyebaran data dengan waktu yang cukup singkat. Namun range juga memiliki kekurangan yaitu range tidak tepat untuk menentukan suatu perhitungan apalagi untuk data yang kompleks.

C. SIMPANGAN RATA - RATA

Simpangan Rata Rata atau biasa disebut Deviasi Rata Rata. Simpangan rata rata didefinisikan sebagai harga mutlak (absoult) setiap nilai negative dianngap positif atau dalam artian ukuran yang menyatakan betapa / seberapa besar penyebaran tiap nilai data terhadap nilai meannya (rata-ratanya). Simpangan rata rata terbagi dua golongan yaitu simpangan rata rata data tunggal dan simpangan rata rata data kelompok.

1. Simpangan rata rata data tunggal

Rumus:

$$\boldsymbol{SR = \frac{\sum|D_i - \bar{D}|}{n}}$$

keterangan :

SR = Simpangan rata rata

Xd = Data pengamatan

$\bar{X}$ = Rata-rata data

Contoh:

Nilai uts mata kuliah statistika di jurusan ekonomi pembangunan kelas c 15 orang adalah sebagai berikut 60, 75, 45, 60, 90,75,75,60,45,90, 40,50, 65,55,60. Carilah simpangan rata-rata

a. Mencari nilai rata-rata

Rumus :

$$\bar{D} = \frac{\sum D_i}{n}$$

$$\bar{D} = \frac{60 + 75 + 45 + 60 + 90 + 75 + 75 + 60 + 45 + 90 + 40 + 50 + 65 + 55 + 60}{15}$$

$$\bar{D} = \frac{945}{63} = 63$$

b. Mencari selisih antara nilai D_i , dengan nilai rata-rata $(\bar{D})$

No	Nilai D_i	Rata-rata $(\bar{D})$	$\lvert D_i - D \rvert$
1	60	63	3
2	75	63	12
3	45	63	18
4	60	63	3
5	90	63	27
6	75	63	12
7	75	63	12
8	60	63	3
9	45	63	18
10	90	63	27
11	40	63	23
12	50	63	13
13	65	63	2
14	55	63	8
15	60	63	3
Jumlah			$\sum$ 184

c. Menghitung nilai simpangan rata-rata

$$SR = \frac{\sum|\boldsymbol{D_i} - \bar{\boldsymbol{D}}|}{n}$$

$SR = \frac{184}{15}$ = 12,26

Nilai simpangan rata-rata sebesar 12,26 dapat diartikan, bahwa terjadi penyimpangan sebesar 12,26 terhadap nilai rata-ratanya.

2. Simpangan rata-rata data kelompok

Rumus :

$$\boldsymbol{SR} = \frac{\sum \boldsymbol{f}|\boldsymbol{D}|}{\sum \boldsymbol{f}}$$

Dimana : $\left|\boldsymbol{X} - \bar{\bar{\boldsymbol{D}}}\right| = |\boldsymbol{D}|$

X = titik tengah

Contoh Kasus

Diketahui nilai ujian Matematika Ekonomi kelas C di Fakultas Ekonomi dan Bisnis yang diikuti 45 orang mahasiswa. Berapa daerah jangkauannya.

No Kelas	Interval Kelas	Frekuensi
1	21 – 30	8
2	31 – 40	7
3	41 – 50	12
4	51 – 60	11
5	61 – 70	6
		$\sum = 44$

Ditanya : Carilah nilai simpangan rata-ratanya

Dalam kasus seperti ini kita harus mencari beberapa poinnya seperti

a. *Menetukan titik tengahnya*

Titik tengah (**X**)

Pada kelas – 1

$$\mathbf{x} = \frac{(21 - 30)}{2} = 25{,}5$$

Begitupun selanjutnya sampai pada kelas – 5

b. Mengalikan frekuensi (*f*) dengan titik tengah **(x)**
Mengalikan **(*f*.x)** pada setiap kelas
Misalnya :
Kelas – 1
***f*.x** = 8 x 25,5
***f*.x** = 204

c. Menentukan rata-rata $(\bar{\mathbf{D}})$
Rumus dalam menentukan rata-rata $(\bar{\mathrm{D}})$ adalah

$$(\bar{\mathbf{D}}) = \frac{\sum f.\mathrm{x}}{\sum f}$$

$$(\bar{\mathbf{D}}) = \frac{2002}{44} = 45{,}5$$

d. Menemukan nilai $|\mathbf{D} - \bar{\mathbf{D}}|$
Menemukan nilai $|\mathbf{D} - (\bar{\mathbf{D}})|$ pada setiap kelas
Misalnya pada kelas – 1 seperti berikut :

$$|\mathbf{x} - \bar{\mathbf{x}}| = 25{,}5 - 45{,}5$$

e. Mengalikan frekuensi (*f*) dengan $|\mathrm{x} - \bar{\mathrm{x}}|$
Dilakukan pada semua kelas
Misalnya pada kelas – 1
$\boldsymbol{f}|\mathbf{x} - \bar{\mathbf{x}}|$ = 8 x 20 = 160
Berikut adalah tabel

Interval kelas	Frekuensi (*f*)	Titik Tengah **(x)**	$\boldsymbol{f}$ **. x**	$\|\mathbf{x} - \bar{\mathbf{x}}\|$	$\boldsymbol{f}\|\mathbf{x} - \bar{\mathbf{x}}\|$
21 – 30	8	25,5	204	20	160
31 – 40	7	35,5	248,5	10	70
41 – 50	12	45,5	546	0	0
51 - 60	11	55,5	610,5	10	110
61 – 70	6	65,5	393	20	120
	∑= 44		∑= 2002		∑= 460

f. Menentukan simpangan rata-rata
Menggunakan rumus : $\boldsymbol{SR} = \frac{\sum \boldsymbol{f}|\mathbf{D}-\bar{\mathbf{D}}|}{\sum \boldsymbol{f}}$

$$\boldsymbol{SR} = \frac{\sum \boldsymbol{f}|\mathbf{D}-\bar{\mathbf{D}}|}{\sum \boldsymbol{f}} = \frac{460}{44} = 10{,}45$$

D. SIMPANGAN BAKU (STANDAR DEVIASI)

Simpangan baku (standar deviasi) adalah nilai yang menunjukkan tingkat variasi kelompok data data atau ukuran standar penyimpangan dari nilai rata-ratanya. Simpangan baku (S) dari sekumpulan bilangan adalah akar dari jumlah deviasi kuadrat dari bilangan-bilangan tersebut dibagi dengan banyaknya bilangan atau akar dari rata-rata deviasi kuadrat**.**

1. Simpangan baku data tunggal

Rumus :

$$\mathbf{S} = \sqrt{\frac{\sum(\mathbf{D_i}-\bar{\mathbf{D}})^2}{n}} \text{ atau } \mathbf{S} = \sqrt{\frac{\sum \mathbf{D^2}}{n} - \left[\frac{\sum \mathbf{D}}{n}\right]^2}$$

Dimana :

S : *Simpangan baku*
D_i : *Data pengukuran*
$\bar{D}$: *Rata-rata nilai*
n : *Jumlah data*

Contoh Kasus

Data usia 7 orang penduduk di Desa Suka Makmur adalah sebagai berikut: 14, 35, 42, 28, 28, 21,21.

Tentukan nilai simpangan bakunya!

Cara kerjanya adalah sebagai berikut

a. Mencari nilai rata-rata ($\bar{\mathbf{D}}$)

   Dengan rumus : $\bar{\mathbf{D}} = \frac{14+35+42+28+28+21+21}{7} = \frac{189}{7} = 27$

b. Mencari selisih antara nilai ($\mathbf{D_i}$) dengan nilai rata-rata ($\bar{\mathbf{D}}$)

   Dengan rumus : $(\mathbf{D_i} - \bar{\mathbf{D}})$

   Untuk semua data, misalnya data – 1

   $(\mathbf{D_i} - \bar{\mathbf{D}})$ = (14 – 27)

   = -13

$\mathbf{D_i}$	$(\mathbf{D_i} - \bar{\mathbf{D}})$	$(\mathbf{D_i} - \bar{\mathbf{D}})^2$
14	-13	169
35	8	64
42	15	225
28	1	1
28	1	1
21	-6	36
21	-6	36
		$\sum$= 532

c. Menghitung nilai simpangan baku tunggal

$$\mathbf{S} = \sqrt{\frac{\sum(\mathbf{D}-\bar{\mathbf{D}})^2}{\boldsymbol{n}}} = \frac{532}{7} = \sqrt{76} = 8{,}7$$

2. Simpangan baku data kelompok

Simpangan baku data berkelompok mempunyai rumus sebagai berikut

$$\mathbf{S} = \sqrt{\frac{\sum \boldsymbol{f}\mathbf{D}}{\sum \boldsymbol{f}} - \left[\frac{\sum \boldsymbol{f}.\mathbf{D}}{\sum \boldsymbol{f}}\right]^2}$$

Dimana :

S : Simpangan Baku

F : Frekuensi

D : Titik Tengah

Contoh kasus :

Perhatikan tabel dibawah ini dan hitung simpangan bakunya

No Kelas	Interval Kelas	Frekuensi
1	21 – 30	8
2	31 – 40	7
3	41 – 50	12
4	51 – 60	11
5	61 – 70	6
		$\sum = 44$

Perhatikan langkah – langkah dalam menjawab

Dalam menyelesaikan hal ini kita perlu mencari beberapa hal penting yaitu sebagai berikut

a. Mencari titik tengah

Cara mencari titik tengah sama seperti mencari titik tengah pada simpangan rata-rata yaitu dengan

D = $\frac{21+30}{2} = 25{,}5$

Begitu pun dengan seterusya.

Interval	Frekuensi (*f*)	Titik tengah**(D)**	f**.D**	$\mathbf{D^2}$	$f.\mathbf{D^2}$
21 – 30	8	25,5	204	650,25	5.202
31 – 40	7	35,5	248,5	1.260,25	8.821,75
41 – 50	12	45,5	546	2.070,25	24.843
51 – 60	11	55,5	610,5	3.080,25	33.882,75
61 – 70	6	65,5	393	4.290,25	25.741,5
	$\sum = 44$		$\sum$= 2002		$\sum$= 98.491

b. Menentukan simpangan bakunya

$$S = \sqrt{\frac{\sum f\mathbf{D^2}}{\sum f} - \left[\frac{\sum f.\mathbf{D}}{\sum f}\right]^2}$$

$$S = \sqrt{\frac{98.491}{44} - \left[\frac{2002}{44}\right]^2}$$

$$S = \sqrt{2.238{,}4 - 2.070{,}25}$$

$$S = \sqrt{168{,}15}$$

$$S = 12{,}9$$

E. KOEFISIEN VARIANS (KV)

Koefisien Variasi (KV) adalah suatu sistem perbandingan antara simpangan standar dengan nilai hitung rata-rata yang dinyatakan dalam bentuk persentase. Sistem ini digunakan dalam mencari nilai rata-rata yang terdapat dalam suatu data kelompok.

Tujuan dilakukan perhitungan koefisien varians dalam suatu rangkaian data adalah untuk mengetahui tingkat keseragaman data semakin kecil nilai koefisien varians semakin seragam data tersebut, begitu juga dengan sebaliknya semakin besar nilai koefisien varians, semakin tidak seragam data tersebut. Adapun rumus dalam menentukan koefisien varians dalam suatu data adalah

$$\mathbf{KV} = \frac{\mathbf{S}}{\bar{\mathbf{D}}} \mathbf{x\ 100\%}$$

Dimana :

KV : Koefisien Varians

S : Simpangan Standar

$\bar{\mathbf{D}}$: Rata-rata

Contoh Kasus

Berikut ini adalah tabel nilai ujian matematika ekonomi di Fakultas Ekonomi Universitas Negeri Makassar. Tentukan koefisien variansnya

Interval kelas	Frekuensi (*f*)
21 – 30	8
31 – 40	7
41 – 50	12
51 – 60	11
61 – 70	6
	$\sum$= 44

Cara pengerjaannya

a. Mencari rata-ratanya

$$(\bar{\mathbf{D}}) = \frac{\sum f.\mathbf{D}}{\sum f}$$

$$(\bar{\mathbf{D}}) = \frac{2002}{44} = 45{,}5$$

b. Mencari nilai$|\mathbf{D} - \bar{\mathbf{D}}|$

Dengan cara mengurangi nilai D dengan $\bar{\mathrm{D}}$

$|\mathbf{D} - \bar{\mathbf{D}}|$= 25,5 – 45,5

= 20

Begitupun dengan seterusnya

c. Memangkat $|\mathbf{D} - \bar{\mathbf{D}}|$

Memangkatkan $|\mathbf{D} - \bar{\mathbf{D}}|$ untuk setiap interval

$|\mathbf{D} - \bar{\mathbf{D}}|^{\mathbf{2}} = (20)^2$

$|\mathbf{D} - \bar{\mathbf{D}}|^{\mathbf{2}} = 400$

d. Mengalikan frekuensi (*f*) dengan $|\mathbf{D} - \bar{\mathbf{D}}|$

Ini dilakukan untuk setiap interval kelas

$f|\mathbf{D} - \bar{\mathbf{D}}|^{\mathbf{2}} = 8 \times 400$

$= 3.200$

Untuk lebih jelasnya perhatikan tabel dibawah ini !

Interval kelas	Frekuensi (*f*)	Titik Tengah **(D)**	$\boldsymbol{f}\,.\,\mathbf{D}$	$\|\mathbf{D} - \bar{\mathbf{D}}\|$	$\|\mathbf{D} - \bar{\mathbf{D}}\|^{\mathbf{2}}$	$f\|\mathbf{D} - \bar{\mathbf{D}}\|^{\mathbf{2}}$
21 – 30	8	25,5	204	20	400	3.200
31 – 40	7	35,5	248,5	10	100	700
41 – 50	12	45,5	546	0	0	0
51 - 60	11	55,5	610,5	10	100	1.100
61 – 70	6	65,5	393	20	400	2.400
	∑= 44		∑= 2002			∑= 7.400

a. Menghitung nilai standar deviasi

$$S = \sqrt{\frac{\sum(\boldsymbol{f}|\mathbf{D} - \bar{\mathbf{D}}|^{\mathbf{2}})}{n - 1}}$$

$$S = \sqrt{\frac{7.400}{43}}$$

$$S = \sqrt{172,09}$$

$$S = 13,11$$

b. Menghitung nilai Koefisien varians

Dengan menggunakan rumus $KV = \frac{S}{\bar{D}} \times 100\%$

$$KV = \frac{13,11}{45,5} \times 100\%$$

$$KV = 28,8\%$$

F. KUARTIL

Kuartil adalah sekumpulan data yang telah disusun mulai dari yang terkecil sampai yang terbesar. Kuartil adalah nilai-nilai yang membagi data yang telah diurutkan kedalam empat bagian yang sama banyak. Kuartil dibagi menjadi 4 sub bab yang sama banyak dan telah disusun dari yang terkecil sampai yang terbesar.

Kuartil dibagi menjadi dua jenis yaitu kuartil data tunggal dan kuartil data kelompok. Dalam kuartil ini ada tiga yang perlu dicari yaitu kuartil bawah (Q1), Kuartil tengah (Q2), dan juga kuartil atas (Q3).Selain dari pada yang ada dibuku, adapun beberapa pandangan para ahli mengenai kuartil.

1. Kuartil adalag nilai yang membagi data yang tersaji atau sering disebut dengan distribusi frekuensi menjadi empat yang sama yaitu kuartil pertama,kedua dan ketiga. (Wirawan, 2001:105)
2. Kuartil berarti membagi sebuah kelompok data menjadi empat bagian yang pertama sampai bagian keempat. (Suliyanto, 2002:106)
3. Kuartil adalah titik atau skor atau nilai yang membagi keseluruhan frekuensi kedalam 5 sub-bagian yang masing masing sebesar 1/4N. (Sudijono,2006:112)

1. Kuartil data tunggal

Langkah-langkah dalam menghitung kuartil dari serangkaian data tunggal

a. Susunlah data mulai dari data yang terkecil ke data yang terbesar
b. Tentukan letak kuartil
c. Tentukan nilai kuartil

Letak kuartil ke *n*, diberi simbol LQ dan niliai kuartil ditentukan dengan rumus sebagai berikut : $\boldsymbol{Q_n = \frac{n(q+1)}{4}}$

Dimana :

Q_n ***: Nilai kuartil***
LQ ***: Letak kuartil***
q ***: data***

Contoh kasus

Data usia 8 pendaftar guru honorer di sekolah XY sebagai berikut : 24,26,24,22,25,25,27,23. Berapakah nilai kuartil bawah, tengah, dan atas data tersebut?

Langkah langkah dalam menjawab

a. Menyusun data mulai dari yang terkecil sampai ke terbesar
   22, 23, 24,24,25,25,26,27
b. Menentukan letak kuartil ke...i
   Letak kuartil (LQ) =1,2,3
c. Menghitung nilai kuartil bawah (Q_1)

   $$Q = \frac{\boldsymbol{n(q+1)}}{\boldsymbol{4}}$$

   $Q_1 = \frac{1(8+1)}{4}$ = 2,25

   Letak K_1 terletak diseperempat dari data kedua dan ketiga sehingga nilai K_1 adalah
   Q_1= 23 + 0,25 (24-23)
   Q_1= 23 + 0,25
   Q_1= 23,25

d. Menghitung nilai kuartil tengah (Q_2)

   $$Q_i = \frac{\boldsymbol{n(q+1)}}{\boldsymbol{4}}$$

   $Q_2 = \frac{2(8+1)}{4}$ = 4,5
   Letak Q_2 terletak di pertengahan data keempat dan data kelima sehinggan nilai Q_2adalah
   Q_2 = 24 + 0,5 (25-24)
   Q_2 = 24 + 0,5
   Q_2 = 24,5

e. Menghitung nilai kuartil atas (Q_3)

   $$Q_i = \frac{\boldsymbol{n(q+1)}}{\boldsymbol{4}}$$

$Q_3 = \frac{3(8+1)}{4}$ = 6,75

Letak Q_3 terletak di antara tiga seperempat dari data keenam dan data ketujuh, sehingga nilai K_3 adalah

Q_3 = 25 + 0,75 (26-25)

Q_3 = 25 + 0,75

Q_3 = 25,75

2. Kuartil data kelompok

Rumus kuratil data kelompok

$$\mathbf{K_i = B_0 + y\left(\frac{\frac{nq}{4} - ft}{f}\right)}$$

Dimana :

K_i = Nilai kuartil

B_0 = Batas bawah kelas kuartil

y = Panjang kelas

ft = Jumlah frekuensi semua kelas sebelum kelas kuartil

f = Frekuensi kelas kuartil

Contoh kasus

Diketahui nilai ujian mahasiswa jurusan ekonomi pembangunan dalam mata kuliah statistika adalah sebagai berikut

Interval kelas	Frekuensi (*f*)
21 – 30	3
31 – 40	8
41 – 50	11
51 – 60	12
61 – 70	7
71 – 80	6
81 – 90	7
	∑= 54

Hitunglah nilai kuartil bawah, tengah, dan atasnya!
Langkah-langkah dalam menjawab :

a. Menghitung nilai kuartil bawah (Q_1)

Mencari interval kuartil dengan menggunakan rumus : $\boldsymbol{K_i = \frac{n}{4}\ (q)}$

$$K_1 = \frac{1}{4}(54) = 13{,}5$$

Maka untuk menghitung nilai kuartil bawah adalah : $\mathbf{K_i = B_0 + y\left(\frac{\frac{n.q}{4} - ft}{f}\right)}$

$\mathrm{K_1} = 40{,}5 + 10\left(\frac{\frac{1.54}{4} - 11}{11}\right)$ = 40,5 + 10 (0,27) = 43,2

b. Menghitung nilai kuartil tengah (K_2)

Mencari interval kuartil dengan menggunakan rumus : $\boldsymbol{K_i = \frac{n}{4}(q)}$

$$K_2 = \frac{2}{4}(54) = 27$$

Maka untuk menghitung nilai kuartil tengah adalah : $\mathbf{K_i = B_0 + y\left(\frac{\frac{n.q}{4} - ft}{f}\right)}$

$$\mathrm{K_i} = 50{,}5 + 10\left(\frac{\frac{2.54}{4} - 22}{12}\right) = 50{,}5 + 10\ (0{,}41) = 54{,}6$$

c. Menghitung nilai kuartil atas (K_3)

Mencari interval kuartil dengan menggunakan rumus : $\boldsymbol{K_i = \frac{n}{4}(q)}$

$$K_3 = \frac{3}{4}(54) = 40{,}5$$

Maka untuk menghitung nilai kuartil atas adalah : $\mathbf{K_i = B_0 + y\left(\frac{\frac{n.q}{4} - ft}{f}\right)}$

$$\mathrm{K_3} = 60{,}5 + 10\left(\frac{\frac{3.54}{4} - 34}{7}\right) = 60{,}5 + 10\ (0{,}92) = 69{,}7$$

G. DESIL

Desil adalah metode kuantitatif untuk membagi satu set data peringkat menjadi 10 sub bagian yang sama besar. Jenis peringkat data ini dilakukan sebagai bagian dari banyak studi akademi dan statistik di bidang keuangan dan ekonomi. Ada beberapa pendapat ahli yaitu sebagai berikut

1. Sugiyono,(2007) adalah sebagai titik atau skor atau nilai yang membagi seluruh distribusi frekuensi dari data yang diselidiki kedalam 10 bagian yang sama besar yaitu masing-masing sebesar 1/10 N
2. Desil adalah sebagai nilai-nilai yang membagi seangkatan dana atau suatu distribusi frekuensi menjadi sepuluh bagian yang sama
3. Desil terjadi apabila sekumpulan data itu dibagi menjadi 10 bagian yang sama maka akan didapat sembilan pembagi dan setiap bagian disebut desil

Rumus yang digunakan adalah

$\boldsymbol{D_n = \frac{n(q+1)}{10}}$ **untuk data tunggal**

$\boldsymbol{D_i = B_0 + y\left(\frac{\frac{n.q}{10} - ft}{f}\right)}$ **untuk data kelompok**

Dimana :

$\boldsymbol{D_n}$ ***: Nilai Desil***
$\boldsymbol{B_0}$ ***: Batas bawah kelas Desil***
$\boldsymbol{y}$ ***: Panjang kelas***
$\boldsymbol{ft}$ ***: Jumlah frekuensi semua kelas sebelum Desil***
$\boldsymbol{f}$ ***: Frekuensi kelas Desil***

Contoh kasus :

1. Berikut ini nilai untuk nilai dari 8 mahasiswa jurusan teknik kimia adalah sebagai berikut : 95, 80,90,65.55,60,85,70,75. Tentukan D_2, D_4, dan D_7!
   a. Menentukan letak desil

$$D_2 = \frac{2(8+1)}{10} = 1{,}8$$

$$D_4 = \frac{4(8+1)}{10} = 3{,}6$$

$$D_7 = \frac{7(8+1)}{10} = 6{,}3$$

b. Mencari nilai desil D_2, D_4, dan D_7

$D_2 = 55 + 0{,}8(60 - 55) = 59$

$D_4 = 65 + 0{,}6(70 - 65) = 68$

$D_7 = 85 + 0{,}3(90 - 85) = 86{,}5$

2. Berikut ini daftar nilai dari mahasiswa jurusan pendidikan akuntansi seperti ini

Interval kelas	Frekuensi (*f*)
21 – 30	3
31 – 40	8
41 – 50	11
51 – 60	12
61 – 70	7
71 – 80	6
81 – 90	7
	∑= 54

Tentukan D_1,D_3, dan D_6!

Langkah-langkah pengerjaan

a. Mencari kelas yang mengandung desil

D_1= 5,4

D_3 = 16,2

D_6 = 32,4

b. Menetukan batas bawah, menentukan panjang kelas, jumlah semua frekuensi sebelum desil, dan menetukan frekuensi dikelas desil.

Batas bawah kelas **(P_0)**

- Batas bawah kelas untuk D_1 = 31 – 0,5 = 30,5
- Batas bawah kelas untuk D_3 = 41 – 0,5 = 40,5
- Batas bawah kelas untuk D_6 = 51 – 0,5 = 50,5

Panjang kelas adalah masing-masing 10 **(y)**

Jumlah semua frekuensi sebelum desil **(ft)**

- Untuk D_1 = 3
- Untuk D_3 = 11
- Untuk D_6 = 22

frekuensi dikelas desil (f)

- Untuk $D_1 = 8$
- Untuk $D_3 = 11$
- Untuk $D_1 = 12$

c. Menghitung desil D_1,D_3, dan D_6

- $\boldsymbol{D_i = B_0 + y\left(\frac{\frac{n.q}{10} - ft}{f}\right)}$

$$D_1 = 30{,}5 + 10\left(\frac{\frac{1.54}{10} - 3}{8}\right) = 30{,}5 + 10\ (0{,}3) = 33{,}5$$

- $\boldsymbol{D_i = B_0 + y\left(\frac{\frac{n.q}{10} - ft}{f}\right)}$

$$D_3 = 40{,}5 + 10\left(\frac{\frac{3.54}{10} - 11}{11}\right) = 40{,}5 + 10\ (0{,}47) = 45{,}2$$

- $\boldsymbol{D_i = B_0 + y\left(\frac{\frac{n.q}{10} - ft}{f}\right)}$

$$D_6 = 50{,}5 + 10\left(\frac{\frac{6.54}{10} - 22}{12}\right) = 50{,}5 + 10\ (0{,}87) = 59{,}2$$

H. PERSENTIL

Persentil adalah istilah dalam statistika untuk yang membagi kelompok data menjadi seratus bagian yang sama rata. Jadi, terdapat 99 nilai persentil yang membagi data menjadi seratus bagian. Persentil biasa dilambangkan dengan huruf P.

Persentil ialah nilai yang terbagi menjadi 100 data bagian yang sama, setelah tersususn dari nilai yang terkecil sampai yang terbesar. **(Riduawan, 2009:114).** Persentil juga merupakan nilai yang sekumpulan data dibagi menjadi seratus sub bagian yang sama. **(Andi, 2007:85**). Adapun rumus nya sebagai berikut :

$\boldsymbol{P_n = \frac{n(q+1)}{100}}$ **Data Tunggal**

$\boldsymbol{P_n = B_0 + y\left(\frac{\frac{n.q}{100} - ft}{f}\right)}$ **Data Kelompok**

Dimana :

P_n : Nilai Persentil

B_0 : Batas bawah kelas Persentil

y : Panjang kelas

ft : Jumlah frekuensi semua kelas sebelum Persentil

f : Frekuensi kelas Persentil

Contoh kasus

Data usia 8 masyarakatberbeda usia untuk dilakukan penelitian oleh mahasiswa tingkat akhir, berikut ini datanya : 9,12,14,16,19,21,23,26

Tentukan P_{20}!

Langkah-langkah menjawab

a. Menentukan interval persentil

$$P_{20} = \frac{\boldsymbol{n(q+1)}}{\boldsymbol{100}} = P_{20} = \frac{20(8+1)}{100} = 1{,}8$$

b. Menentukan nilai persentil P_{20}

$P_{20} = 9 + 0{,}8\ (12 - 9)$

$P_{20} = 9 + 2{,}4$

$P_{20} = 11{,}4$

Berikut ini tabel nilai terendah di kelas Ibu Darma selaku pengampu materi Matematika ekonomi sebagai berikut

Interval Kelas	Frekuensi
10 – 19	3
20 – 29	6
30 – 39	11
40 – 49	14
50 – 59	12
60 – 69	8
70 – 79	3
Jumlah	∑=57

Tentukan nilai P_{45},P_{61}, dan P_{80}!

Cara menjawab contoh kasus berikut

b. Menentukan interval kelas

$$P_{45} = \frac{45}{100}(57) = 25{,}65$$
$$P_{61} = \frac{61}{100}(57) = 34{,}77$$
$$P_{80} = \frac{80}{100}(57) = 45{,}6$$

c. Menentukan batas bawah kelas persentil, panjang kelas, jumlah semua frekuensi sebelum persentil, dan menetukan frekuensi dikelas persentil

d. Batas bawah kelas**(B_0)**
   - Batas bawah untuk P_{45} = 40 – 0,5 = 39,5
   - Batas bawah untuk P_{61} = 50 – 0,5 = 49,5
   - Batas bawah untuk P_{80} = 50 – 0,5 = 49,5

e. Panjang masing-masing kelas adalah 10 **(y)**

f. Jumlah semua frekuensi sebelum persentil **(ft)**
   - Untuk P_{45} = 20
   - Untuk P_{61} = 34
   - Untuk P_{45} = 34

g. Frekuensi dikelas persentil **(f)**
   - Untuk P_{45} = 14
   - Untuk P_{61} = 46
   - Untuk P_{45} = 46

h. Menentukan nilai persentil menggunakan rumus : $\boldsymbol{P_n = B_0 + y\left(\frac{\frac{n.q}{100} - ft}{f}\right)}$
   - $P_{45} = 39{,}5 + 10\left(\frac{\frac{45.57}{100} - 20}{14}\right) = 39{,}5 + 10(0{,}4) = 43{,}5$
   - $P_{61} = 49{,}5 + 10\left(\frac{\frac{61.57}{100} - 34}{46}\right) = 49{,}5 + 10(0{,}01) = 49{,}6$
   - $P_{45} = 49{,}5 + 10\left(\frac{\frac{80.57}{100} - 34}{46}\right) = 49{,}5 + 10(0{,}2) = 51{,}5$

DAFTAR PUSTAKA

Alvin, M. A. (2021, juni 14). Pengertian Range dalam Statistik dan Penjelasan Lebih Lanjutnya.

Andi Supangat (2007), *Statistik Dalam Kajian Deskriptif, Inferensial, Dan Nonparametric*, Edisi Pertama, Jakarta : Kencana

Anwari, A. (2016, oktober 06). Makalah Statistik : Kuartil, Desil, dan Persentil.

Ir. Syofian Siregar, M. (2018). Statistika Deskriptif untuk Penelitian. (hal. 41). Jakarta: Rajawali Pers.

Kho, D. (2020). Pengertian Kuartil beserta Rumus dan contoh Kuartil

Riduawan, (2009), *Skala Pengukuran Variabel-Variabel Penelitian*, Bandung CV:Alfabeta

Sugiyono,(2007) *Memahami Penelitian Kualitatif. Bandung* CV: Alfabeta

PROFIL PENULIS

Dr. Sri Astuty SE, M.Si adalah Dosen Jurusan Ilmu Ekonomi Prodi Ekonomi Pembangunan Fakultas Ekonomi Dan Bisnis Universitas Negeri Makassar. Lulus Sarjana Ilmu Ekonomi Dan Studi Pembangunan Universitas Hasanuddin tahun 2002, Magister Ekonomi Sumberdaya Universitas Hasanuddin tahun 2006, dan Doktor Ilmu Ekonomi Universitas Hasanuddin tahun 2017.

Beberapa matakuliah yang pernah diampu adalah: Pengantar Statistik, Matematika Ekonomi I, Matematika Ekonomi II, Ekonomi Sumber Manusia dan Ketenagakerjaan, Ekonomi Sumberdaya Alam, Ekonometrika II, Pasar Modal, Keuangan International, dan Ekonomi Moneter. Beberapa artikel yang pernah dipublikasikan adalah: Do You Trust Your Transformation Leader? A study of Civil State Apparatus, (2022), Impact Assessment of the Covid-19 Outbreak on Indonesian Tourism (2021), Does Service Quality In Education And Training Process Matters? Study Of Government"s Human Resource Agencies In Indonesia (2020), Menanamkan Karakter Abad 21 Untuk siswa SMA (2019),Analisis Faktor-faktor Yang Mempegaruhi Take Home Pay Dosen Di Kota Makassar (2019)

Beberapa penelitian yang pernah dilakukan adalah: Strategi Pemberdayaan Wanita Nelayan dalam Rangka Peningkatan Ekonomi Rumah Tangga Nelayan Tradisional Pantai Selatan Kabupaten Jeneponto (Hibah Bersaing 2014), Pola Konsumsi Dosen Wanita Pada Universitas Negeri Makassar Di Kota Makassar (Dosen Pemula 2015) Pengembangan Model Pemanfaatan Waktu Luang / Leisure Time Dosen di Perguruan Tinggi Negeri di Kota Makassar (Hibah Disertasi Doctor 2016), PKM Diversifikasi Ikan Bandeng (2020). Buku yang pernah ditulis adalah: Matematika Ekonomi II (2021). Bookchapter: Pengantar Statistik (2022), Ekonomi Teknik (2022), dan Manajemen Pemasaran (2022).

BAB 5
DISTRIBUSI FREKUENSI

Dodiet Aditya Setyawan, SKM.,MPH.
Politeknik Kesehatan Kementerian Kesehatan Surakarta

A. PENGERTIAN

Distribusi Frekuensi merupakan bagian dari teknik analisis deskriptif pada Statistik Deskriptif. Sedangkan Statistik Deskriptif pada dasarnya adalah statistik yang digunakan untuk menggambarkan atau menganalisis data statistik hasil survey atau penelitian tetapi tidak digunakan atau ditujukan untuk membuat kesimpulan yang lebih luas atau dalam arti kata tidak untuk melakukan Generalisasi atau Inferensi. Sehingga Statistik Deskriptif hanya berfungsi untuk mengorganisasi, menganalisa serta memberikan pengertian tentang data baik yang menunjukkan tentang kedaan, gejala atau persoalan/permasalahan dalam bentuk angka supaya dapat meberikan gambaran yang teratur, jelas dan ringkas. Teknik Analisis dan Penyajian Data pada Statistik Deskriptif dapat berupa:

1. Distribusi Frekuensi
2. Tendensi Sentral atau Ukuran Pemusatan
3. Dispersi Data atau Ukuran Persebaran data

Secara khusus pada Bab V Buku ini akan diuraikan secara lengkap dan jelas tentang Tabel Distribusi Frekuensi berikut cara melakukan penghitungan atau menentukan nilai-nilai dalam pembuatan Tabel Distribusi Frekuensi.

Distribusi Frekuensi merupakan teknik penyusunan data dalam bentuk kelompok mulai dari yang terkecil sampai yang terbesar berdasarkan kelas-kelas interval dan kategori tertentu (Hasibuan,dkk.2009). Manfaat dari penyajian data dalam bentuk Distribusi Frekuensi ini adalah untuk menyederhanakan teknik penyajian data sehingga menjadi lebih mudah untuk dibaca dan dipahami sebagai bahan informasi. Tabel Distribusi Frekuensi disusun apabila jumlah data yang akan disajikan cukup banyak, sehingga bila data tersebut disajikan dengan menggunakan tabel biasa menjadi tidak efektif dan efisien serta kurang komunikatif. Disamping itu tabel distribusi frekuensi juga dapat dibuat untuk bahan dalam melakukan uji normalitas data. Beberapa hal yang perlu diperhatikan dalam Tabel Distribusi frekuensi adalah:

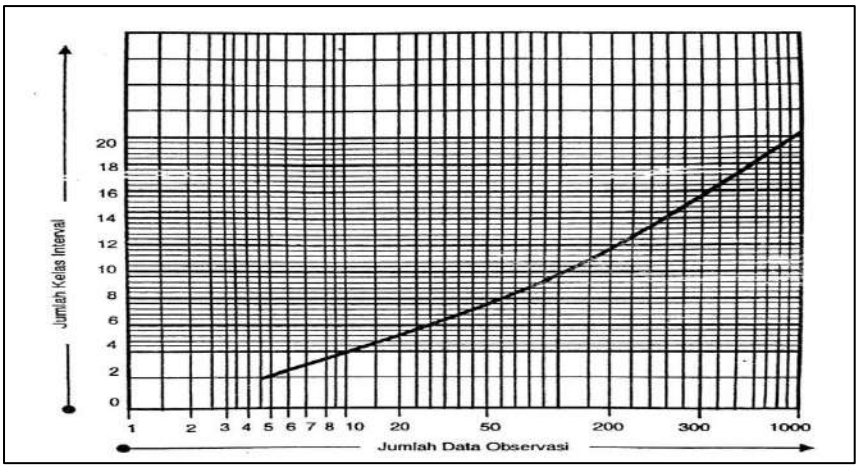
1. Tabel distribusi frekuensi mempunyai beberapa kelas yang jumlahnya disesuaikan dengan banyaknya data.
2. Setiap kelas pada tabel distribusi frekuensi mempunyai kelas interval. Setiap kelas interval mempunya panjang kelas yaitu interval kelas bawah sampai dengan interval kelas atas. Jadi panjang kelas interval merupakan jarak antara nilai batas bawah dengan batas atas pada setiap kelas.

3. Setiap kelas interval mempunyai jumlah atau yang sering disebut frekuensi.

B. PEDOMAN UMUM MEMBUAT TABEL DISTRIBUSI FREKUENSI

Membuat tabel distribusi frekuensi diawali dengan menentukan kelas interval dari sejumlah data yang sudah dikumpulkan atau di tabulasi. Berikut ini adalah bagian-bagian yang harus dibuat terlebih dahulu dalam sebuah tabel distribusi frekuensi:

1. Kelas Interval/Jumlah Kelas Interval (*Class*)
   Kelas interval merupakan kelompok-kelompok nilai atau variabel. Jumlah kelas menunjukkan jumlah kelompok nilai/variabel dari data yang diobservasi. Dalam menentukan Jumlah Kelas Interval terdapat 3 pedoman sebagai berikut:
   a. Ditentukan Berdasarkan Pengalaman
      Pada umumnya jumlah kelas interval yang dipergunakan dalam penyusunan Tabel Distribusi Frekuensi berkisar antara 6-15 kelas. Makin banyak data, maka makin banyak pula jumlah kelas intervalnya, tetapi jumlah yang paling banyak atau maksimal adalah 15 kelas interval dalam satu tabel distribusi frekuensi, sehingga tabel distribusi frekuensi tidak terlalu panjang.
   b. Ditentukan dengan Membaca Grafik `Jumlah Interval Kelas'
      Dengan menggunakan Grafik yang menunjukkan hubungan antara banyaknya data (n) dengan jumlah kelas interval yang diperlukan, maka penentuan jumlah kelas interval akan lebih cepat. Dimana dalam grafik tersebut, Garis Vertikal menunjukkan Jumlah Kelas Interval dan Garis Horisontal menunjukkan Jumlah Data Observasi. Misalnya, bila jumlah data yang diobservasi 50, maka berdasarkan Tabel, Jumlah Kelas Intervanya kurang lebih 8. Selanjutnya apabila jumlah data yang diobservasi sebanyak 200, maka jumlah kelas intervalnya kurang lebih 12, dan seterusnya.
      Contoh untuk menentukan jumlah kelas interval dengan cara tersebut dapat dilihat pada gambar grafik menentukan jumlah Kelas Interval berikut ini: (Sugiyono, 2015)



c. Ditentukan dengan Rumus Sturges ;

Jumlah Interval Kelas Interval juga dapat dihitung dengan menggunakan rumus Sturges sebagai berikut:

K = 1 + 3,3 Log. N

Dimana :

K = Jumlah Kelas Interval

n = Jumlah Data Observasi

Log = Logaritma

Contoh:

Misalnya Jumlah Data yang diobservasi sebanyak 150, maka jumlah Kelas Intervalnya dapat ditentukan dengan cara sebagai berikut:

K = 1 + 3,3. Log 150

K = 1 + 3,3. 2,17

K = 1 + 7,161

K = 8,161 --> Dibulatkan menjadi 8

Jadi dari sejumlah 150 data yang dikumpulkan dan diobservasi, maka jumlah kelas interval yang diperlukan adalah 8 kelas.

Berikut ini adalah gambaran tentang Kelas Interval (Jumlah Kelas Interval) dalam Tabel Distribusi Frekuensi:

Nilai UAS Statistik	Frekuensi (f)
50 – 55	3
56 – 61	7
62 – 67	2
68 – 73	8
74 – 79	4
80 – 85	6
Jumlah (n)	30

Jumlah Kelas Interval

2. Batas Kelas (*Class Limits*)

Batas Kelas merupakan nilai-nilai yang membatasi antara kelas yang satu dengan kelas berikutnya. Batas Kelas terdiri atas 2 macam, yaitu:

a. Batas Kelas Bawah (*Lower Class Limits*)

Batas Kelas Bawah adalah nilai atau angka yang terdapat pada bagian sebelah kiri dari setiap kelas.

b. Batas Kelas Atas (*Upper Class Limits*)

Batas Kelas Atas adalah nilai atau angka yang berada pada bagian sebelah kanan dari setiap kelas.

Gambaran tentang yang dimaksud dengan Batas Kelas Atas (*Upper Class Limits*) dan Batas Kelas Bawah (*Lower Class Limits*) pada Tabel Distribusi Frekuensi adalah sebagai berikut:

Nilai UAS Statistik	Frekuensi (f)
50 – 55	3
56 – 61	7
62 – 67	2
68 – 73	8
74 – 79	4
80 – 85	6
Jumlah (n)	30

Batas Kelas Atas

Batas Kelas Bawah

3. Rentang Data (*Range*)

Rentang Data (*Range*) adalah selisih antara data tertinggi dengsan data terendah (Data terbesar dikurangi Data terkecil). Perhatikan contoh tabel berikut ini:

Nilai UAS Statistik	Frekuensi (f)
50 – 55	3
56 – 61	7
62 – 67	2
68 – 73	8
74 – 79	4
80 – 85	6
Jumlah (n)	30

Pada Contoh Tabel di atas, dimana data terendah adalah 50 dan data tertinggi adalah 85. Sehingga untuk menentukan Rentang Data (*Range*) adalah 85-50 = 35. Jadi rentang data pada tabel tersebut adalah 35.

4. Panjang Interval Kelas

Panjang Interval Kelas atau disebut juga Panjang Kelas atau Interval Size merupakan jarak antara tepi kelas atas dengan tepi kelas bawah. Dapat dihitung dengan cara membagi Rentang Data dengan Jumlah Kelas. Perhatikan contoh tabel berikut ini:

Nilai UAS Statistik	Frekuensi (f)
50 – 55	3
56 – 61	7
62 – 67	2
68 – 73	8
74 – 79	4
80 – 85	6
Jumlah (n)	30

Pada contoh Tabel di atas, sudah diketahui bahwa rentang data adalah 35 dan jumlah kelas adalah 6, maka Panjang Interval kelasnya dapat dihitung dengan cara Rentang dibagi Jumlah kelas, sehingga didapatkan 35/6 = 5,8 (Dibulatkan = 6). Jadi Panjang Kelas interval adalah 6 pada setiap kelas.

5. Frekuensi Kelas (*Class Frequency*)
   Frekuensi kelas merupakan banyaknya jumlah data yang terdapat pada kelas tertentu. Misalnya pada contoh tabel di atas, Frekuensi pada kelas interval 50-55 adalah 3; frekuensi pada kelas interval 56-61 adalah 7, frekuensi pada kelas interval 80-85 adalah 6, dan seterusnya. Perhatikan contoh berikut ini:

Nilai UAS Statistik	Frekuensi (f)
50 – 55	3
56 – 61	7
62 – 67	2
68 – 73	8
74 – 79	4
80 – 85	6
Jumlah (n)	30

Batas Kelas Atas

C. TEKNIK PENYUSUNAN TABEL DISTRIBUSI FREKUENSI

Penyusunan atau pembuatan tabel distribusi frekuensi harus dilakukan secara berurutan atau sistematis. Langkah-langkah untuk membuat sebuah Tabel Distribusi Frekuensi secara sistematis dapat dilakukan sebagai berikut:

1. Mengurutkan semua data yang diobservasi mulai dari yang terkecil sampai yang terbesar.
2. Menghitung Rentang/Range (R), yaitu Data terbesar dikurangi dengan Data terkecil.
3. Menentukan jumlah kelas, dengan menggunakan rumus Sturges:K = 1 + 3,3. Log n
4. Menghitung Panjang Kelas atau Interval, dengan rumus: Panjang Kelas (P) = Rentang (R) : Jumlah Kelas

5. Membuat tabel distribusi frekuensi sementara yang terdiri atas kolom Interval Kelas, Tally, dan Frekuensi.

Kelas Interval	Tally	Frekuensi (f)

6. Menghitung jumlah Frekuensi dengan Tally atau melidi dalam Kolom Tally sesuai dengan banyaknya data.

Kelas Interval	Tally	Frekuensi (f)
50 – 55	III	3
56 - 61	~~IIII~~ II	7
		dan seterusnya

7. Setelah jumlah keseluruhan Frekuensi ditemukan, kemudian kolom Tally dihilangkan dalam Penyajian Data sehingga terbentuk Tabel Distribusi Frekuensi yang dimaksud.

Kelas Interval	Frekuensi (f)
50- 55	3
56 – 61	7
	dan seterusnya

D. MACAM-MACAM TABEL DISTRIBUSI FREKUENSI

Secara Umum pembuatan Tabel Distribusi Frekuensi dapat dilakukan dengan beberapa macam, tergantung pada banyaknya data dan tujuan dari penyajian data. Adapun macam-macam tabel distribusi frekuensi adalah sebagai berikut:

1. Tabel Distribusi Frekuensi Data Tunggal

   Tabel Distribusi Frekuensi Data Tunggal yaitu jenis tabel distribusi frekuensi yang menyajikan frekuensi dari data tunggal yang berdiri sendiri atau tidak dikelompokkan menjadi beberapa kelas atau interval.

Contoh:

Tabel Distribusi Frekuensi Nilai Ujian Akhir Semester Mata Kuliah Statistik Semester II adalah sebagai berikut:

Nilai	Frekuensi (f)
50	5
60	10
70	15
80	10
90	5
100	5
Jumlah (n)	50

Data Tunggal

2. Tabel Distribusi Frekuensi Data Kelompok

Tabel Distribusi Frekuensi Data Kelompok merupakan tabel distribusi frekuensi yang menyajikan frekuensi dari data yang telah dikelompokkan dalam beberapa kategori atau kelas.

Contoh:

Tabel Distribusi Frekuensi Umur Mahasiswa pada Perguruan Tinggi

Umur	Frekuensi (f)
22 – 27	40
28 – 33	20
34 – 39	5
dst	dst
Jumlah (n)	

Data Dikelompokkan

3. Tabel Distribusi Frekuensi Kumulatif

Distribusi Frekuensi Kumulatif merupakan tabel statistik yang menyajikan frekuensi dari data yang dihitung dengan ditambah-tambahkan baik dari atas ke bawah maupun dari bawah ke atas. Dibedakan menjadi 2, yaitu:

a. Frekuensi Kumulatif Atas atau $\mathbf{fk_{(a)}}$

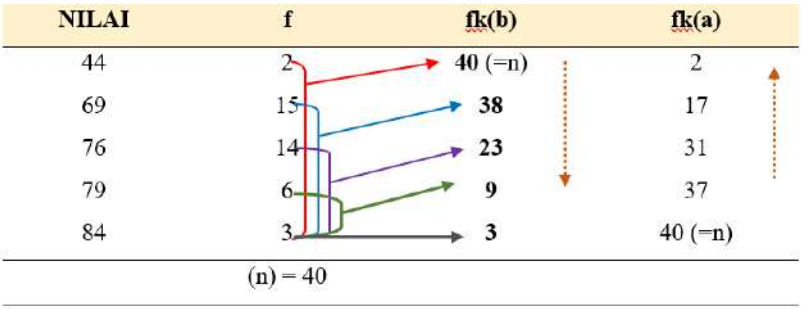
Yaitu Frekuensi yang angka-angkanya ditambahkan dari Bawah ke Atas.

b. Frekuensi Kumulatif Bawah atau **$fk_{(b)}$**

Yaitu Frekuensi yang angka-angkanya ditambahkan dari Atas ke Bawah, Berikut ini disajikan contoh Tabel Distribusi Frekeunsi Kumulatif beserta cara menentukan **$fk_{(a)}$** dan **$fk_{(b)}$**.

Tabel Distribusi Frekuensi Kumulatif Nilai Statistik

NILAI	f	fk(b)	fk(a)
44	2	40 (=n)	2
69	15	38	17
76	14	23	31
79	6	9	37
84	3	3	40 (=n)
	(n) = 40		



Dari contoh tabel di atas, cara mendapatkan fk(b) dapat dijelaskan sebagai berikut:

- fk(b) = 40, didapatkan dari penjumlahan semua frekuensi (f): 2+15+14+6+3 = 40
- fk(b) = 38, didapatkan dari penjumlahan (f) ke-2 - ke-5: 15+14+6+3 = 38
- fk(b) = 23, didapatkan dari penjumlahan frekuensi (f) ke-3 - ke-5: 14+6+3 = 23
- fk(b) = 9, didapatkan dari penjumlahan frekuensi (f) ke-4 - ke-5: 6+3 = 9
- fk(b) = 3, didapatkan dari penjumlahan frekuensi (f) ke-5: yaitu = 3.

Sedangkan cara untuk mendapatkan fk(a) dapat dijelaskan sebagai berikut:

- fk(a) = 40, didapatkan dari penjumlahan semua frekuensi (f): 3+6+14+15+2= 40
- fk(a) = 37, didapatkan dari penjumlahan frekuensi (f) ke-4- ke-1: 6+14+15+2 = 37
- fk(a) = 31, didapatkan dari penjumlahan frekuensi (f) ke-3- ke-1: 14+15+2 = 31
- fk(a) = 17, didapatkan dari penjumlahan frekuensi (f) ke-2- ke-1: 15+2 = 17

❖ fk(a) = 2, didapatkan dari penjumlahan frekuensi (f) ke-1: yaitu = 2

4. Tabel Distribusi Frekuensi Relatif (Prosentase)
   Tabel Distribusi Frekuensi Relatif adalah jenis tabel statistik yang di dalamnya menyajikan frekuensi dalam bentuk angka persentasi (p). Nilai Persentase dihitung dengan menggunkan rumus sebagai berikut:

$$\mathbf{P\ (\%) = \frac{Frekuensi}{Jml.\ Data} X100 \quad ATAU \quad \frac{f}{n} X100}$$

   Contoh:
   Tabel Distribusi Frekuensi Relatif Umur Mahasiswa

Umur	Frekuensi (f)	Persentase (%)
22-27	15	25
28-33	29	48
34-39	16	27
	(n) = 60	100

   Berdasarkan contoh tabel diatas, maka cara mendapatkan persentase tersebut dapat dijelaskan sebagai berikut:
   - ❑ Pesentase 25 (25%) didapatkan dengan cara (15/60)x100 = 25%
   - ❑ Pesentase 48 (48%) didapatkan dengan cara (29/60)x100 = 48%
   - ❑ Pesentase 27 (27%) didapatkan dengan cara (16/60)x100 = 27%

E. TEKNIK MENENTUKAN DISTRIBUSI FREKUENSI DATA MENGGUNAKAN SPSS

Kegiatan statistika telah banyak digunakan dalam bebagai bidang atau disiplin ilmu, seperti pendidikan, kesehatan, ekonomi, industri, pertanian, dan bidang-bidang lain. Banyak fenomena-fenomena pada bidang-bidang tersebut yang tidak terlepas dari kegiatan statistika. Statistika sangat membantu dalam pengambilan data, pengolahan dan penyajian data serta pengujian hipotesis yang pada akhirnya dapat memberikan kesimpulan sebagai informasi yang sangat bermanfaat bagi bidang-bidang tersebut. Selama proses kegiatan statistika tersebut diperlukan ketepatan dan juga kecepatan dalam analisis data. Saat ini telah banyak dikembangkan program atau software yang mampu

membuat kegiatan statistika itu lebih tepat, akurat dan juga cepat. Salah satunya adalah SPSS (*Statistical package for the social sciences)*, yang merupakan *software* berbasis windows yang digunakan untuk menyelesaikan berbagai macam teknik pengolahan data dalam statistika.

Gambaran tentang distribusi frekuensi data juga dapat dilakukan menggunakan software SPSS tersebut. Untuk itu pada Bab ini juga dijelaskan secara teknis cara menentukan distribusi frekeunsi data menggunakan SPSS.

Contoh:

Hasil pengumpulan data pada suatu survey adalah sebagai berikut:

NOMOR	NAMA	JENIS KELAMIN	UMUR (Th)	PENDIDIKAN
1	Aditya	Laki-laki	30	S2
2	Probowati	Perempuan	25	S1
3	Indah	Perempuan	17	SMA
4	Vena	Perempuan	18	SMA
5	Valdi	Laki-laki	20	S1
6	Valen	Laki-laki	17	SMA
7	Doni	Laki-laki	30	S2
8	Fitria	Perempuan	25	S1
9	Setyawan	Laki-laki	30	S2
10	Angel	Perempuan	18	SMA

Berdasarkan hasil survey tersebut, maka kita dapat menentukan Distribusi Frekuensi data menggunakan SPSS dengan cara:

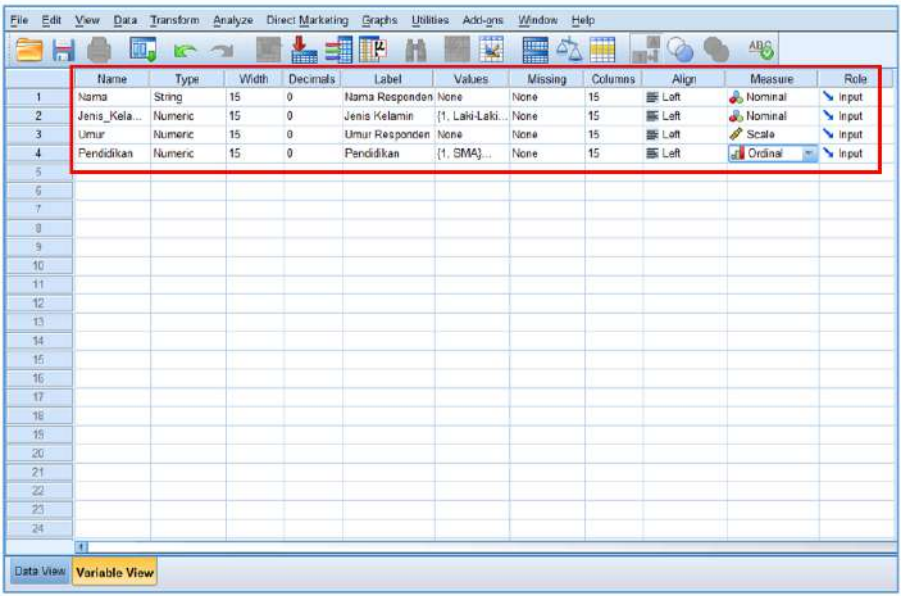
1. Entry Data tersebut ke dalam Program SPSS dengan langkah-langkah sebagai berikut:
   a. Mengisi Variabel View

      Pada contoh data di atas, terdapat variabel yang harus dimasukkan dalam Variabel View, yaitu: Nama (Type data String), Jenis Kelamin (Nominal), Umur (Scale), dan Pendidikan (Ordinal).

Kolom NAME	:	Isi dengan mengetikkan UMUR
Kolom TYPE	:	Isi dengan mengaktifkan pilihan NUMERIC
Kolom WIDTH	:	Isi dengan 15 (untuk keseragaman). Tergantung pada karakter variabel yg terpanjang.

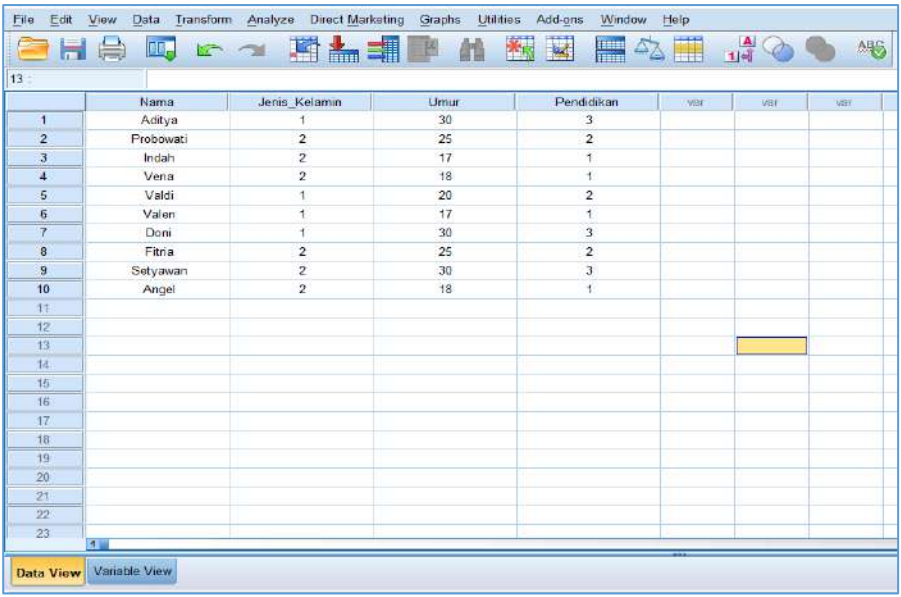
Kolom DECIMALS	:	Isi dengan 0
Kolom LABELS	:	Isi dengan mengetikkan UMUR RESPONDEN
Kolom VALUES	:	Tidak perlu diisi (Tidak ada Kategori)
Kolom COLUMN WIDTH	:	Isi dengan 15 (untuk keseragaman)
Kolom ALIGNMENT	:	Isi dengan Pilihan LEFT (untuk keseragaman)
Kolom MEASURES	:	Isi dengan pilihan SCALE (Istilah Numerik pada SPSS)

Lakukan entry data ke dalam Variabel View dari program SPSS tersebut, sebagaimana tampilan berikut ini:



b. Mengisi Data View

Setelah semua nama-nama variabel tersebut dimasukkan pada Variabel View, langkah selanjutnya adalah memasukkan data ke dalam worksheet Data View. Selanjutnya memasukkan data ke dalam kolom-kolom yang telah tersedia pada Data View sebagaimana terlihat pada tampilan gambar berikut:

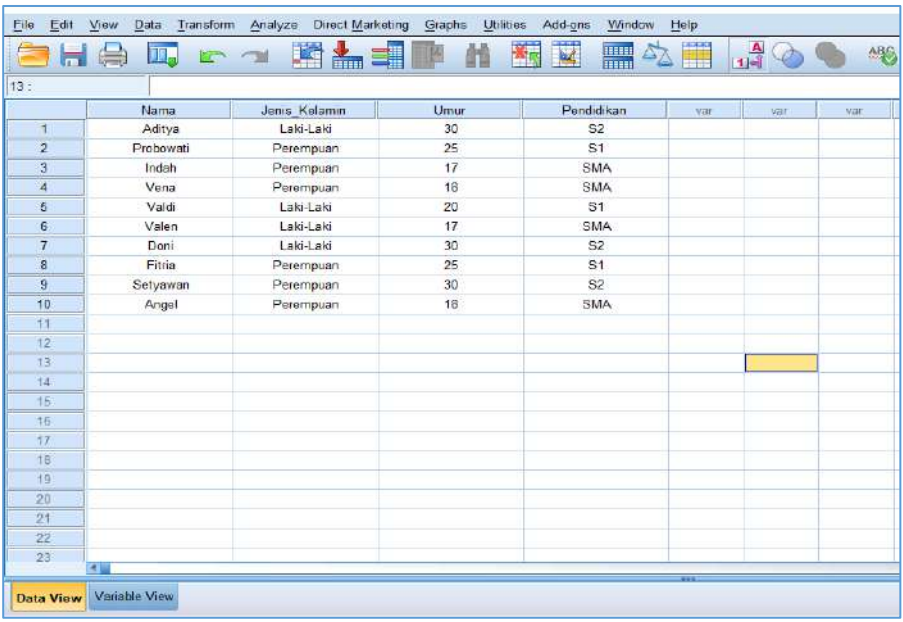


File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

13 :

	Nama	Jenis_Kelamin	Umur	Pendidikan	var	var	var
1	Aditya	1	30	3			
2	Probowati	2	25	2			
3	Indah	2	17	1			
4	Vena	2	18	1			
5	Valdi	1	20	2			
6	Valen	1	17	1			
7	Doni	1	30	3			
8	Fitria	2	25	2			
9	Setyawan	2	30	3			
10	Angel	2	18	1			
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							
23							

Data View | Variable View



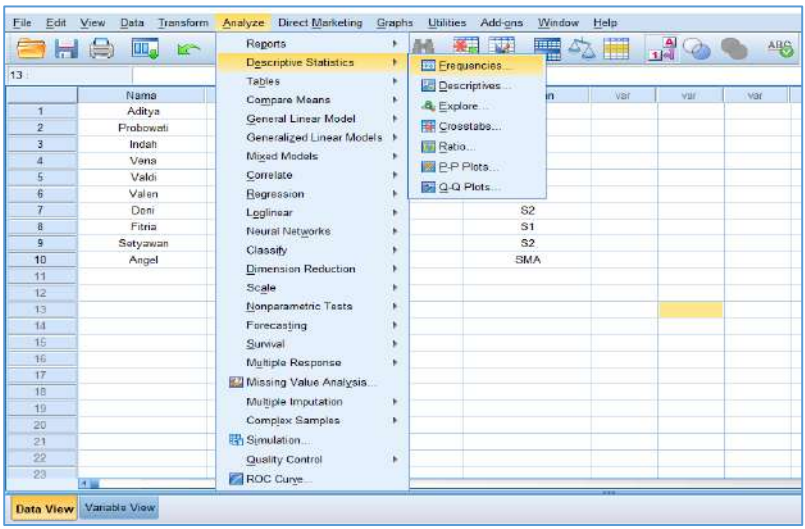
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

13 :

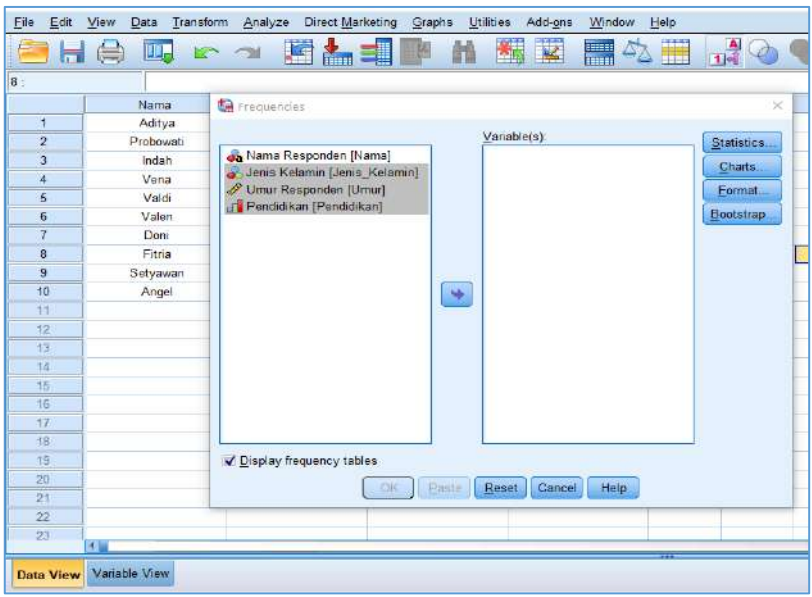
	Nama	Jenis_Kelamin	Umur	Pendidikan	var	var	var
1	Aditya	Laki-Laki	30	S2			
2	Probowati	Perempuan	25	S1			
3	Indah	Perempuan	17	SMA			
4	Vena	Perempuan	18	SMA			
5	Valdi	Laki-Laki	20	S1			
6	Valen	Laki-Laki	17	SMA			
7	Doni	Laki-Laki	30	S2			
8	Fitria	Perempuan	25	S1			
9	Setyawan	Perempuan	30	S2			
10	Angel	Perempuan	18	SMA			
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							
23							

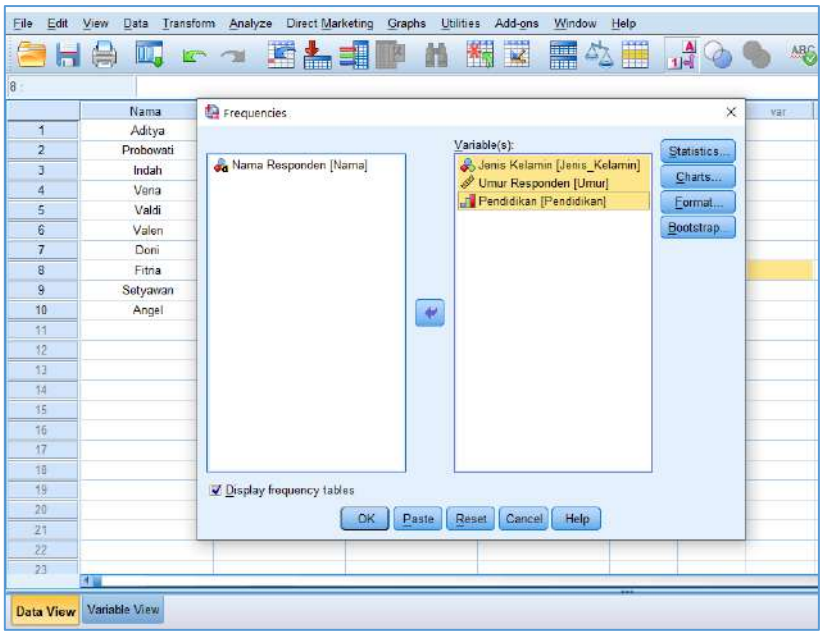
Data View | Variable View

2. Lakukan Analisis dengan langkah-langkah sebagai berikut:
   a. Klik Analyze kemudian pilih Descriptive Statistics dan selanjutnya klik Frequencies, sehingga muncul tampilan seperti berikut:

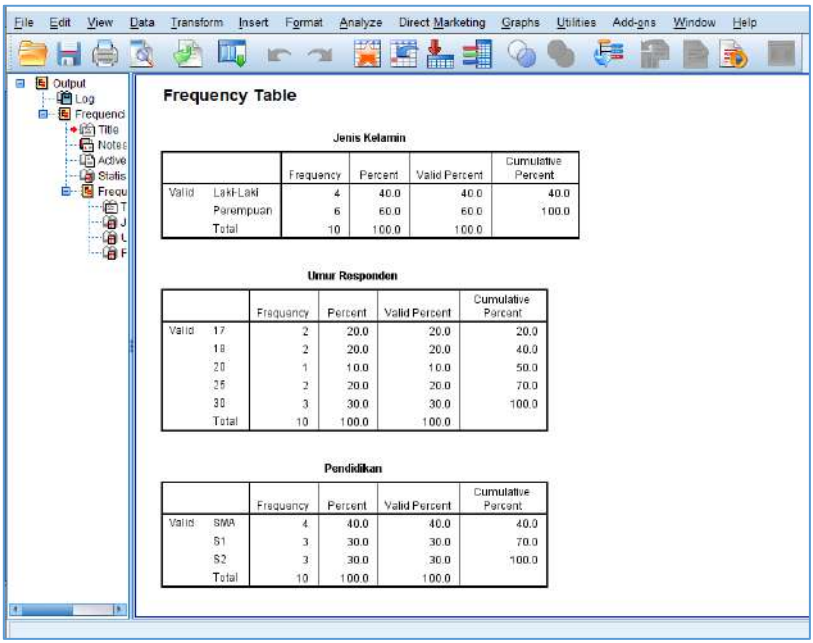


   b. Setelah klik Frequencies selanjutnya masukkan variabel yang akan dianalisis pada kotak Variable, sebagaimana dalam tampilan gambar berikut:





c. Kemudian Klik OK dan akan muncul hasil sebagai berikut:



d. Berdasarkan hasil atau Output SPSS tersebut, dapat dijelaskan bahwa:
   - ❑ Frequency: menunjukkan jumlah. Misalnya: Laki-Laki 4 orang.
   - ❑ Percent: menunjukkan prosentase. Misalnya: Perempuan 60%.
   - ❑ Valid Percent: menunjukkan prosentase untuk data yang terisi.
   - ❑ Cumulative Percent: menunjukkan jumlah kumulatif dari valid percent.

DAFTAR PUSTAKA

Amin.I., Aswin.A., Fajar.I., Isnaeni, Iwan.S., Pudjirahaju.A., Sunindya.R.. 2009. *Statistika untuk Praktisi Kesehatan*. Yogyakarta. Graha Ilmu

Dahlan.M.S. 2014. *Statistik untuk Kedokteran dan Kesehatan: Deskriptif, Bivariat dan Multivariat Dielngkapi Aplikasi dengan Menggunakan SPSS*, Edisi 6. Jakarta. Epidemiologi Indonesia.

Dahlan.M.S. 2012. *Statistik untuk Kedokteran dan Kesehatan*. Jakarta. Salemba Medika.

Dahlan.M.S. 2017. *Multiaksial Statistik Diagnostik dan Multiaksial Substansi Diagnosis Pintu Gerbang Memahami Epidemiologi, Biostatistik dan Metode Penelitian*. Edisi 2. Jakarta. PT. Epidemiologi Indonesia.

Hasibuan.A.A.,Supardi, Syah.D. 2009. *Pengantar Statistik Pendidikan*. Jakarta. Gaung Persada Press.

Riduwan.2010. *Dasar-dasar Statistika*. Bandung. Alfabeta.

Riwidikdo, H., 2013. *Statistik Kesehatan: Belajar mudah teknik analisis data dalam Penelitian Kesehatan (Plus Aplikasi Software SPSS)*. Yogyakarta. Mitra Cendikia Press.

Riwidikdo,H., 2012. *Statistik Kesehatan*. Yogyakarta. Nuha Medika

Santjaka, A. 2011. *Statistik Untuk Penelitian Kesehatan: Deskriptif, Inferensial, Parametrik dan Non Parametrik*. Yogyakarta. Nuha Medika.

Setyawan, D.A. 2022. *Buku Ajar Statistika Kesehatan: Analisis Bivariat pada Hipotesis Penelitian*. Klaten. Tahta Media Group

Siswandari. 2009. *Statistika (Komputer Based)*. Surakarta. LPP UNS dan UNS Press

Stang. 2018. Cara Praktis Penentuan Uji Statistik dalam Penelitian Kesehatan dan Kedokteran. Edisi 2. Jakarta. Mitra Wacana Media.

Sugiyono. 2015. *Statistika untuk Penelitian*. Bandung. Alfabeta

PROFIL PENULIS

Dodiet Aditya Setyawan, SKM.,MPH., lahir di Sragen, 12 Januari 1974. Penulis bertempat tinggal di Jalan Sukowati No. 164, Sragen Kulon, Sragen, Jawa Tengah. Mendapatkan gelar *Master of Public Health* (M.P.H) pada Program Studi S2 Ilmu Kesehatan Masyarakat di Fakultas Kedokteran Universitas Gadjah Mada Yogyakarta, pada Tahun 2014.

Berkarir sebagai Dosen di Politeknik Kesehatan Kementerian Kesehatan Surakarta (Polkesta) dengan Jabatan Lektor sampai dengan saat ini. Selain sebagai Dosen, penulis juga merangkap jabatan sebagai Sekretaris Jurusan Terapi Wicara sejak Tahun 2014 sampai sekarang.

Mata kuliah yang diampu oleh penulis pada saat ini diantaranya adalah Metodologi Penelitian, Statistika, Biostatistika, Sistem Informasi Kesehatan (SIK), Ilmu Kesehatan Masyarakat dan Promosi Kesehatan. Selain mengampu mata kuliah tersebut, penulis juga sangat tertarik dengan pemanfaatan aplikasi Sistem Informasi Geografis (SIG) pada Kesehatan Masyarakat. Hal ini ditunjukkan dengan dihasilkannya berbagai artikel ilmiah hasil penelitian terkait dengan SIG yang dimuat pada Jurnal Internasional bereputasi maupun Jurnal-Jurnal Nasional Terakreditasi. Karya-karya ilmiah dari penulis juga sudah banyak yang mendapatkan HKI baik dalam bentuk Artikel Ilmiah, Laporan Kegiatan Pengabdian Masyarakat, Poster Ringkasan Hasil Penelitian, Buku Ajar, Petunjuk Praktikum dan Modul.

BAB 6
PENGERTIAN POPULASI DAN SAMPEL

dr Prasaja STrKes., M.Kes
Politeknik Kesehatan Kementerian Kesehatan Surakarta

Pada suatu riset kesehatan masyarakat ditujukan untuk memperoleh kesimpulan umum yang valid tentang populasi manusia, bukan orang per orang atau kelompok kecil manusia. Persoalannya, tidak mungkin peneliti mengamati semua subyek dalam populasi yang sangat besar untuk membuat kesimpulan tentang karakteristik maupun fenomena yang ada pada populasi itu. Peneliti hanya dapat mengamati sebagian dari populasi besar, yang dinamakan sampel. Jika peneliti memilih sampel dengan tepat, maka penaksiran tentang distribusi dan hubungan paparan-penyakit tidak jauh meleset.

A. PENGERTIAN POPULASI

Populasi adalah wilayah generalisasi yang terdiri dari atas obyek/subyek yang mempunyai kualitas dan karakteristik tertentu yang ditetapkan oleh peneliti untuk dipelajari dan kemudian ditarik kesimpulannya. Jadi populasi bukan hanya orang, tetapi juga obyek dan benda – benda alam yang lain. Populasi bukan hanya jumlah obyek atau subyek, tetapi meliputi seluruh karakteristik dimiliki oleh obyek atau subyek tersebut (Salim, 2018); (Muhyi et al., 2018).

Misalnya akan melakukan penelitian di sekolah X, maka sekolah X ini merupakan populasi. Sekolah X mempunyai sejumlah orang/subyek dan obyek yang lain. Hal ini berarti populasi dalam arti jumlah/kuantitas. Tetapi sekolah X juga mempunyai karakteristik orang orangnya, misalnya motivasi kerjanya, disiplin kerjanya, kepemimpinannya, iklim organisasinya dan lain-lain; dan juga mempunyai karakteristik obyek yang lain, misalnya kebijakan, prosedur kerja, tata ruang kelas, lulusan yang dihasilkan dan lain-lain. Yang terakhir berarti populasi dalam arti karakteristik. Satu orang pun dapat digunakan sebagai populasi, karena satu orang itu memiliki berbagai karakteristik, misalnya gaya bicaranya, disiplin pribadi, hobi, cara bergaul, kepemimpinannya dan lain-lain (Muhyi et al., 2018).

Populasi (universe) ialah jumlah keseluruhan dari unit analisis yang ciri-cirinya akan diduga. Populasi dapat dibedakan antara populasi sampling dengan populasi sasaran. Sebagai contoh, jika seorang peneliti mengambil rumah tangga sebagai sampel; sedangkan yang diteliti hanya anggota rumahtangga (misalnya ayah atau suami), maka seluruh rumahtangga dalam wilayah penelitian disebut sebagai populasi sampling; sedangkan seluruh

suami atau ayah dalam wilayah penelitian itu dinamakan populasi sasaran (target population). Dalam setiap penelitian, populasi yang dipilih erat kaitannya dengan masalah yang akan diteliti. Sebagai contoh, (1) untuk penelitian tentang tenaga kerja, mestinya populasi yang dipilih adalah penduduk usia kerja, (2) untuk penelitian tentang pemilihan umum, mestinya populasi yang dipilih adalah penduduk yang memiliki hak pilih, (3) untuk penelitian tentang fertilitas, populasi yang dipilih adalah penduduk perempuan usia 15-49 tahun yang pernah kawin (Triyono, 2003).

Populasi adalah keseluruhan elemen/subyek riset (misalnya manusia). Populasi dapat terbatas atau tak terbatas. Populasi terbatas jika elemen-elemen dapat dihitung. Contoh: semua pria di Indonesia; semua wanita umur 15–49 tahun. Populasi tak terbatas jika elemen-elemen penelitian tak terhitung banyaknya. Contoh: jumlah eritrosit dalam tubuh manusia; jumlah orang yang positif HIV di Indonesia. Sesungguhnya tidak ada populasi yang tak terbatas. Persoalannya hanya ketidakmampuan menghitung elemen-elemen di dalam populasi, paling tidak dalam jangka waktu yang tersedia (Murti, 2015). Terdapat beberapa istilah populasi :

1. Populasi Sasaran

   Populasi sasaran (populasi target, reference population) merupakan keseluruhan subyek, item, pengukuran, yang ingin ditarik kesimpulan oleh peneliti melalui inferensi. Tujuan utama riset adalah untuk memperoleh gambaran distribusi (epidemiologi deskriptif) atau penjelasan tentang fenomena hubungan paparan-penyakit (epidmiologi analitik) yang terjadi pada populasi sasaran. Sejauh mana temuan-temuan tentang distribusi atau hubungan paparan-penyakit seperti ditunjukkan oleh statistik sampel adalah sahih untuk digunakan menarik inferensi tentang parameter yang sama pada populasi sasaran disebut validitas internal. Contoh: Dalam suatu studi kohor pengaruh kontrasepsi oral (OC) terhadap infark otot jantung (MI), populasi sasarannya adalah semua wanita Indonesia keturunan Cina berusia 25-49 tahun yang tinggal di perkotaan (Murti, 2015).

2. Populasi Sumber

   Populasi sumber (source population, actual population) merupakan himpunan subyek dari populasi sasaran yang digunakan sebagai sumber pencuplikan subyek penelitian. Contoh: Populasi sasaran studi OC dan

MI adalah wanita Indonesia keturunan Cina berusia 25□49 tahun yang tinggal di perkotaan. Maka peneliti dapat menentukan populasi sumber berupa populasi yang memenuhi kriteria tersebut, tinggal di sejumlah kota yang terpilih (dus tidak semua kota di Indonesia), dan mengunjungi klinik keluarga berencana (KB).

Prinsipnya, populasi sumber memiliki karakteristik yang sama dengan populasi sasaran. Jikalau terdapat karakteristik yang berbeda (misalnya, mengunjungi atau tidak mengunjungi klinik KB) maka harus diyakinkan perbedaan itu tidak berhubungan dengan pemakaian OC maupun kejadian MI. Jika kunjungan ke klinik KB berhubungan dengan pemakaian OC, dan peneliti memutuskan untuk menggunakan pengunjung klinik KB saja (dan tidak menggunakan bukan pengunjung klinik KB) sebagai populasi sumber, maka inferensi hasil studi hanya valid untuk "subset" populasi sasaran tersebut, yakni populasi wanita Indonesia keturunan Cina berusia 25-49 tahun yang tinggal di perkotaan dan pengunjung klinik KB (Murti, 2015).

3. Populasi Eksternal

Populasi eksternal (external population) adalah populasi yang lebih luas atau di luar populasi sasaran tetapi peneliti masih berminat membuat generalisasi (ekstrapolasi) temuan riset. Tujuan utama riset epidemiologi adalah memperoleh gambaran distribusi penyakit, atau penjelasan tentang hubungan paparan-penyakit pada populasi sasaran. Kadang-kadang peneliti masih ingin mengekstrapolasikan hasil risetnya di luar populasi sasaran – disebut populasi eksternal. Sejauh mana temuan-temuan tentang karakteristik (epidemiologi deskriptif) atau hubungan paparan-penyakit (epidemiologi analitik) seperti ditunjukkan oleh statistik sampel adalah sahih untuk digunakan menarik inferensi tentang parameter yang sama pada populasi eksternal disebut validitas eksternal (generalisasi).

Contoh: Andaikata sebuah studi menemukan bahwa pemakaian OC meningkatkan risiko terkena MI pada wanita Indonesia perkotaan keturunan Cina berusia 25–49 tahun, ada kemungkinan peneliti ingin memperluas kesimpulan itu bagi populasi wanita Indonesia pribumi perkotaan berusia 25–49 tahun – maka populasi ini merupakan populasi eksternal. Satu hal perlu diingat, generalisasi temuan penelitian kepada

populasi eksternal dilakukan berdasarkan keputusan (judgment) peneliti. Tidak ada satu perangkat statistikpun dapat digunakan untuk menentukan validitas eksternal (Murti, 2015).

Dari populasi sumber dapat dibuat kerangka pencuplikan (sampling frame), yaitu daftar semua subyek dalam populasi sumber yang digunakan sebagai basis pemilihan subyek ke dalam sampel.

Dalam menganalisis hubungan paparan-penyakit, peneliti umumnya tidak menyelidiki seluruh populasi, melainkan sebuah sampel dari populasi, untuk menarik kesimpulan (inferensi) tentang populasi itu. Pencuplikan (sampling) memberikan sejumlah keuntungan : (1) Mengurangi biaya penelitian; (2) Meningkatkan kecepatan pengumpulan dan analisis data; (3) Meningkatkan akurasi pengumpulan data karena berkurangnya volume kerja; (4) Memperluas perolehan informasi tentang berbagai faktor. Pendeknya pencuplikan memberikan cara praktis, cepat, dan ekonomis untuk memperoleh informasi yang diinginkan peneliti. Tetapi sebelum mengupas cara mencuplik sampel, perlu dipahami dulu konsep-konsep dasar terkait pencuplikan (Murti, 2015).

B. PENGERTIAN SAMPEL

Sampel adalah bagian dari jumlah dan karakteristik yang dimiliki oleh populasi tersebut. Bila populasi besar, dan peneliti tidak mungkin mempelajari semua yang ada pada populasi, misalnya karena keterbatasan dana, waktu dan tenaga (Muhyi et al., 2018). Sampel (study population) merupakan sebuah subset yang dicuplik dari populasi, yang akan diamati atau diukur peneliti. Pencuplikan sampel dapat dilakukan secara random atau non-random, dengan restriksi atau tanpa restriksi pemilihan subyek. Dalam studi epidemiologi dikenal kriteria restriksi pemilihan subyek yang disebut eligibility criteria – pernyataan eksplisit tentang syarat-syarat subyek untuk dapat dimasukkan ke dalam sampel. Kriteria eligibilitas terdiri dari kriteria inklusi (kriteria dimasukkan) dan kriteria eksklusi (kriteria dikeluarkan) (Murti, 2015).

Dalam suatu penelitian survei, tidak perlu untuk meneliti semua individu dalam suatu populasi, sebab di samping memakan biaya yang banyak, juga membutuhkan waktu yang lama. Dengan meneliti sebagian dari populasi, diharapkan hasil yang diperoleh akan dapat menggambarkan sifat populasi

yang bersangkutan. Untuk dapat mencapai tujuan ini, maka cara-cara pengambilan sebuah sampel harus memenuhi syarat-syarat tertentu (Triyono, 2003).

Sampel adalah bagian dari populasi yang ingin diteliti, yang ciri-ciri dan keberadaannya diharapkan dapat mewakili atau menggambarkan ciri-ciri dan keberadaan populasi yang sebenarnya. Sampling adalah suatu proses yang dilakukan untuk memilih dan mengambil sampel secara benar dari suatu populasi sehingga sampel tersebut dapat mewakili populasinya (Rawung, 2020).

C. METODE/ TEKNIK SAMPLING

Teknik sampling adalah teknik pengambilan sampel (Muhyi et al., 2018). Teknik pengambilan sampel yang ideal mempunyai sifat-sifat (1) dapat menghasilkan gambaran yang dapat dipercaya dari seluruh populasi yang diteliti, (2) dapat menentukan presisi (presicion) dari hasil penelitian dengan menentukan simpangan baku (standard deviation) dari taksiran yang diperoleh, (3) sederhana, sehingga mudah dilaksanakan, dan (4) dapat memberikan keterangan sebanyak mungkin, dengan biaya yang serendah-rendahnya (Triyono, 2003).

Dalam menentukan teknik pengambilan sampel yang akan diterapkan dalam suatu penelitian, seorang peneliti harus memperhatikan hubungan antara biaya, tenaga, dan waktu di satu pihak, serta tingkat presisi di pihak lain. Jika jumlah biaya, tenaga, dan waktu sudah dibatasi sejak semula, seorang peneliti harus berusaha mendapatkan teknik pengambilan sampel yang menghasilkan presisi tertinggi. Perlu disadari bahwa tingkat presisi yang tinggi tidak mungkin dapat dicapai dengan biaya, tenaga, dan waktu yang terbatas. Di samping itu, perlu diperhatikan pula masalah "efisiensi" dalam memilih teknik pengambilan sampel (Triyono, 2003).

Desain pencuplikan (sampling design) merupakan rancangan yang dibuat peneliti untuk memperoleh sampel dari seluruh anggota populasi. Desain pencuplikan merupakan bagian penting dari desain penelitian (research design), karena itu keduanya harus konsisten. Mengapa repot-repot merancang pencuplikan? Ada dua alasan untuk "repot". Pertama, memilih subyek penelitian secara gegabah akan mengakibatkan kesalahan sistematis

yang disebut bias seleksi (selection bias). Contoh: kelompok-kelompok studi yang akan diperbandingkan dalam studi analitik - baik studi potong-lintang, kasus-kontrol, kohor, maupun eksperimen – intinya harus sebanding dalam hal distribusi faktor-faktor di luar paparan, agar penilaian hubungan antara paparan dan penyakit tersebut valid. Kedua, ukuran sampel mempengaruhi presisi penelitian; ukuran sampel yang tidak cukup besar akan memperbesar kesalahan random (random error) (Murti, 2015).

Kategori desain pencuplikan berdasarkan dua kriteria – randomness dan restriksi pemilihan subyek. Berdasarkan kriteria random, cara pencuplikan dapat dibagi dua - pencuplikan random (pencuplikan probabilitas) dan pencuplikan non-random (pencuplikan non-probabilitas). Berdasarkan kriteria restriksi pemilihan subyek, cara pencuplikan dibagi dua – pencuplikan tanpa kriteria retriksi, dan pencuplikan dengan kriteria restriksi (Murti, 2015).

Teknik sampling pada dasarnya dapat dikelompokkan menjadi dua yaitu probability sampling (random) dan nonprobability sampling (non random) (Muhyi et al., 2018).

1. Pencuplikan random

   Probability sampling (pencuplikan random) adalah teknik pengambilan sampel yang memberikan peluang yang sama bagi setiap unsur (anggota) populasi untuk dipilih menjadi anggota sampel (Muhyi et al., 2018). Penggunaan prosedur pencuplikan random mengandung implikasi, setiap elemen dari populasi diketahui peluangnya untuk terpilih ke dalam sampel. Peluang tersebut tidak ditentukan dengan sengaja oleh peneliti, melainkan suatu "peluang buta" ("blind chance") seperti mengambil gulungan kertas lotere. Dengan cara demikian peneliti dapat mengetahui probabilitas hasil pencuplikan dan besarnya kesalahan estimasi – disebut sampling error atau sampling variation. Karakteristik itu merupakan keunggulan relatif pencuplikan random dibandingkan desain pencuplikan purposif. (Murti, 2015). Sebuah sampel harus dipilih sedemikian rupa sehingga setiap satuan unsur mempunyai kesempatan dan peluang yang sama untuk dipilih, dan besarnya peluang tersebut tidak boleh sama dengan nol. Di samping itu, pengambilan sampel secara acak (random) harus menggunakan teknik yang tepat sesuai dengan ciri-ciri populasi dan tujuan penelitian (Triyono, 2003).

Dalam pencuplikan random berlaku Hukum Regularitas Statistik (The Law of Statistical Regularity). Artinya, jika secara rata-rata sampel terpilih merupakan sampel random, maka sampel itu akan memiliki komposisi dan karakteristik populasi. Jadi prosedur pencuplikan random menghasilkan sampel yang representatif terhadap populasi (Murti, 2015). Berikut ini teknik probability sampling :

a) Simple Random Sampling

Dikatakan simple (sederhana) karena pengambilan anggota sampel dari populasi dilakukan secara acak tanpa memperhatikan strata yang ada dalam populasi itu. Cara demikian dilakukan bila anggota populasi dianggap homogeny (Muhyi et al., 2018). Pencuplikan random sederhana (simple random sampling) dari suatu populasi terbatas (finite population) merupakan metode pemilihan sampel dimana masing-masing item (elemen) dari keseluruhan populasi memiliki peluang yang sama dan independen (baca: tidak bergantung!) untuk terpilih ke dalam sampel. Pencuplikan random banyak digunakan dalam studi deskriptif untuk memberikan sampel yang representatif terhadap populasi. Pencuplikan itu biasanya dilakukan tanpa pengembalian (without replacement). Artinya, sekali sebuah item terpilih ke dalam sampel, maka item tersebut tidak dapat lagi muncul dalam proses pencuplikan item berikutnya. Sebaliknya, dalam pencuplikan dengan pengembalian (with replacement) elemen yang telah terpilih ke dalam sampel dikembalikan ke dalam populasi sebelum proses pencuplikan elemen berikutnya. Jadi elemen yang sama dapat muncul dua kali dalam satu sampel sebelum elemen berikutnya terpilih. Pencuplikan dengan pengembalian jarang dilakukan (Murti, 2015).

Sampel acak sederhana (simple random sampling) ialah suatu sampel yang diambil sedemikian rupa sehingga tiap unit penelitian dari suatu populasi mempunyai kesempatan yang sama untuk dipiih sebagai sampel. Dalam prakteknya, sampel acak sederhana dapat dilakukan dengan (a) undian, atau (b) bilangan acak (Triyono, 2003).

Satu cara untuk menjalankan prosedur random adalah memberikan nomer kepada setiap individu, mulai dari 0, 1, 2, 3, dan seterusnya. Lalu nomer-nomer itu dipilih secara random dengan menggunakan tabel angka

random atau program komputer sampai ukuran sampel diinginkan tercapai (Murti, 2015).

b) Pencuplikan random kompleks

Pencuplikan random kompleks (complex/mixed random sampling) merupakan pencuplikan random dengan restriksi, biasanya memadukan prosedur pencuplikan random dengan pencuplikan non-random. Studi deskriptif menggunakan pencuplikan random kompleks untuk mendapatkan sampel representatif dengan lebih efisien ketimbang pencuplikan random sederhana. Sedang studi analitik umumnya memadukan pencuplikan random dengan pencuplikan purposif (misalnya, fixed-exposure sampling, fixed-disease sampling). Motif di balik pencuplikan random kompleks dalam studi analitik adalah untuk memperoleh kelompok-kelompok penelitian yang emang "comparable" untuk diperbandingkan, dengan demikian meminimalkan bias pemilihan subyek penelitian (selection bias) (Murti, 2015).

Beberapa desain pencuplikan random kompleks yang populer sebagai berikut: (1) pencuplikan sistematis; (2) pencuplikan, klaster; (3) pencuplikan area; (4) pencuplikan berstrata (stratified sampling); (5) pencuplikan bertingkat (multi-stage sampling).

(1) Pencuplikan sistematis

Apabila banyaknya satuan elementer yang akan dipilih cukup besar, maka pemilihan sampel acak sederhana akan berat mengerjakannya. Dalam keadaan seperti ini ahli statistik cenderung memakai metode lain. Pengambilan sampel acak sistematis (systematic random sampling) ialah suatu metode pengambilan sampel, dimana hanya unsur pertama saja dari sampel dipilih secara acak, sedangkan unsur-unsur selanjutnya dipilih secara sistematis menurut pola tertentu (Triyono, 2003). Unsur prosedur random dapat diterapkan ke dalam jenis pencuplikan ini untuk memilih elemen pertama sampel. Setelah itu elemen-elemen berikut dipilih setiap interval 5, atau 10, atau 20, dan sebagainya, hingga jumlah elemen sampel yang diinginkan tercapai. Prosedur ini sangat praktis ketika kerangka pencuplikan sudah tersedia dalam bentuk daftar (Murti, 2015).

Ada pendapat bahwa pengambilan sampel dengan metode ini tidak acak, karena yang diambil secara acak unsur pertama saja, sedangkan unsur selanjutnya diurutkan berdasarkan interval yang sudah tertentu dan tetap. Karena itu, untuk dapat mempergunakan metode ini, harus dipenuhi beberapa syarat yakni (1) populasi harus besar, (2) harus tersedia daftar kerangka sampel, (3). populasi harus bersifat homogen (Triyono, 2003).

(2) Pencuplikan klaster

Pencuplikan klaster (cluster sampling) adalah metode pencuplikan dimana unit pencuplikan merupakan kelompok (baca: klaster) subyek (misalnya dukuh, atau anggota keluarga), bukannya individu. Pengamatan dilakukan terhadap seluruh individu dalam klaster terpilih. Dengan kata lain, variabel tetap diukur pada level individu. Pencuplikan klaster cocok digunakan jika populasi menempati area luas (Murti, 2015).

Jika seorang peneliti ingin meneliti besarnya pendapatan per bulan dari tiap-tiap keluarga di suatu kecamatan, sedangkan data mengenai jumlah keluarga di kecamatan tersebut tidak tersedia, maka kecamatan tersebut dibagi menjadi desa-desa. Desa-desa itu dijadikan gugus atau unsur sampling. Semua desa yang ada diberi nomor dan dipilih secara acak sebuah desa atau lebih sebagai sampel. Karena unsur penelitian adalah keluarga atau rumahtangga, maka semua rumahtangga yang ada dalam desa tersebut yang diteliti (Triyono, 2003).

(3) Pencuplikan area

Pencuplikan area (area sampling) merupakan metode pencuplikan yang dapat digunakan ketika anggota populasi tersebar dalam area luas. Dalam metode ini, seluruh area yang akan dicuplik dibagi dulu dalam sub-area, lalu sub-area diberi nomer dan dicuplik dengan menggunakan tabel angka random. Selanjutnya anggota-anggota dalam area yang dicuplik diberi nomer untuk menjalani pencuplikan tahap kedua. Pencuplikan area sebenarnya analog dengan pencuplikan klaster yang ditentukan berdasarkan pembagian geografis (Murti, 2015).

Teknik sampling daerah digunakan untuk menentukan sampel bila obyek yang akan diteliti atau sumber data sangat luas, missal penduduk dari suatu Negara, provinsi atau kabupaten. Untuk menentukan penduduk mana yang akan dijadikan sumber data, maka pengambilan sampelnya berdasarkan daerah populasi yang telah ditetapkan. Misalnya di Indonesia

terdapat 30 propinsi, dan sampelnya akan menggunakan 15 propinsi, maka pengambilan 15 propinsi itu dilakukan secara random. Tetapi perlu diingat, karena propinsi-propinsidi Indonesia berstrata (tidak sama) maka pengambilan sampelnya perlu menggunakan stratified random sampling. Propinsi di Indonesia ada yang penduduknya padat semacam ini perlu diperhatikan sehingga pengambilan sampel menurut strata populasi itu dapat diterapkan (Muhyi et al., 2018).

(4) Pencuplikan Berstrata

Pencuplikan berstrata (stratified sampling) merupakan teknik pencuplikan subyek di mana populasi sasaran pertama-tama dibagi dalam strata (subpopulasi) yang berbeda menurut karakteristik penting tertentu untuk penelitian bersangkutan, misalnya umur, status sosio-ekonomi, lalu dilakukan pencuplikan dari masing-masing stratum. Pencuplikan pada masing-masing stratum populasi biasanya dilakukan secara random, sehingga prosedur pencuplikan keseluruhan disebut pencuplikan random berstrata (stratified random sampling) (Murti, 2015).

Tujuan pencuplikan berstrata adalah untuk memperoleh kasus (studi kasus kontrol) atau subyek terpapar (studi kohor) dalam jumlah yang cukup pada masing-masing strata, sehingga kelak dapat dianalisis secara statistik. Variabel yang dilakukan stratifikasi adalah variabel yang berhubungan dengan penyakit atau paparan, misalnya umur, etnik, dan ras. Contoh: Peneliti berminat meneliti penyakit jantung koroner (PJK) di semua usia. Insidensi PJK pada usia muda lebih rendah daripada usia dewasa. Agar memperoleh jumlah kasus PJK yang cukup dari kelompok usia muda, maka peneliti perlu membagi populasi sasaran dalam strata umur. Kalau saja tidak dilakukan stratifikasi tetapi langsung melakukan pencuplikan random, maka - karena peran peluang - peneliti akan memperoleh jumlah PJK usia muda terlalu sedikit untuk bisa dianalisis secara statistik (Murti, 2015).

Dalam praktek sering dijumpai populasi yang tidak homogen. Makin heterogen suatu populasi, makin besar pula perbedaan sifat antara lapisan-lapisan tersebut. Presisi dan hasil yang dapat dicapai dengan penggunaan suatu metode pengambilan sampel, antara lain dipengaruhi oleh derajat keseragaman dari populasi yang bersangkutan.

Untuk dapat menggambarkan secara tepat mengenai sifat-sifat populasi yang heterogen, maka populasi yang bersangkutan dibagi ke dalam lapisan-lapisan (stratum) yang seragam dan dari setiap lapisan diambil sampel secara acak. Dalam sampel berlapis, peluang untuk terpilih satu strata dengan yang lain mungkin sama, mungkin pula berbeda.

Ada dua syarat yag harus terpenuhi untuk dapat mempergunakan metode pengambilan sampel acak berlapis, yaitu (a) ada kriteria jelas yang akan dipergunakan sebagai dasar untuk menstratifikasi populasi, dan (b) diketahui dengan tepat jumlah satuan-satuan elementer dari tiap lapisan dalam populasi itu. Besarnya sampel yang diambil dari tiap-tiap strata dapat berimbang dan dapat pula tidak berimbang. Dalam pengambilan sampel yang berimbang, unsur-unsur satuan yang diambil dari setiap strata berbanding lurus dengan jumlah satuan-satuan elementer dalam strata yang bersangkutan. Kalau peneliti akan mempergunakan metode tidak berimbang, maka ia dapat menentukan sendiri jumlah unsur-unsur sampel yang akan diambilnya (Triyono, 2003).

(5) Multi-stage sampling

Multi-stage sampling (pencuplikan bertingkat) merupakan teknik pencuplikan dimana peneliti mencuplik sampel melalui proses bertingkat-tingkat (strata hirarkis). Tahap pertama, peneliti membagi populasi ke dalam strata, dan mencuplik sampel dari strata di tingkat pertama tersebut. Tahap kedua, peneliti mencuplik dari sampel tingkat pertama untuk mendapatkan sampel tingkat kedua. Demikian seterusnya hingga terpilih unit-unit pencuplikan dari strata hirarkis terakhir. Tergantung jumlah tingkat, desain pencuplikan dapat bertingkat dua (two-stage sampling), bertingkat tiga (three-stage sampling), dan seterusnya.

Umumnya peneliti mencuplik unit-unit pencuplikan secara random di tiap-tiap tingkat – disebut multi-stage random sampling. Bila unit-unit pencuplikan itu merupakan klaster maka desain itu menjadi multi-stage random cluster sampling. Bila klaster ditentukan berdasarkan wilayah geografis, maka desain itu menjadi multi-stage random area sampling. Pencuplikan bertingkat pada umumnya memang dipilih tatkala populasi sasaran menempati suatu area geografis yang sangat luas, misalnya sebuah negara (Murti, 2015).

Tiga keuntungan pencuplikan bertingkat. Pertama, lebih mudah dilakukan daripada teknik satu tingkat umumnya, sebab kerangka pencuplikan bertingkat dibuat dalam unit-unit terpisah. Kedua, untuk anggaran yang sama, pencuplikan bertingkat menghasilkan jumlah sampel lebih besar daripada teknik pencuplikan sederhana. Ketiga, dapat memberikan data hirarkis yang selanjutnya dianalisis dengan analisis multilevel (multilevel analysis) menggunakan model multilevel (multilevel modelling).

2. Pencuplikan non-random

Nonprobability sampling pencuplikan non random) adalah teknik pengambilan sampel yang tidak memberi peluang atau kesempatan sama bagi setiap unsur atau anggota populasi untuk di pilih menjadi sampel (Muhyi et al., 2018). Prosedur pencuplikan non-random (non-random sampling, non-probability sampling) memilih subyek-subyek populasi ke dalam sampel tidak secara random, dengan kata lain tidak menggunakan Hukum Regularitas Statistik (Murti, 2015).

Pencuplikan non-random dapat dibagi dalam dua kategori: (1) Pencuplikan seenaknya (convenience sampling); dan (2) Pencuplikan purposif (purposive sampling).

(1) Pencuplikan seenaknya (convenience sampling)

Pencuplikan seenaknya (convenience sampling, haphazard sampling, grab sampling, accidental sampling) merupakan metode pencuplikan non-random yang dilakukan dengan "bebas" tanpa restriksi atau rencana khusus dari pihak peneliti. Pencuplikan "liberal" ini mudah dilakukan, semudah mencuplik sampel dari orang yang ditemui di jalan ("man-in-the-street") atau pengunjung sebuah stan bazar. Karena tidak ada diskresi obyektif dari pihak peneliti dalam mendesain sampel, maka teknik pencuplikan seenaknya cenderung mengintroduksi bias pencuplikan (sampling bias), dengan demikian validitas penarikan kesimpulan hasil kepada populasi sasaran lemah. Selain itu, sampel melalui pencuplikan seenaknya tidak representatif terhadap populasi (Murti, 2015).

Sampling insidental adalah teknik penentuan sampel berdasarkan kebetulan, yaitu siapa saja yang secara kebetulan atau insidental bertemu dengan peneliti dapat di gunakan sebagai sampel ,bila di

pandang orang yang kebetulan di temui itu cocok sebagai sumber data (Muhyi et al., 2018).

(2) Pencuplikan purposif (purposive sampling)

Sampling purposive adalah teknik penentuan sampel dengan pertimbangan tertentu,misalnya akan melakukan penelitian tentang kualitas makanan,maka sampel sumber datanya adalah orang yang ahli makanan,sampel ini lebih cocok di gunakan untuk penelitian kualitatif,atau penelitian-penelitian yang tidak melakukan generalisasi (Muhyi et al., 2018).

Pencuplikan purposif (purposive sampling, deliberate sampling) merupakan metode pencuplikan non-random dimana peneliti melakukan pendekatan terhadap masalah pencuplikan dengan rencana spesifik tertentu dalam benaknya sesuai dengan masalah dan hipotesis penelitian. Peneliti memiliki diskresi untuk memilih elemen dengan sengaja, tetapi pemilihan itu tidak dilakukan sembarangan melainkan dengan rencana tertentu sesuai dengan tujuan penelitian.

Karena unsur subyektif peneliti sangat kental dalam prosedur pencuplikan ini, dan probabilitas masing-masing elemen dalam populasi untuk terpilih ke dalam sampel tidak diketahui, maka prosedur ini tidak tepat untuk dipilih jika tujuan penelitian adalah mendeskripsikan karakteristik populasi dalam studi deskriptif. Jika dilakukan dengan hati-hati, prosedur ini sangat bermanfaat untuk mendapatkan kelompok-kelompok studi yang memiliki karakteristik “comparable” untuk diperbandingkan dalam studi analitik.

Jika seorang peneliti memilih metode pencuplikan purposif, maka ia harus bisa menjelaskan maksud “purposif” dan tujuan yang diharapkan dari memilih metode itu untuk penelitiannya.

Pencuplikan purposif mencakup sejumlah teknik pemilihan subyek sebagai berikut : (a) Fixed-exposure sampling; (b) Fixed-disease sampling; (c) Restriksi; (d) Pencocokan (matching); (e) Pencuplikan kuota; (f) Expert sampling; (g) Pencuplikan bola salju (snowball sampling) (Murti, 2015).

(a) Fixed-exposure sampling

Fixed-exposure sampling merupakan prosedur pencuplikan berdasarkan status paparan subyek, sedang status penyakit subyek

bervariasi mengikuti status paparan subyek yang sudah "fixed" tersebut (Gerstman, 1998). Ketika paparan di alam langka, maka prosedur pencuplikan berdasarkan status paparan anggota-anggota populasi akan memastikan jumlah subyek penelitian yang cukup dalam kelompok-kelompok terpapar dan tak terpapar. Fixed-exposure sampling paling umum dilakukan pada studi kohor.

(b) Fixed-disease sampling

Fixed-disease sampling merupakan prosedur pencuplikan berdasarkan status penyakit subyek, sedang status paparan subyek bervariasi mengikuti status penyakit subyek yang sudah "fixed" tersebut. Ketika penyakit di alam langka, maka prosedur fixed-disease sampling akan memastikan jumlah subyek penelitian yang cukup dalam kelompok-kelompok berpenyakit dan tak berpenyakit. Fixed-disease sampling paling umum dilakukan pada studi kasus-kontrol.

(c) Restriksi

Restriksi (restriction) merupakan proses mempersempit eligibilitas subyek potensial ke dalam sampel penelitian dengan menggunakan kriteria restriksi (kriteria eligibilitas, admissibility criteria). Dua jenis kriteria restriksi: (1) Kriteria inklusi menentukan subyek- subyek yang boleh dimasukkan ke dalam sampel penelitian; dan (2) Kriteria eksklusi menentukan subyek-subyek yang harus digusur ke luar sampel. Sampel dapat diperoleh dengan atau tanpa kriteria restriksi. Sampel yang diperoleh dengan restriksi disebut sampel dengan pembatasan (restricted sample). Karena ada peneliti mengintervensi pemilihan sampel, maka pencuplikan dengan restriksi dikategorikan pencuplikan purposif. Sedang sampel yang diperoleh tanpa restriksi disebut sampel tanpa pembatasan (unrestricted sample).

(d) Pencocokan (matching)

Pencocokan (matching) adalah teknik memilih kelompok pembanding agar sebanding dengan kelompok indeks dalam hal faktor-faktor perancu. Yang dimaksudkan dengan subyek/ kelompok indeks adalah subyek/ kelompok yang dibandingkan dengan kelompok pembanding. Pada studi kasus kontrol, subyek indeks adalah kasus, sedang pada studi kohor, subyek indeks adalah subyek terpapar. Yang dimaksudkan dengan subyek pembanding adalah kontrol pada studi kasus kontrol, dan subyek tak

terpapar pada studi kohor. Pencocokan digunakan pada studi observasional dan eksperimen kuasi. Pada studi kohor dan eksperimen kuasi, tujuan pencocokan untuk mengontrol pengaruh faktor perancu dalam menilai pengaruh paparan terhadap penyakit, atau pengaruh perlakuan terhadap hasil. Pada studi kasus kontrol, tujuan pencocokan untuk meningkatkan efisiensi penaksiran pengaruh paparan terhadap penyakit.

(e) Pencuplikan kuota

Sampling kuota adalah teknik untuk menentukan sampel dari populasi yang mempunyai ciri-ciri tertentu sampai jumlah (kuota) yang di inginkan (Muhyi et al., 2018). Pencuplikan kuota (quota sampling) merupakan teknik pencuplikan non-random dimana peneliti membagi populasi ke dalam kategori (strata), lalu memberikan "jatah" jumlah subyek untuk masing-masing stratum tersebut. Subyek dalam masing-masing kategori tidak dipilih secara random, melainkan berdasarkan kemudahan, dan mungkin sedikit restriksi. Jenis pencuplikan ini jelas mudah dilakukan dan relatif murah. Meskipun mirip dengan pencuplikan random berstrata, tetapi sampel yang dicuplik dengan pencuplikan kuota tidak memiliki karakteristik sampel random, sehingga tidak reliabel untuk digunakan penarikan kesimpulan. Pencuplikan kuota terdiri dari dua jenis – proporsional dan non-proporsional. Pada pencuplikan kuota proporsional, peneliti mencuplik subyek untuk masing-masing kategori karakteristik sampel dalam jumlah proporsional sesuai komposisi karakteristik tersebut pada populasi. Pencuplikan kuota non-proporsional lebih restriktif. Dalam metode ini, peneliti menentukan jumlah minimum unit pencuplikan sesuai yang diinginkan peneliti dalam masing-masing kategori.

(f) Expert sampling

Expert sampling (judgment sampling) merupakan teknik pencuplikan dimana peneliti mewawancarai sekelompok individu yang diketahui merupakan pakar di bidang yang sedang diteliti. Kepakaran tersebut tidak harus berarti pernah mengenyam pendidikan formal, melainkan merujuk kepada suatu pengetahuan khusus.

(g) Pencuplikan bola salju (snowball sampling)

Snowball sampling adalah teknik penentuan sampel yang mula-mula jumlahnya kecil, kemudian membesar (Muhyi et al., 2018). Pencuplikan bola salju (snowball sampling, chain referral sampling, network sampling) dimulai dengan mengidentifikasi seorang atau dua orang subyek yang memenuhi kriteria inklusi untuk suatu penelitian. Subyek tersebut kemudian diminta memberikan keterangan tentang subyek-subyek lainnya yang menurut subyek pertama tadi memenuhi kriteria inklusi. Meskipun sulit untuk dapat memberikan sampel representatif, metode ini bermanfaat untuk mencuplik populasi yang sulit dijangkau.

D. FAKTOR-FAKTOR YANG MENENTUKAN DESAIN PENCUPLIKAN

Sejumlah faktor menentukan desain pencuplikan yang baik: (1) Desain penelitian; (2) Parameter yang diinginkan; (3) Unit pencuplikan; (4) Kerangka pencuplikan; (5) Ukuran sampel; (6) Anggaran penelitian.

E. CIRI-CIRI DESAIN PENCUPLIKAN YANG BAIK

1. Menghasilkan sampel yang representatif dalam studi deskriptif, atau sampel-sampel yang dapat diperbandingkan dengan valid dalam studi analitik.
2. Mampu meminimalkan kesalahan pencuplikan (sampling error).
3. Mampu mengontrol bias sistematis dalam studi analitik.
4. Menghasilkan sampel yang hasil-hasil pengamatan pada sampel dapat diterapkan kepada populasi sasaran dengan tingkat keyakinan yang cukup baik.

DAFTAR PUSTAKA

Muhyi, M., Hartono, Budiyono, S. C., Satianingsih, R., Sumardi, Rifai, I., Zaman, A. Q., Astutik, E. P., & Fitriatien, S. R. (2018). *Metodologi Penelitian*. 1–83. www.unipasby.ac.id

Murti, B. (2015). Populasi, Sampel, Dan Pemilihan Subyek. *Naskah Tutorial Pengembangan Bahan Pengajaran*, 1–26. http://fkm.malahayati.ac.id/wp-content/uploads/2015/12/Populasi-Sampel-Prof-Bhisma-Murti.pdf

Rawung, D. T. (2020). Bahan ajar Diklat Statistisi Ahli BPS Angkatan XXI Tahun 2020 Mata diklat: Metode penarikan sampel. *Pusat Pendidikan Dan Pelatihan Badan Pusat Statistik RI*, 17. https://pusdiklat.bps.go.id/diklat/bahan_diklat/BA_2144.pdf

Salim. (2018). *Metodologi Penelitian*.

Triyono. (2003). Teknik Sampling Dalam Pelaksanaan Penelitian. *Info Kesehatan*, *7*(1), 64. https://osf.io/preprints/inarxiv/dcq8u/download

PROFIL PENULIS

Prasaja, lahir di Karanganyar pada 09 Juli 1973, menyelesaikan pendidikan dasar dan menengah di Karanganyar, S-1 Kedokteran Umum Fakultas Kedokteran UNS tahun 1999, S-2 Magister Kedokteran keluarga UNS tahun 2013, dan Diploma 4 Okupasi Terapi Politeknik Kesehatan Surakarta tahun 2018. Riwayat Pekerjaan mulai tahun 2000 sampai 2003 bekerja sebagai dokter PTT di Puskesmas Tawangmangu Karanganyar Surakarta, tahun 2003 sampai 2004 bekerja sebagai dokter fungsional di RS PKU Muhammadiyah Delanggu, tahun 2004 sampai 2018 bekerja sebagai dokter fungsional di RS Umum Islam Kustati Surakarta.

Sejak tahun 2006 sebagai Dosen di Jurusan Okupasi Terapi Politeknik Kesehatan Kementerian Kesehatan Surakarta. Beberapa mata kuliah yang diampunya dalam bidang metodologi penelitian, statistik kesehatan, Konsep Kesehatan dan Patologi, OT pada Ortopedi serta OT pada Penyakit dalam dan Bedah.

Beberapa penelitian yang disusun telah dipublikasikan di jurnal nasional terakreditasi di bidang kesehatan. Fokus penelitian di bidang okupasi terapi serta aktif menulis Modul Training sesuai topik penelitiannya serta Modul Ajar sesuai mata kuliah yang diampunya.

Email penulis : prasajaahmad@gmail.com

BAB 7
DISTRIBUSI PROPORSI SAMPLING

Nova Fahradina, M. Pd
Universitas Iskandarmuda (Unida)
Banda Aceh

Distribusi sampling adalah distribusi dari rata-rata sampel atau proporsi sampel yang diambil secara berulang-ulang (n kali) dari suatu populasi. Untuk mempelajari populasi kita memerlukan sampel yang diambil dari populasi yang bersangkutan. Populasi dapat dilihat karakteristiknya dari sampel yang diambil. Dalam analisa statistik, pendugaan parameter populasi dilakukan atas dasar statistik sampel (Dajan, 1996).

Agar estimasi atau uji hipotesis mendekati kondisi sebenarnya pada populasi maka perlu diambil sampel-sampel yang dapat mewakili populasi. Hal ini dapat dilakukan dengan cara random sampling, dimana setiap elemen dari populasi memiliki peluang yang sama untuk terpilih menjadi sampel. Berdasarkan sifat-sifat sampel yang diambil dari sebuah populasi, statistika akan membuat kesimpulan umum yang diharapkan berlaku untuk populasi itu. Jika nilai-nilai statistik yang sejenis dikumpulkan, lalu disusun dalam suatu daftar sehingga terdapat hubungan antar nilai statistik dan frekuensi statistik yang didapat, maka diperoleh kumpulan statistik yang disebut distribusi sampling.

Distribusi sampling biasanya diberi nama berdasarkan pada nama statistik yang digunakan. Misalnya distribusi sampling rata-rata, distribusi sampling proporsi, distribusi sampling simpangan baku, dan lain-lain. Nama-nama tersebut biasanya disingkat lagi berturut-turut menjadi distribusi rata-rata, distribusi proporsi, distribusi simpangan baku, dan lain-lain (Sudjana, 2002).

A. DISTRIBUSI SAMPLING RATA-RATA

Distribusi sampel dari rata-rata hitung sampel adalah Suatu distribusi probabilitas yang terdiri dari seluruh kemungkinan rata-rata hitung sampel dari suatu ukuran sampel tertentu yang dipilih dari populasi (suharyadi, 2009).

Misalkan sebuah populasi berukuran terhingga N dengan parameter rata-rata μ dan simpangan baku σ. Dari populasi ini diambil sampel random berukuran n. Jika pengambilan sampel dilakukan tanpa pengembalian maka ada $\binom{N}{n}$ buah sampel yang berlainan. Dari sampel yang didapat kemudian masing-masing dihitung rata-ratanya. Rata-rata yang didapat merupakan data baru diberi simbol $\mu_{\bar{x}}$ dan simpangan bakunya adalah $\sigma_{\bar{x}}$. Distribusi semacam ini disebut distribusi sampel rata-rata (*sample distribution of means*).

Selanjutnya dapat dihitung rata-rata dari distribusi sampel rata-rata (*means of sample distribution*). Rata-rata distribusi ini akan sama dengan rata-rata populasi (Partino, 2010).

Untuk pengambilan sampel dengan pengembalian atau bila N banyaknya tak terhingga atau N besar sekali relative terhadap n (n/N ≤ 5%):

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

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

Untuk pengambilan sampel tanpa pengembalian (n/N > 5%):

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

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

Untuk distribusi normal table z yang digunakan adalah:

$Z = \frac{\bar{x}-\mu}{\sigma_{\bar{x}}}$

Contoh 1

Suatu populasi dengan N = 5, yakni terdiri atas angka-angka 4, 6, 8, 11, 14 dengan σ = 3,56. Dari populasi tersebut akan diambil sampel yang terdiri dari dua angka. Hitung mean dan standar deviasi bila sampel diambil dengan pengembalian dan tanpa pengembalian!

Penyelesaian:

Bila pengambilan sampel dilakukan dengan pengembalian (with replacement) maka berlaku: $N^n = 5^2 = 25$. Yaitu terdapat 25 sampel dengan dua angka yang mungkin diambil dari populasi tersebut.

4, 6, 8, 11, 14

Sampel	**Mean**	**Sampel**	**Mean**	**Sampel**	**Mean**
4, 4	$\bar{X}_1 = 4$	6, 4	$\bar{X}_6 = 5$	8, 4	$\bar{X}_{11} = 6$
4, 6	$\bar{X}_2 = 5$	6, 6	$\bar{X}_7 = 6$	8, 6	$\bar{X}_{12} = 7$
4, 8	$\bar{X}_3 = 6$	6, 8	$\bar{X}_8 = 7$	8, 8	$\bar{X}_{13} = 8$
4, 11	$\bar{X}_4 = 7{,}5$	6, 11	$\bar{X}_9 = 8{,}5$	8, 11	$\bar{X}_{14} = 9{,}5$
4, 14	$\bar{X}_5 = 9$	6, 14	$\bar{X}_{10} = 10$	8, 14	$\bar{X}_{15} = 11$

Sampel	Mean	Sampel	Mean
11, 4	$\bar{X}_{16} = 7{,}5$	14, 4	$\bar{X}_{21} = 9$
11, 6	$\bar{X}_{17} = 8{,}5$	14, 6	$\bar{X}_{22} = 10$
11, 8	$\bar{X}_{18} = 9{,}5$	14, 8	$\bar{X}_{23} = 11$
11, 11	$\bar{X}_{19} = 11$	14, 11	$\bar{X}_{24} = 12{,}5$
11, 14	$\bar{X}_{20} = 12{,}5$	14, 14	$\bar{X}_{25} = 14$

$\mu_x = \frac{4+6+8+11+14}{5} = \frac{43}{5} = 8{,}6$ atau $\mu_{\bar{x}} = \frac{4+5+6+\cdots+14}{25} = 8{,}6$

$\sigma_{\bar{x}} = \sqrt{\frac{(4-8{,}6)^2+(5-8{,}6)^2+\cdots+(14-8{,}6)^2}{25}} = \sqrt{\frac{158}{25}} = \sqrt{6{,}35} = 2{,}52$ atau

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3{,}56}{\sqrt{2}} \quad \frac{3{,}56}{1{,}41} = 2{,}52$

Bila pengambilan sampel dilakukan tanpa pengembalian (without replacement) maka berlaku: $\binom{5}{2} = 10$. Yaitu terdapat 10 sampel dengan dua angka yang mungkin diambil dari populasi tersebut.

4, 6, 8, 11, 14

Sampel	Mean	Sampel	Mean
4, 6	$\bar{X}_1 = 10$	6,11	$\bar{X}_6 = 8{,}5$
4, 8	$\bar{X}_2 = 6$	6, 14	$\bar{X}_7 = 10$
4, 11	$\bar{X}_3 = 7{,}5$	8, 11	$\bar{X}_8 = 9{,}5$
4, 14	$\bar{X}_4 = 9$	8, 14	$\bar{X}_9 = 11$
6,8	$\bar{X}_5 = 7$	11, 14	$\bar{X}_{10} = 12{,}5$

$\mu = \mu_x = \frac{10+6+\cdots+12{,}5}{10} = \frac{91}{10} = 9{,}1$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{3{,}56}{\sqrt{2}}\sqrt{\frac{5-2}{5-1}} = 2{,}18$

Dari contoh terlihat bahwa rata-rata untuk sampel membentuk sebuah distribusi peluang dan berlaku sebuah dalil yang dinamakan Dalil Limit Pusat:

Jika sebuah populasi mempunyai rata-rata μ dan simpangan baku σ yang besarnya terhingga, maka untuk ukuran sampel acak n cukup besar, distribusi

rata-rata sampel mendekati distribusi normal dengan rata-rata $\mu = \mu_x$ dan simpangan baku $\sigma\bar{x} = \frac{\sigma}{\sqrt{n}}$.

Dalil ini berlaku untuk sebarang bentuk atau model populasi yang simpangan bakunya terhingga besarnya. Untuk populasi yang simpangan bakunya terhingga maka rata-rata sampelnya akan mendekati distribusi normal. Pendekatan normal ini akan lebih baik untuk ukuran sampel n yang makin besar. Biasanya batas untuk dikatakan sebuah sampel besar adan $n \geq 30$. Apabila populasi yang disampel sudah berdistribusi normal, maka rata-rata sampel jugan berdistribusi normal meskipun ukuran sampel $n < 30$.

B. DISTRIBUSI SAMPLING PROPORSI

Distribusi sampling proporsi adalah distribusi dari proporsi (persentase) yang diperoleh dari semua sampel sama besar yang mungkin dari suatu populasi (Hasan: 2008). Distribusi sampling proporsi dapat digunakan untuk mengetahui persentase atau perbandingan antara dua hal yang berkomplemen (peristiwa binomial). Pada distribusi sampling proporsi berlaku rumus-rumus berikut:

1. Jika ukuran populasi besar ($n/N \leq 5\%$) atau pengambilan sampel dengan pengembalian, maka berlaku:

   $\mu_p = p$ $\qquad\qquad$ $\sigma_p = \sqrt{\frac{P\,(1-n)}{n}}$

2. Jika ukuran populasi kecil ($n/N > 5\%$) atau pengambilan tanpa pengembalian, maka berlaku:

   $\mu_p = p$ $\qquad\qquad$ $\sigma_p = \sqrt{\frac{P\,(1-n)}{n}}\sqrt{\frac{N-n}{N-1}}$

3. Untuk perhitungan daftar distribusi normal baku pada distribusi sampling proporsi, dapat digunakan rumus berikut:

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

Contoh 2

Ada petunjuk kuat bahwa 20% anggota masyarakat di suatu wilayah merupakan golongan perokok. Sebuah sampel acak terdiri atas 100 orang telah

diambil. Tentukan peluang bahwa dari 100 orang itu akan ada paling sedikit 10 orang dari golongan perokok!

Penyelesaian:
Populasi berukuran cukup besar dengan p = 0,2 dan 1-p = 0,8
Untuk ukuran sampel 100, diantaranya paling sedikit 10 orang tergolong golongan perokok, maka paling sedikit x/n = 0,10.
Kekeliruan bakunya adalah:

$\sigma_p = \sqrt{\frac{P\ (1-n)}{n}} = \sqrt{\frac{0{,}2\ (0{,}8)}{100}} = 0{,}04$

bilangan z paling sedikit = $\frac{0{,}10-0{,}2}{0{,}04} = -2{,}5$

Dari daftar normal baku, luasnya 0,5 – 0,4938 = 0,0062.
Peluang dalam sampel itu akan ada paling sedikit 10 orang perokok adalah 0,0062.

C. DISTRIBUSI SELISIH DAN JUMLAH RATA-RATA

Misalkan, kita mempunyai dua populasi yang sedang kita teliti, yaitu populasi 1 dan populasi 2 yang masing-masing memiliki ukuran N1 dan N2. Populasi kesatu mempunyai rata-rata m_{s1} dan simpangan baku s_{s1}. Kemudian dari masing-masing populasi tersebut kita mengambil sampel dengan ukuran n1 dan n2. Pengambilan sampel n1 dari populasi N1 menghasilkan kombinasi sampel sebanyak C(N1, n1) dan pengambilan sampel n2 dari populasi N2 menghasilkan kombinasi sampel sebanyak C(N2, n2).

Distribusi selisih rata-rata:

$m_{s1-s2} = m_{s1} - m_{s2}$

$s_{s1-s2} = \sqrt{\frac{s_{s1}^2}{n1} + \frac{s_{s2}^2}{n2}}$

Distribusi selisih rata-rata:

$m_{s1+s2} = m_{s1} + m_{s2}$

$s_{s1+s2} = \sqrt{\frac{s_{s1}^2}{n1} + \frac{s_{s2}^2}{n2}}$

variable normal standar z:

$$z = \frac{(s_1 - s_2) - m_{s1-s2}}{s_{s1-s2}}$$

Contoh 3

Rata-rata tinggi badan mahasiswa perempuan adalah 150 cm dan simpangan bakunya 4,7 cm. sementara itu untuk rata-rata tinggi badan mahasiswa laki-laki adalah 162 dan simpangan bakunya 5 cm. dari kedua kelopok mahasiswa itu masing-masing diambil sebuah sampel acak secara independen sebanyak 130 orang. Berapa peluang rata-rata tinggi badan mahasiswa laki-laki 11 cm lebihnya dari rata-rata tinggi badan mahasiswa perempuan?

Penyelesaian:

Misalkan m_{s1} merupakan rata-rata tinggi badan dari sampel mahasiswa laki-laki dan m_{s2} merupakan rata-rata tinggi badan dari sampel mahasiswa perempuan. Simpangan baku mahasiswa laki-laki adalah s_{s1} dan Simpangan baku mahasiswa perempuan $s_{s2,}$ dan n1 = n2 = 140.

$$m_{s1-s2} = m_{s1} - m_{s2}$$
$$= 162 - 150$$
$$= 12 \text{ cm}$$

$$s_{s1-s2} = \sqrt{\frac{s_{s1}^2}{n1} + \frac{s_{s2}^2}{n2}}$$
$$= \sqrt{\frac{(5)^2}{140} + \frac{(4,7)^2}{140}}$$
$$= 0,58 \text{ cm}$$

Variable normal z:

$$z = \frac{(s_1 - s_2) - m_{s1-s2}}{s_{s1-s2}}$$
$$= \frac{11-12}{0,58}$$
$$= -1,72$$

Luas daerah normal baku adalah 0,5 + 0,4573 = 0,9573.

Jadi peluang rata-rata tinggi badan mahasiswa laki-laki 11 cm lebihnya dari rata-rata tinggi badan mahasiswa perempuan adalah 0,9573.

D. DISTRIBUSI SAMPLING SELISIH PROPORSI

Pada prinsipnya, proses terbentuknya distribusi sampling selisih proporsi ini sama dengan pembentukan distribusi sampling selisih dan jumlah rata-rata. Misalnya terdapat 2 populasi binomial (populasi yang dibedakan menjadi 2 kelompok, seperti merokok dan tidak merokok, setuju dan tidak setuju, dsb). Dari kedua populasi tersebut, secara independen masing-masing diambil sampel acak berukuran n_1 dan n_2. Jika p adalah selisih p_1 dan p_2, maka akan didapat $p_1 - p_2$ yang membentuk distribusi normal dengan rata-rata μ_{p1-p2} dan deviasi standar σ_{p1-p2}.

Untuk rata-rata:

$\mu_{p1-p2} = p_1 - p_2$

Untuk simpangan baku:

$$\sigma_{p1-p2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

Jika n_1 dan n_2 ($n_1, n_2 \geq 30$) cukup besar, distribusi sampling proporsi akan mendekati distribusi normal z:

$$Z = \frac{(p_1 - p_2) - (p_1 - p_2)}{\sigma_{p1-p2}}$$

Contoh 4

Ada petunjuk kuat bahwa calon A akan mendapatkan suara sebanyak 65% dalam pemilihan. Sebanyak masing-masing 300 orang telah diambil dari dua buah sampel acak. Tentukan peluangnya akan terjadi perbedaan persentase tidak lebih dari 10% yang akan memilih A!

Penyelesaian:

Kedua sampel diambil dari sebuah populasi, jadi kita anggap dua populasi yang sama, sehingga $p_1 = p_2 = 0{,}65$.

$\mu_{p1-p2} = p_1 - p_2$

$= 0{,}65 - 0{,}65 = 0$

$\sigma_{p1-p2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$

$= \sqrt{\frac{0{,}65(0{,}35)}{300} + \frac{0{,}65(0{,}35)}{300}}$

$= 0{,}04$

Peluangnya adalah $-10\% < x/n_1 - y/n_2 < 10\%$

$Z_1 = \frac{(p_1 - p_2) - (p_1 - p_2)}{\sigma_{p1-p2}}$ $\quad$ $Z_2 = \frac{(p_1 - p_2) - (p_1 - p_2)}{\sigma_{p1-p2}}$

$= \frac{-0{,}10 - 0}{0{,}04}$ $\quad$ $= \frac{0{,}10 - 0}{0{,}04}$

$= -2{,}5$ $\quad$ $= 2{,}5$

Luas daerah z adalah 2(0,4938) = 0,9876

E. DISTRIBUSI SAMPLING SIMPANGAN BAKU

Misalkan kita mempunyai populasi berukuran N. Dari populasi N diambil sampel acak berukuran n, kemudian untuk setiap sampel yang diambil dihitung simpangan bakunya yaitu s. jika populasi berdistribusi normal, maka distribusi sampling simpangan baku untuk n besar, $n \geq 100$, berlaku:

$\mu_s = \sigma$

$\sigma_s = \frac{\sigma}{\sqrt{2n}}$

untuk daerah z adalah:

$z = \frac{s - \sigma}{\sigma_s}$

contoh 5

varians sebuah populasi yang berdistribusi normal 6,25. Diambil sampel berukuran 215. Tentukan peluang sampel tersebut akan mempunyai simpangan baku kurang dari 3,5.

Penyelesaian:

Varians = 6,25 maka:

$\sigma = \sqrt{6{,}25} = 2{,}5$

ukuran sampel cukup besar, maka $\mu_s = 2{,}5$

simpangan baku:

$$\sigma_s = \frac{\sigma}{\sqrt{2n}}$$

$$= \frac{2{,}5}{\sqrt{2x215}}$$

$$= 0{,}12$$

Bilangan z untuk s = 3,5 adalah:

$$z = \frac{s - \sigma}{\sigma_s}$$

$$= \frac{3{,}5 - 2{,}5}{0{,}12}$$

$$= 8{,}33$$

Tidak terjadi sampel berukuran 215 dengan simpangan baku kurang dari 3,5.

DAFTAR PUSTAKA

Dajan, A. (1996). *Pengantar Metode Statistik*. Jakarta: LP3ES.

Harinaldi, (2005). *Prinsip-Prinsip Statistik untuk Teknik dan Sains*. Jakarta: Erlangga.

Hasan, M. (2008). *Pokok-Pokok Materi Statistik 2*. Jakarta: PT. Bumi Aksara.

Partino dan Idrus. (2010). *Statistika Inferensial*. Yogyakarta: Safiria Insania Press.

Sudjana. (2002). *Metode statistika*. Bandung: Tarsito.

Suryadi dan Purwanto. (2015). *Statistika*. Jakarta: Salemba Empat.

PROFIL PENULIS

Nova Fahradina, M. Pd Lahir di Medan tanggal 02 November 1986. Lulus S1 Pendidikan Matematika Universitas Syiah Kuala (Unsyiah) Banda Aceh tahun 2009, dan Program Pascasarjana Magister Pendidikan Matematika Unsyiah tahun 2014. Penulis saat ini merupakan dosen tetap di FKIP Pendidikan Matematika Universitas Iskandarmuda (Unida) Banda Aceh. Penulis merupakan pengelola e-jurnal DikMas: Jurnal Pendidikan Matematika dan Sains dengan alamat https://ejournal.unida-aceh.ac.id/index.php/dikmas/index. Artikel yang pernah ditulis oleh penulis dapat dilihat melalui https://scholar.google.com/citations?hl=id&user=c4_XeEMAAAAJ.
Email: nova.farahdina@unida-aceh.ac.id

BAB 8
KESALAHAN SAMPLING DAN NON SAMPLING

Dr. La One ST, MT.
Univerrsitas Haluoleo

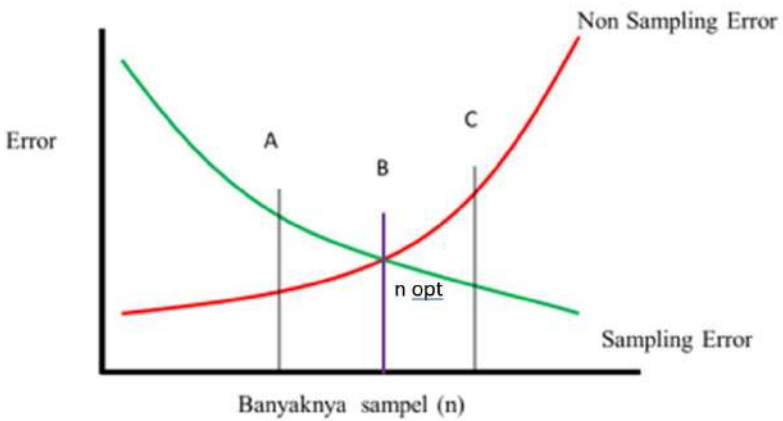
A. PENDAHULUAN

Asumsi umum dalam teori sampling bahwa nilai sebenarnya dari setiap unit dalam populasi dapat diperoleh dan ditabulasi tanpa kesalahan. Dalam praktiknya, asumsi tersebut sulit diterapkan karena sangat tidak mungkin menghindari adanya kendala praktis dalam pelaksanaan survey atau sensus. Pada dasarnya angka statistik merupakan nilai pendugaan yang diupayakan sedapat mungkin tidak menyimpang jauh berbeda dengan nilai yang diharapkan melalui penerapan metode ilmiah.

Angka statistik hasil survey atau sensus yang dirilis oleh lembaga-lembaga statistik dan peneliti di seluruh dunia pasti tidak akan mendapatkan hasil yang sempurna atau bebas dari kesalahan. Kesalahan tersebut itu bukan hanya bersumber dari pihak yang melaksanakan survey atau sensus (lembaga statistik dan Peneliti) tapi juga responden (masyarakat) memiliki andil dalam hal tersebut. Kesalahan survei mungkin disebabkan oleh pewawancara, responden, pengolah data, dan personel survei lainnya. Umumnya, penyebab kesalahan pengukuran adalah pertanyaan atau desain kuesioner yang buruk, pelatihan atau pengawasan pribadi yang tidak memadai, dan kontrol kualitas yang tidak memadai.

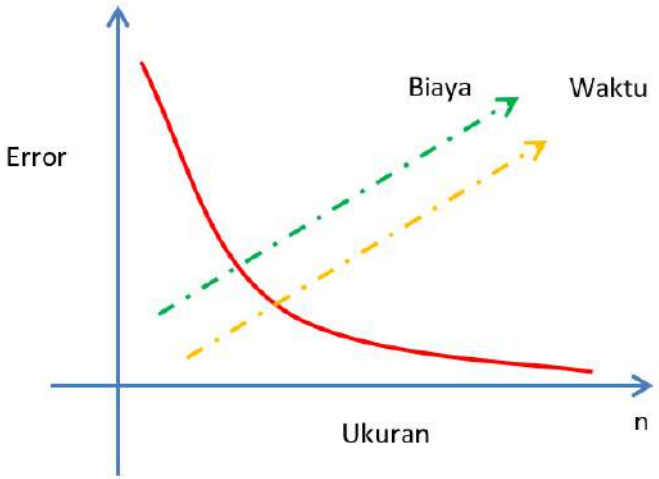
Kesalahan survey dapat diklasifikasikan menjadi kesalahan sampling dan kesalahan non-sampling. Kesalahan sampling berkaitan dengan pengambilan sample acak yang kurang menggambarkan keseluruhan populasi secara sempurna. Sementara kesalahan non-sampling akibat *human error* tidak dapat dihindari baik dalam sensus maupun survei. Data yang terkumnpul secara lengkap dalam pencacahan sensus akan bebas dari kesalahan sampling tetapi tidak akan bebas dari kesalahan non-sampling. Data yang dikumpulkan melalui sampel survei dapat memiliki kesalahan sampling dan kesalahan non-sampling

Secara umum, kesalahan sampling dapat direduksi dengan bertambahnya ukuran sampel, sedangkan kesalahan non-sampling meningkat seiring dengan bertambahnya ukuran sampel. Kesalahan non-sampling memiliki kontribusi utama terhadap total kesalahan survei, sementara kesalahan sampling lebuh mudah untuk diminimalisir. Untuk mereduksu kesalahan total, perlu upaya untuk meningkatkan kualitas ketersediaan data survey. Hubungan antara kesalahan sampling dan kesalahan non-sampling seperti Gambar 8.1.



Gambar 8.1 Hubungan Kesalahan Sampling dan Kesalahan Non-Sampling

Keandalan survei adalah fungsi dari kesalahan total, yang terdiri dari kesalahan sampling dan kesalahan non-sampling. Reduksi kesalahan survey total dengan memastikan kualitas data survey akan terkait dengan peningkatan biaya dan waktu. Hubungan antara ukuran sampel, kesalahan sampling, biaya yang dibutuhkan dan waktu pelaksanaan survey ditunjukan pada Gambar 8.2.



Gambar 8.2. Hubungan antara ukuran sampel, kesalahan, biaya dan waktu survey.

B. KESALAHAN SAMPLING

Kesalahan sampling disebabkan oleh pengamatan sampel yang tidak mempertimbangkan seluruh populasi. Teknik sampling melihat sebagian karakteristik populasi dan mengacu pada perbedaan antara perkiraan yang berasal dari survei sampel dan nilai sebenarnya yang akan dihasilkan jika sensus dari seluruh populasi dilakukan dan diambil dalam kondisi yang sama. Kesalahan sampling timbul akibat sampel yang dipilih bukan merupakan representasi sempurna dari populasi uji. Kesalahan sampling dapat direduksi dengan pemilihan sampel yang cermat dan meningkatkan ukuran sampel. Pada pelaksanaan sensus, tidak ada kesalahan pengambilan sampel karena seluruh populasi diperhitungkan.

Tujuan dari survei sampel adalah untuk membuat perkiraan atau kesimpulan suatu populasi berdasarkan pengamatan yang dilakukan pada sejumlah sampel yang terbatas. Perbedaan antara nilai aktual (parameter) keseluruhan populasi (biasanya tidak diketahui) dan nilai (statistik) yang diperkirakan dari sampel yang diamati disebut kesalahan sampling. Kesalahan sampling ditentukan dengan rumus:

$$\text{Kesalahan Sampling} = \bar{X} - \bar{x}$$

$\bar{X}$ = Rata-rata populasi ; $\bar{X} = \frac{\sum x_i}{N}$

$\bar{x}$= Rata-rata sampel; $\bar{x} = \frac{\sum x_i}{n}$

Contoh 1.

Indeks Polutasi Udara pada 132 Kota di Negara-negara Asia Tahun 2022 sebagai berikut:

95.37; 94.38; 93.79; 93.63; 93.5; 93.08; 92.43; 92.35[; 92; 91.57; 91.51; 91.34; 90.52; 89.97; 89.86; 89.58; 89.42; 89.2[89.15; 88.45; 88.36; 87.36; 86.11; 86.04; 84.42; 84.33; 83.79; 83.54; 83.34; 83.19; 83.01; 82.89; 82.73; 82.59; 81.43; 81.29; 81.14; 80.51; 80.5; 78.65; 78.62; 78.45; 78.33; 78.11; 77.28; 76.97; 76.61; 76.5; 76.41; 76.09; 75.68; 75.61; 75.12; 74.79; 74.17; 73.66; 73.38; 73.36; 73.29; 73.26; 73.19; 73.1; 72.97; 71.99; 71.76; 70.85; 70; 69.74; 69.17; 68.18; 67.58; 67.57; 67.52; 67.24;66.59; 66.01; 65.81; 65.77; 65.67; 65.6; 65.42; 64.68; 64.58; 64.39; 64.01; 63.1; 62.7; 62.39; 62.31; 61.97; 61.51; 61.38; 61.24; 60.46; 60.41; 60.26; 60.12; 59.66; 59.29; 58.75; 58.35; 57.66; 57.44; 57.21; 56.98; 56.82; 56.72; 56.54; 55.84; 5.29; 54.83; 53.43;

52.47; 52.45; 51.41; 51.4; 51.36; 50.36; 50.11; 48.98; 48.54; 48.47; 48.04; 47.01; 46.25; 44.54; 43.65; 42.31; 41.71; 40.39; 36.31; 33.31 (https://www.numbeo.com/pollution/region_rankings.jsp?title=2022&region=142)

Tentukan kesalahan Sampling, dari:

a. 30 data sampel dengan pemilihan secara acak sebagai berikut: 92.43; 92.00 ; 90.52 ; 89.86 ; 88.45 ; 87.36 ; 83.34 ; 82.89 ; 81.43; 78.45; 76.97; 76.61; 76.09; 74.79 ; 73.66; 73.36 ; 72.97 ; 67.57; 65.81; 64.58 ; 62.39 ; 61.97 ; 59.29 ; 58.35 ; 56.72; 52.47 48.54; 47.01 ; 44.54 ; 42.31
b. 30 data yang dipilih dari data indeks polusi tertinggi
c. 30 data yang dipilih dari data indeks polusi terendah
d. Range kesalahan sampling dari 30 data sampel.

Jawab:

Rata-rata indeks polusi udara dari populasi 132 Kota,

$\bar{X} = \frac{\sum x_i}{N} = \frac{9.233,13}{132} = 69,95$

Rata-rata indeks polusi udara dari 30 sampel acak,

$\bar{x} = \frac{\sum x_i}{n} = \frac{2.122,73}{30} = 70,76$

Rata-rata indeks polusi udara dari 30 sampel tertinggi,

$\bar{x} = \frac{\sum x_i}{30} = \frac{2.681,58}{30} = 89,39$

Rata-rata indeks polusi udara dari 30 sampel terendah,

$\bar{x} = \frac{\sum x_i}{30} = \frac{1.490,17}{30} = 49,67$

a. Kesalahan Sampling dari 30 sampel acak, $\bar{X} - \bar{x}$ = 69,95-70,76 = - 0,81
b. Kesalahan Sampling dari 30 sampel dengan indeks polutan tertinggi = 69,95-89,39 = -19,44
c. Kesalahan Sampling dari 30 sampel dengan indeks polutan terendah = 69,95-49,67 = 20,28
d. Range kesalahan sampling dengan ukuran sampel 30 adalah (-19,44; 20,28)

Pemilihan sampel dengan indeks polusi tertinggi maupun indeks polusi terendah tidak mempresentasekan indeks polusi udarapada 132 Kota di Asia, namun pemilihan 30 sampel acak lebih mempresentasekan indeks polusi udara pada 132 Kota di Negara-negara Asia. Hal ini ditunjukan dengan kesalahan sampling melalui pemilihan sampel dengan indeks polusi tertinggi maupun indeks polusi terendah memiliki kesalahan sampel jauh lebih besar dibanding dengan pemilihan sampel secara acak. Pemilihan sampel pada keseluruhan populasi akan menghasilkan kesalahan sampel menjadi nol.

1. Estimasi Kesalahan Sampling

Setiap perkiraan yang berasal dari sampel tunduk pada kesalahan pengambilan sampel karena hanya sebagian dari populasi yang diamati. Sampel yang berbeda dapat menghasilkan perkiraan yang berbeda. Kesalahan pengambilan sampel menyebabkan variabilitas di antara perkiraan yang berasal dari sampel yang berbeda ketika menjaga ukuran dan desain sampel yang sama, dan metode estimasi yang sama digunakan.

Kesalahan Sampling dari suatu estimasi yang dihasilkan dari sampel acak sederhana (SRS) dapat dihitung dengan mudah. Akan tetapi Susenas 2007 dan survei-survei BPS lainnya menggunakan rancangan multi-stage cluster sampling sebagai pengganti SRS. Multi-stage cluster sampling merupakan pilihan terbaik dalam berbagai situasi lapangan dimana frame yang memadai untuk pemilihan unit terkecil (ultimate-stage units, SSU) tidak tersedia, dan biaya untuk membentuk frame semacam itu sangat mahal. Multi-stage cluster sampling juga mampu menekan biaya dibanding metode SRS, karena adanya penghematan waktu perjalanan dan biaya operasional di lapangan. Metode estimasi standard error yang digunakan harus mempertimbangkan rancangan multi-stage cluster sampling ini.

Rancangan multi-stage cluster sampling yang kompleks ini memerlukan pendekatan khusus dalam estimasi standard error-nya, karena penghitungan estimasi standard error secara langsung berdasarkan rancangan ini sangat sulit sehingga tidak mungkin dilakukan. Untuk subkelompok/domain populasi dengan ukuran sampel yang relatif kecil, pendekatan untuk mengestimasi standard error ini juga menjadi tidak

tepat. Susenas 2007 dirancang sedemikian rupa sehingga statistik kesejahteraan rakyat dan statistik perumahan dan permukiman dapat diestimasi dengan tepat dan estimasi pendekatan sampling error-nya juga dapat dihitung dari survei itu sendiri. Hal ini berimplikasi pada ukuran sampel yang cukup besar agar estimasi yang dihasilkan memenuhi persyaratan teori statistik untuk sampel berukuran besar yang dikembangkan dalam survei sampel.

Estimasi kesalahan sampling biasanya diukur dengan varians sampling, yang tergantung pada banyak hal, termasuk metode sampling, metode estimasi, ukuran sampel dan variabilitas karakteristik yang diestimasi. Disamping menggunakan varian sampling untuk mengukur kesalahan sampling, metode yang biasa digunakan antara lain adlah interval kepercayaan, standard error, koefisien variasi, dan margin error.

a. Varians Sampling

Dalam desain sampel sederhana, seperti Simple Random Sampling, varians sampling dapat dihitung secara langsung menggunakan rumus. Namun, untuk desain sampling yang lebih kompleks biasanya, estimasi varians sampling biasanya dihitung dengan metode linearisasi Taylor atau metode resampling seperti jackknife dan bootstrap.

Terlepas dari metode mana yang digunakan untuk estimasi varians, metode ini harus menggabungkan sifat desain sampel seperti stratifikasi, pengelompokan, dan pemilihan multitahap atau multifase. Faktor-faktor lain yang mempengaruhi besarnya varians sampling meliputi:

1) Secara umum, varians sampling menurun dengan bertambahnya ukuran sampel tetapi perubahannya tidak proporsional.
2) Ukuran populasi berdampak pada varians sampel untuk populasi berukuran kecil hingga sedang,, sedangkan untuk populasi besar memiliki pengeruh yang kecil.
3) Variabilitas karakteristik yang diambil dalam populasi juga mempengaruhi besar kecilnya sampling error. Semakin besar perbedaan antara unit populasi, semakin besar ukuran sampel yang diperlukan untuk mencapai tingkat presisi tertentu.
4) Rencana pengambilan sampel, yang mencakup desain sampel dan prosedur estimasi, juga mempengaruhi besarnya kesalahan pengambilan sampel. Metode pengambilan sampel (desain sampel)

memengaruhi ukuran kesalahan pengambilan sampel. Survei yang melibatkan desain sampel yang kompleks dapat menyebabkan kesalahan pengambilan sampel yang lebih besar daripada yang lebih sederhana. Prosedur estimasi juga memiliki dampak besar pada kesalahan sampling. Konsep-konsep perlu dipastikan dalam pengambilan sampel.

Penghitungan estimasi varians suatu karakteristik pada prinsipnya dapat dilakukan dengan metode replikasi atau dengan metode linearisasi Taylor. Langkah-langkah metode replikasi:

1) Membagi sampel menjadi replikasi-replikasi subsampel yang mencerminkan rancangan dari sampel tersebut dengan menentukan variabel strata dan PSU.
2) Menghitung penimbang (weight) untuk masing-masing replikasi, dengan menggunakan prosedur yang sama yang digunakan untuk penimbang sampel.
3) Menghitung estimasi-estimasi untuk masing-masing replikasi dengan menggunakan metode yang sama yang digunakan untuk estimasi sampel yang lengkap.
4) Mengestimasi varians dari estimasi sampel yang lengkap, dengan menggunakan hasil-hasil estimasi dari sampel yang lengkap dan replikasi.

Varians replikasi dihitung dengan Rumus:

$$v\left(\hat{\bar{\theta}}\right) = \sum_{g=1}^{G} c(\hat{\bar{\theta}}_{(g)} - \hat{\bar{\theta}})^2$$

dengan:

$\hat{\bar{\theta}}$ = estimasi keseluruhan sampel

$\hat{\bar{\theta}}_{(g)}$ = estimasi dari replikasi ke-g

G = jumlah replikasi

c = konstanta yang tergantung pada metode replikasi yang digunakan

Misalkan w_i adalah penimbang observasi ke-i, $w_{(g)i}$ adalah penimbang untuk replikasi ke-g, dan $\hat{\bar{\theta}}$ adalah total, maka

$\hat{\bar{\theta}} = \sum_{i=1}^{n} w_i y_i$ dan $\hat{\bar{\theta}}_{(g)} = \sum_{i=1}^{n_g} w_{(g)i} y_i$

Dengan n_g= jumlah observasi untuk replikasi ke-g.

Formula yang lebih spesifik yang mengandung faktor JKn, $h_{(g)}$, dan faktor koreksi populasi, $f_{(g)}$, yaitu:

$$v\left(\hat{\bar{\theta}}\right) = c \sum_{g=1}^{G} f_{(g)} h_{(g)} (\hat{\bar{\theta}}_{(g)} - \hat{\bar{\theta}})^2$$

dengan:

c = konstanta yang tergantung pada metode replikasi yang digunakan

$f_{(g)}$ = finite population correction (fpc) yang ditentukan untuk masing- masing replikasi,

$$f_{(g)} = \frac{1 - n'_h}{N_h}$$

$h_{(g)} = \frac{n'_h - 1}{n_h}$

h' = strata yang bersesuaian dengan replikasi ke-g

n'_h= jumlah PSU yang unique dalam strata ke h'

Ada beberapa metode replikasi, diantaranya BRR, Fay, dan Jacknife. Masing-masing metode replikasi yang digunakan akan memiliki nilai konstanta c dan penghitungan varians replikasi yang berbeda. Nilai-nilai konstanta ini diringkas dalam tabel 1 berikut ini:

Tabel 8. 1. Nilai c dari berbagai metode replikasi

Metode	Nama	Nilai c
Balanced Repeated Replication	BRR	1/G
Fay's Method	FAY	$1/[G(1-K)^2]$
Jacknife 1	JK1	(G-1)/G
Jacknife 2	JK2	1
Jacknife n	JKn	1

Selain metode diatas, metode linierisasi Taylor (Taylor Series Linearization) juga merupakan salah satu metode untuk mengestimasi karakteristik baik total maupun rata-rata. Metode linierisasi Taylor memperlakukan persentase atau rata-rata sebagai suatu estimasi rasio, r = y/x, dengan y sebagai total nilai sampel untuk variabel y, dan x adalah jumlah kasus dalam grup atau subgrup yang diperhitungkan. Ragam dari r dihitung

menggunakan rumus di bawah ini, dengan galat baku adalah akar pangkat dua dari ragam tersebut.

$$\text{Var (r)} = \frac{1-f}{x^2}\sum_{h=1}^{H}\left[\frac{m_h}{m_h-1}\left(\sum_{i=1}^{m_h} z_{hi}^2 - \frac{Z_h^2}{m_h}\right)\right]$$

dengan

$$Z_{hi} = y_{hi} - r\,x_{hi} \text{ dan } Z_h = y_h - r\,x_h$$

h : adalah strata yang mempunyai nilai antara 1 dan H,

m_h: adalah jumlah blok sensus terpilih dalam strata h,

y_{hi}: adalah jumlah tertimbang nilai dari variabel y dalam blok sensus i strata h,

x_{hi}: adalah jumlah kasus dalam blok sensus i dan strata h, dan

f : adalah fraksi sampling, yang karena nilainya kecil, tidak diperhitungkan.

Untuk sampel kecil, penghitungan dengan menggunakan metode replikasi akan menghasilkan estimasi yang jauh lebih tepat dibandingkan dengan metode linierisasi. Namun metode replikasi akan membutuhkan waktu yang jauh lebih lama dibandingkan dengan metode linierisasi jika jumlah sampelnya besar. Penghitungan kesalahan sampling dapat menggunakan program Wesvar 4.2 dan Stata version 9. Wesvar 4.2 hanya dapat menghitung kesalahan sampling dengan metode replikasi, sedangkan Stata version 9 dapat digunakan baik untuk metode replikasi dan metode linierisasi.

b. Interval kepercayaan atau Confidence Interval (CI)

Interval kepercayaan (CI) memberikan rentang nilai di sekitar perkiraan yang kemungkinan mencakup nilai populasi yang tidak diketahui dengan probabilitas tertentu. Probabilitas ini adalah tingkat kepercayaan CI. Untuk perkiraan tertentu dalam sampel tertentu, menggunakan tingkat kepercayaan yang lebih tinggi menghasilkan CI yang lebih luas, yang berarti CI yang kurang tepat. Tingkat kepercayaan yang paling umum digunakan adalah 95%, tetapi tingkat kepercayaan 99% atau 90% juga digunakan dalam keadaan tertentu.

Interval kepercayaan bagi nilai populasi yang sebenarnya dengan besaran peluang tertentu diperoleh dari nilai estimasi beserta standard error-nya. Apabila proses pengambilan sampel diulang berkali-kali dan

nilai estimasi serta standard error karakteristik dihitung untuk setiap sampel, maka kira-kira 95% selang kepercayaan dengan 1,96 standard error di bawah dan di atas nilai estimasi akan mencakup nilai populasi sebenarnya. Dengan kondisi biasa, pendekatan (1-α) 100% selang kepercayaan bagi parameter θ adalah

$$\hat{\theta} - z_{\frac{\alpha}{2}} se(\hat{\theta}) < \theta < \hat{\theta} + z_{\frac{\alpha}{2}} se(\hat{\theta})$$

c. Standard error

Standard errror (se) merupakan ukuran statistik yang menyatakan keragaman antar estimasi parameter populasi yang diturunkan dari seluruh kemungkinan sampel yang berbeda dan disurvei dengan kondisi yang sama. Ukuran ini lebih mudah diinterpretasikan karena memberikan indikasi kesalahan pengambilan sampel menggunakan skala yang sama dengan perkiraan sedangkan varians didasarkan pada perbedaan kuadrat. Nilai standard error ini dapat didekati dari sembarang sampel tunggal, yang menyatakan ukuran presisi sejauh mana estimasi yang dihasilkan akan mendekati rata-rata estimasi dari seluruh kemungkinan sampel. Standard error merupakan akar kuadrat dari varian sampling, yang ditulis dengan rumus.

$$se(\hat{\theta}) = \sqrt{v(\hat{\theta})}$$

d. Koefisien variasi (relative standar error)

Koefisien variasi didefinisikan sebagai rasio antara nilai kesalahan standard terhadap nilai perkiraaan karateristik yang diukur. Koefisien variasi hanya mengukur varians sampling dan tidak mengukur bias nilai estimasi.

$$rse(\hat{\theta}) = \frac{se(\hat{\theta})}{\hat{\theta}}$$

Koefisien variasi sangat berguna dalam membandingkan ketepatan estimasi sampel, di mana ukuran atau skalanya berbeda satu sama lain. Meskipun rse banyak digunakan namun tidak disarankan untuk mengukur presisi proporsi, terutama ketika proporsi yang diperkirakan mendekati 0 atau 1. Dalam hal ini, interval kepercayaan lebih tepat digunakan.

e. Margin error

Margin error adalah setengah dari lebar CI. Semakin besar margin error, semakin sedikit kepercayaan yang harus dimiliki seseorang bahwa suatu hasil akan mencerminkan hasil survei terhadap seluruh populasi. Ini sering digunakan untuk melaporkan kesalahan pengambilan sampel oleh lembaga survei atau jurnalis.

C. KESALAHAN NON SAMPLING

1. Sumber/ penyebab kesalahan non-sampling

Kesalahan non-sampling merupakan masalah utama yang muncul dalam suatu survei dan secara langsung mempengaruhi hasil survei. Dilema yang sering dihadapi peneliti survei adalah apakah akan memilih sampel besar untuk meminimalkan kesalahan, atau sampel yang lebih sedikit dengan memastikan kualitas pewawancara yang lebih baik, tingkat respons yang lebih tinggi, dan respons yang lebih akurat. Idealnya, seorang peneliti memusatkan upaya untuk mengurangi kesalahan ssurvey. Mengingat kendala biaya dan waktu, yang ideal jarang terwujud. Dilema ini semakin diperumit oleh fakta bahwa setelah melakukan survey, peneliti jarang dapat mengukur kesalahan survei total atau membandingkan ukuran relatif komponennya.

Sumber kesalahan non-sampling adalah kesalahan yang tidak terkait dengan pengambilan sampel. Kesalahan non sampling dapat terjadi pada setiap tahap perencanaan dan pelaksanaan survei atau sensus. Hal ini terjadi pada tahap perencanaan, tahap pelaksanaan di lapangan serta pada tahap tabulasi dan komputasi. Sumber-sumber kesalahan non-sampling:

a. Kegagalan mengukur beberapa unit dalam sampel terpilih karena kerangka survey tidak lengkap dan cakupan populasi atau sampel yang tidak lengkap,
b. Kesalahan pengamatan karena teknik pengukuran yang tidak sempurna, kurangnya spesifikasi yang tepat dari domain studi dan ruang lingkup penyelidikan, serta definisi yang salah, dan
c. Kesalahan mengedit, memberi kode, mengentri data, dan tabulasi hasil-hasil survei.

Lebih khusus, alasan berikut dapat menimbulkan kesalahan non-sampling atau menunjukkan kehadirannya:

a. spesifikasi data mungkin tidak memadai dan tidak konsisten dengan tujuan survei atau sensus.
b. definisi satuan yang tidak tepat, identifikasi satuan yang tidak lengkap atau salah,
c. metode pencacahan yang salah.
d. metode pengumpulan wawancara dan observasi mungkin tidak akurat atau tidak tepat.
e. kuesioner, definisi dan instruksi mungkin ambigu.
f. para peneliti mungkin tidak berpengalaman atau tidak terlatih dengan baik.
g. kesalahan penarikan dapat menimbulkan kesulitan dalam melaporkan data yang sebenarnya.
h. pengawasan data tidak memadai.
i. pengkodean, tabulasi, dll. dari data mungkin salah.
j. ada responden yang tidak menjawab secara akurat
k. ada data responden yang digandakan
l. ada data responden yang hilang
m. ada kesalahan dalam menyajikan dan mencetak hasil tabulasi, grafik, dll.

2. Jenis Kesalahan Non-Sampling

Kesalahan non-sampling dapat secara luas diklasifikasikan ke dalam tiga kategori.

a. Kesalahan spesifikasi: Kesalahan ini terjadi pada tahap perencanaan karena berbagai alasan, misalnya, spesifikasi data yang tidak memadai dan tidak konsisten sehubungan dengan tujuan survei/sensus, penghilangan atau duplikasi satuan karena definisi yang tidak tepat, metode pencacahan yang salah/ wawancara/jadwal ambigu dll.
b. Kesalahan penetapan: Kesalahan ini terjadi pada tahap pelaksanaan di lapangan karena berbagai alasan misalnya, kurangnya tenaga investigasi yang terlatih dan berpengalaman, kesalahan penarikan

kembali dan jenis kesalahan lain dalam pengumpulan data, kurangnya inspeksi yang memadai dan kurangnya pengawasan, dll.

c. Kesalahan tabulasi: Kesalahan ini terjadi pada tahap tabulasi karena berbagai alasan, misalnya, pemeriksaan data yang tidak memadai, kesalahan dalam memproses data, kesalahan dalam mempublikasikan hasil tabulasi, grafik, dll.

Kesalahan non-sampling dapat terjadi di semua aspek proses survei, dan dapat diklasifikasikan ke dalam kategori berikut:

a. Kesalahan cakupan.

Kesalahan cakupan terdiri dari kelalaian (undercoverage), kesalahan penyertaan, duplikasi dan kesalahan klasifikasi (overcoverage) unit dalam kerangka survei. Karena mempengaruhi setiap perkiraan yang dihasilkan oleh survei, mereka adalah salah satu jenis kesalahan yang paling penting. Dalam kasus sensus, itu mungkin menjadi sumber kesalahan utama. Kesalahan cakupan dapat memiliki dimensi spasial dan temporal, dan dapat menyebabkan bias dalam estimasi. Efeknya dapat bervariasi untuk berbagai subkelompok populasi. Kesalahan ini cenderung sistematis dan biasanya karena under coverage, oleh karena itu penting untuk menguranginya sebanyak mungkin.

b. Kesalahan Pengukuran.

Kesalahan pengukuran, juga disebut kesalahan respons, adalah perbedaan antara nilai terukur dan nilai sebenarnya. Ini terdiri dari bias dan varians, dan hal tersebut terjadi ketika data yang diperoleh secara tidak benar. Kesalahan ini dapat terjadi karena inefisiensi dengan kuesioner, interviewer, responden atau proses survei.

c. Rancangan Kuisioner yang baik

Sangat penting bahwa pertanyaan kuisioner disajikan dengan hati-hati untuk menghindari bias. Jika pertanyaan membingungkan, maka tanggapannya mungkin terdistorsi. Kuisioner yang baik dibuat simple dan mudah dimengerti, memiliki ukuran yang jelas sesuai pemahaman responden, Pertanyaan yang tidak terlalu banyak/panjang agar responden memberi jawaban yang serius, pertanyaan bersifat tertutup sesuai informasi yang dibutuhkan, dan pertanyaan tidak bersifat ambigu agar responden memberikan jawaban yang sesuai.

d. Interviewer bias

Seorang pewawancara dapat mempengaruhi bagaimana responden menjawab pertanyaan survei. Ini mungkin terjadi ketika pewawancara terlalu ramah atau menyendiri atau mendorong responden. Untuk mencegah hal ini, pewawancara harus dilatih untuk tetap netral selama wawancara. Mereka juga harus memperhatikan cara mereka mengajukan setiap pertanyaan. Jika pewawancara mengubah cara pertanyaan diucapkan, hal itu dapat memengaruhi jawaban responden.

e. Kesalahan Responden

Responden juga dapat memberikan jawaban yang salah. Ingatan yang salah, kecenderungan untuk melebih-lebihkan atau meremehkan peristiwa, dan kecenderungan untuk memberikan jawaban yang tampak lebih dapat diterima secara sosial adalah beberapa alasan mengapa responden dapat memberikan jawaban yang salah.

f. Permasalahan terhadap Proses Survey

Kesalahan juga dapat terjadi karena masalah dengan proses survei yang sebenarnya. Menggunakan tanggapan proxy, yang berarti mengambil jawaban dari orang lain selain responden, atau kurang kontrol atas prosedur survei hanyalah beberapa cara untuk meningkatkan risiko kesalahan tanggapan.

g. Kesalahan Non-Respon

Estimasi yang diperoleh setelah nonresponse telah diamati dan imputasi telah digunakan untuk menangani nonresponse ini biasanya tidak setara dengan estimasi yang akan diperoleh jika semua nilai yang diinginkan telah diamati tanpa kesalahan. Perbedaan antara kedua jenis perkiraan ini disebut kesalahan nonresponse. Ada dua jenis kesalahan non-respons: total dan parsial.

Kesalahan nonrespons total terjadi ketika semua atau hampir semua data untuk unit sampling hilang. Hal ini dapat terjadi jika responden tidak ada atau tidak hadir untuk sementara waktu, responden tidak dapat berpartisipasi atau menolak untuk berpartisipasi dalam survei, atau jika tempat tinggal kosong. Jika sejumlah besar unit sampel tidak menanggapi survei, maka hasilnya mungkin bias karena karakteristik non-responden mungkin berbeda dari mereka yang telah berpartisipasi.

Kesalahan nonrespon sebagian terjadi ketika responden memberikan informasi yang tidak lengkap. Bagi orang-orang tertentu, beberapa pertanyaan mungkin sulit dipahami, mereka mungkin menolak atau lupa menjawab pertanyaan. Kuesioner yang dirancang dengan buruk atau teknik wawancara yang buruk juga dapat menjadi alasan yang menghasilkan kesalahan nonrespon sebagian. Untuk mengurangi bentuk kesalahan ini, perhatian harus diberikan dalam merancang dan menguji kuesioner. Pelatihan pewawancara yang memadai dan strategi edit dan imputasi yang tepat juga akan membantu meminimalkan kesalahan ini.

h. Kesalaham Pemrosesan

Kesalahan pemrosesan terjadi selama pemrosesan data. Ini mencakup semua kegiatan pemrosesan data setelah pengumpulan dan sebelum estimasi, seperti kesalahan dalam pengambilan data, pengkodean, pengeditan dan tabulasi data serta dalam penetapan bobot survei.

Kesalahan pengkodean terjadi ketika pembuat kode yang berbeda mengkodekan jawaban yang sama secara berbeda, yang dapat disebabkan oleh pelatihan yang buruk, instruksi yang tidak lengkap, perbedaan dalam kinerja pembuat kode (yaitu kelelahan, sakit), kesalahan entri data, atau kerusakan mesin (beberapa kesalahan pemrosesan disebabkan oleh kesalahan dalam program komputer).

Kesalahan pengambilan data terjadi ketika data tidak dimasukkan ke komputer persis seperti yang muncul pada kuesioner. Hal ini dapat disebabkan oleh kompleksitas data alfanumerik dan kurangnya kejelasan jawaban yang diberikan. Tata letak fisik kuesioner itu sendiri atau dokumen pengkodean dapat menyebabkan kesalahan pengambilan data. Metode pengambilan data, manual atau otomatis (misalnya, menggunakan pemindai optik), juga dapat menyebabkan kesalahan.

Kesalahan pengeditan dan imputasi dapat disebabkan oleh kualitas data asli yang buruk atau oleh strukturnya yang kompleks. Ketika proses pengeditan dan imputasi diotomatisasi, kesalahan juga dapat disebabkan oleh program yang salah yang tidak cukup diuji. Pilihan metode imputasi yang tidak tepat dapat menimbulkan bias. Kesalahan juga dapat terjadi akibat salah mengubah data yang ternyata salah, atau karena salah mengubah data yang benar.

3. Estimasi/ Perkiraan Kesalahan Non-Sampling

a. Kompoinen total kesalahan survey

Angka perkiraan kesalahan non-sampling sangat rumit untuk diukur dan diketahui dengan tepat. Namun salah satu upaya yang dilaksanakan oleh BPS untuk mereduksi kesalahan non-sampling melalui penggunaan aplikasi Post Enumeration Survey (PES). Kegiatan PES mencakup pencacahan ulang pada sampel yang mewakili sensus, pencocokan dua arah (two-way matching) pada setiap individu yang dicatat di PES dengan informasi dari sensus, dan kunjungan rekonsiliasi lapangan untuk mengkonfirmasi keberadaan individu yang keberadaannya diragukan, baik di sensus maupun sutrvey. PES dilakukan dengan upaya penyempurnaan dalam berbagai tahap kegiatan mulai dartahapan persiapan, pelaksanaan lapang hingga pengolahan dan penyajian iantara lain melalui: pengerahan petugas yang berkualitas, pelatihan petugas lapangan yang intensif, pengawasan yang cukup ketat di lapangan, dan pengolahan data yang cermat (BPS, 2007).

Kesalahan survei adalah fungsi dari perbedaan antara nilai sebenarnya rata-rata populasi secara keseluruhan dan nilai rata-rata yang diamati yang diperoleh dari responden sampel tertentu. Total kesalahan Survey = f($\bar{X}$- $\bar{x}_{\text{true}}$)

Total kesalahan survei dapat dibagi menjadi dua komponen, kesalahan sampling dan kesalahan non-sampling. Kesalahan sampling menunjukkan seberapa akurat nilai rata-rata sebenarnya dari sampel yang dipilih ($\bar{x}_s$) mewakili nilai rata-rata sebenarnya dari populasi ($\bar{X}$). Kesalahan non-sampling, di sisi lain menunjukkan seberapa baik nilai rata-rata yang diamati yang diperoleh dari responden sampel ($\bar{x}_{\text{true}}$) mewakili nilai rata-rata sampel yang sebenarnya ($\bar{x}_s$)

Dengan demikian dapat ditulis bahwa, total kesalahan survey = kesalahan sampling + kesalahan nonsampling, atau:

$$f(\bar{X} - \bar{x}_{\text{true}}) = f(\bar{X} - \bar{x}_s) + f(x_s - \bar{x}_{true})$$

b. Bias non-sampling

Untuk tujuan simplifikasi, diasumsikan dua tahap pengambilan random. Misalkan, $\bar{x}_{sr}$ adalah estimasi rata-rata populasi X

berdasarkan sampel tertentu. Nilai harapan $\bar{x}_{sr}$ yang diambil dari tahap kedua random untuk unit sampel tetap adalah :

$$E(\bar{x}_{sr}) = \bar{x}_{so}$$

Nilai tersebut mungkin berbeda dengan rata-rata sampel acak $\bar{x}_s$ yang merupakan nilai dari sampel dengan ukuran yang mewakili populasi. Nilai yang diharapkan dari $\bar{x}_{so}$_setelah tahap pertama ramdom.

$$E(\bar{x}_{so}) = \bar{x}_{true}$$

yang merupakan nilai yang diperoleh penduga tak bias melalui proses survei yang ditentukan. Nilai $\bar{x}^*$ mungkin berbeda dari rata-rata populasi sebenarnya $\bar{X}$. Total bias adalah:

$$Bias_t(\bar{x}_{sr}) = \bar{x}_{true} - \bar{X}$$

Bias sampling :

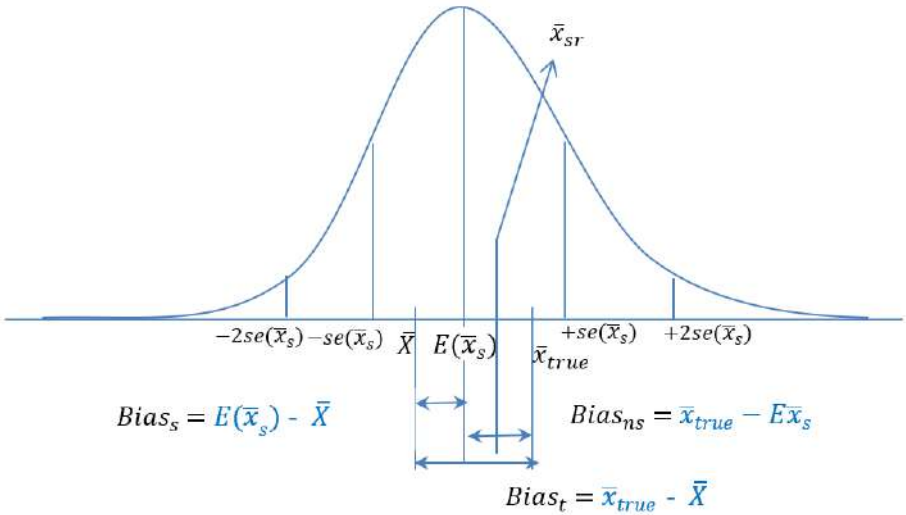
$$Bias_s(\bar{X}) = E(\bar{x}_s) - \bar{X}$$

Bias non-sampling:

$$Bias_{ns}(\bar{x}_{sr}) = Bias_t(\bar{x}_{sr}) - Bias_s(\bar{x}_s) = \bar{x}_{true} - E(\bar{x}_s) = E(\bar{x}_{so} - \bar{x}_s)$$

Ini merupakan nilai ekspektasi dari deviasi non-sampling.

Hubungan antara bias total, bias sampling, dan bias non sampling ditunjukan pada gambar distribusi normal dari sampel sebenarnya berikut.



Gambar 8.3. Hubungan antara bias total, bias sampling, dan bias non sampling

DAFTAR PUSTAKA

BPS, 2007, Sampling error survei sosial ekonomi nasional, Jakarta, Indonesia

***BPS, 2020**, Metode penarikan sampeL,* Bahan Ajar Diklat Statistisi Ahli BPS Angkatan XXI, *Jakarta, Indonesia*

Khadka J., 2019, *Sampling error in survey research*, International Journal of Science and Research (IJSR), Vol . 8 (1).

Nasional Science Board-US, 2016, *Science & engineering indicators*, Alexandria-Virginia, US. diakses dari: https://www.nsf.gov/statistics/2016/nsb20161/#/report/appendix-methodology/data-accuracy/sampling-error

Shalabh, 2020. *Lecture note*: Sampling theory, Department of Mathematics & Statistics IIT Kanpur – India, diakses dari: http://home.iitk.ac.in/~shalab/sampling/chapter13-sampling-non-sampling-errors.pdf

Srimali, Wickramasinghe, dan Aponsu, 2015, *Identification of factors affecting to Non-Sampling Errors*: Study on Economic Census 2013/2014, Sri Langka.

Statistics Canada, 2021, *Data gathering and processing*, diakses dari: https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch6/5214809-eng.htm

PROFIL PENULIS

Dr. La One ST, MT., Lahir di Raha Kabupaten Muna Provinsi Sulawesi tenggara pada tanggal 14 Juli 1971. Menyelesiakan Pendidikan SD Negeri 10 Raha Tahun 1995, SMP 1 Raha Tahun 1988, SMA 1 Raha Tahun 1991. Kemudian menamatkan Pendidikan S1 Jurusan Teknik Sipil Fakultas Teknik Universitas Hasanuddin tahun 1996, Menyelesaikan Pendidikan S2 Magister Sistem dan Teknik Transportasi Universitas Gadjah Mada tahun 2002 dengan masa studi selama 13 bulan. Pendidikan S3 ditempuh di Program Studi Teknik Sipil Universitas Hasanuddin Tahun 2016-2020. Tahun 1997-1998 menjadi Dosen Honorer pada D3 Teknik Sipil Universitras Haluoleo. Sejak tahun 1998 hingga tahun 2020 bekerja di Pemda Kabupaten Muna dengan jabatan yang pernah diemban sebagai Sekretaris Bappeda Kabupaten Muna tahun 2014-2016, dan Tahun 2020-2021 menduduki Jabatan Fungsional Perencana Madya. Terhitung Mulai 1 Maret 2021 beralih tugas sebagai Dosen Jurusan Teknik Sipil Univerrsitas Haluoleo.

BAB 9
UJI NORMALITAS

Rahma Faelasofi, S.Si., M.Sc.
Universitas Muhammadiyah Pringsewu

A. PENGANTAR

Uji normalitas dimaksudkan untuk mendeteksi apakah data yang akan digunakan sebagai bahan acuan dari terbukti memenuhi asumsi normalitas atau sebaliknya, atau uji normalitas merupakan salah satu uji yang dilakukan untuk mengukur apakah data yang dimiliki berdistribusi normal atau tidak. Dampak dari pelaksanaan uji normalitas, berakibat pada penentuan uji statistik yang digunakan. Jika suatu data tersebut memenuhi asumsi data berdistribusi normal maka uji statistik yang digunakan yaitu statistik parametrik, sebaliknya jika asumsi uji normalitas data tidak terpenuhi maka yang digunakan adalah uji statistik non parametrik.

Suatu data dikatakan normal, jika data yang tersedia mempunyai distribusi data yang normal. Suatu data yang dinyatakan berdistribusi normal berdasarkan keterpenuhan nilai dari rata-rata dan standar deviasi yang sama. Sehingga bisa dinyatakan bahwa, uji normalitas pada dasarnya melakukan perbandingan antara data yang dimiliki dengan data yang berdistribusi normal memiliki nilai rata-rata dan standar deviasi yang sama dengan data yang diujikan.

Perolehan data yang akan diujikan memenuhi distribusi normal atau tidak, berdasarkan pada proses randomisasi pengambilan sampel, dengan harapan bahwa data yang diperoleh menggambarkan kesesuaian dari kondisi yang wajar dari fenomena alami aspek yang diukur. Data yang diperoleh dengan memperhatikan proses pengambilan sampel yang memenuhi aspek random, respon, dari sampel penelitian sebagai wakil populasi, diharapkan diasumsikan memiliki kewajaran. Kecenderungan fenomena alami yang berpola seragam dan respon yang wajar tersebut memberikan data yang tidak jauh menyimpang dari kecenderungannya, yaitu kecenderungan terpola/terpusat.

Banyak cara yang dapat digunakan untuk membuktikan data berdistribusi normal atau tidak. Mulai dari metode klasik dalam pengujian normalitas suatu data dengan menggunakan sesuatu yang tidak begitu rumit. Selanjutnya berdasarkan pengalaman empiris, data yang banyaknya lebih dari 30 ($n > 30$), sudah dapat diasumsikan data memenuhi asumsi berdistribusi normal, atau semakin besar sampel yang digunakan dalam penelitian, semakin besar peluangnya untuk data tersebut memenuhi asumsi berdistribusi normal,

berdasarkan kajian teori pada teorema limit pusat, yang menyatakan data dengan ukuran sampel yang besar akan terdistribusi normal.

Namun, untuk membuktikan sebagai upaya untuk menunjukkan kepastian dari data tersebut, bahwa data yang diujikan berdistribusi normal atau tidak, sebaiknya digunakan uji statistik normalitas data, dikarenakan untuk lebih memastikan apakah data dengan jumlah sampel n>30 berdistribusi normal atau tidak. Atau sebaliknya, pada saat data sampel yang anggota sampelnya n < 30, belum tentu peluang data untuk memenuhi asumsi berdistribusi normal menjadi kecil, Oleh karena itu untuk data dengan sampel besar n>30 ataupun untuk data dengan sampel kecil n<30, tetap memerlukan proses pembuktian data memenuhi asumsi berdistribusi normal atau tidak.

B. MACAM-MACAM UJI NORMALITAS

Uji normalitas dalam membuktikannya, dilakukan dengan beberapa cara, yaitu:

1. Uji grafik

   Pada uji grafik yang dilakukan adalah dengan memperhatikan penyebaran data pada sumber diagonal pada grafik p-p plot of regression standardized residual. Data yang dinyatakan berdistribusi normal apabila sebaran titik-titik berada pada sekitaran garis dan mengikuti garis diagonal, maka data tersebut dinyatakan normal.
2. Metode chi-square

   Metode chi square atau X^2 untuk uji Goodness of Fit Distribusi Normal menggunakan pendekatan penjumlahan penyimpangan data observasi tiap kelas dengan nilai yang diharapkan. Persyaratan metode chi-square (uji goodness of fit Distribusi Normal)
   a. Data tersusun berkelompok atau dikelompokkan dalam tabel distribusi frekuensi
   b. Sesuai, untuk data dengan banyaknya jumlah sampel, n>30
3. Metode Liliefors

   Metode liliefors menggunakan data dasar yang belum diolah dalam tabel distribusi frekuensi. Data ditransformasikan dalam nilai Z untuk dapat dihitung luasan kurva normal sebagai probabilitas kumulatif normal.

Persyaratan metode liliefors:

a. Data berskala interval atau rasio (kuantitatif)
b. Data tunggal/belum dikelompokkan pada tabel distribusi frekuensi
c. Dapat digunakan untuk sampel dengan n besar maupun n kecil

4. Metode Kolmogorov-Smirnov

   Metode Kolmogorov-Smirnov tidak jauh berbeda dengan metode Liliefors. Metode Kolmogorov-Smirnov merupakan pengujian normalitas yang banyak dipakai, kelebihan dari uji dengan menggunakan metode ini adalah kesederhanaannya dan tidak menimbulkan perbedaan persepsi di antara satu pengamat dengan pengamat yang lain, yang sering terjadi pada uji normalitas dengan menggunakan grafik. Sedangkan kelemahan dari metode Kolmogorov-Smirnov kesimpulan yang dihasilkan dapat memberikan hasil yang tidak normal, maka kita tidak bisa menentukan transformasi seperti apa yang harus kita gunakan untuk normalisasi. Langkah-langkah penyelesaian dan penggunaan rumus sama, namun pada penggunaan nilai signifikansi yang berbeda. Signifikansi metode Kolomogorov-Smirnov menggunakan tabel pembanding Kolmogorov-Smirnov, sedangkan metode Liliefors menggunakan tabel pembanding metode Liliefors. Persyaratan metode Kolmogorov-Smirnov:

   a. Data berskala interval atau ratio (kuantitatif)
   b. Data tunggal/belum dikelompokkan pada tabel distribusi frekuensi
   c. Dapat digunakan untuk sampel dengan n besar maupun n kecil

5. Metode Shapiro Wilk

   Metode Shapiro Wilk menggunakan data dasar yang belum diolah dalam tabel distribusi frekuensi. Data diurut, kemudian dibagi menjadi dua kelompok untuk dikonversi dalam Shapiro-Wilk, dan dilanjutkan dengan melakukan transformasi dalam nilai Z untuk dapat dihitung luasan kurva normal. Persyaratan metode Shapiro-Wilk:

   b. Data berskala interval atau ratio (kuantitatif)
   c. Data tunggal/belum dikelompokkan pada tabel distribusi frekuensi
   d. Dapat dari sampel random

C. PROSES DAN CONTOH UJI NORMALITAS

Proses dan contoh uji normalitas dengan beberapa cara, yaitu:

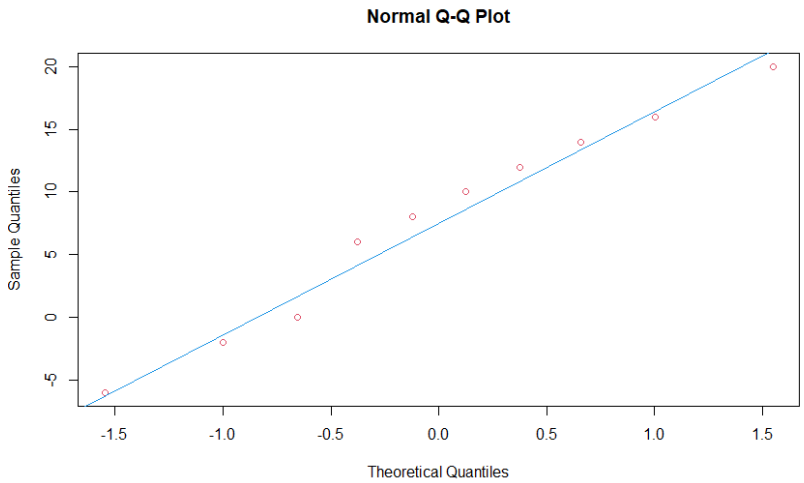
1. Uji grafik

   Berikut ini disajikan uji normalitas dalam bentuk grafik melalui perolehan software R, berikut ini beberapa langkah yang dapat dilakukan saat menyajikan normal Q-Q plot dengan menggunakan R:

   a. Entry data pada lembar kerja R
   b. Gunakan perintah syntax:

      #visual methode
      qqplot(x,y,col=”red”)
      #membuat qq plot
      qqnorm(y,col=2) #membuat lingkaran warna dengan perwanaan warna merah
      qqline(y,col=4) #membuat tampilan line/garis dengan pewarnaan warna biru

      Berdasarkan perintah syntax tersebut diperoleh hasil berikut ini:



Grafik q-q plot menunjukkan sebaran data sebanyak 10 item (lingkaran merah) menyebar di sepanjang garis diagonal, sehingga data dapat dikatakan berdistribusi normal. Namun, sebaliknya, jika data tidak menyebar di sepanjang garis diagonal, maka data tidak

berdistribusi normal. Namun untuk mendapatkan hasil yang lebih akuratf yaitu dengan dilengkapi uji normalitas dengan menggunakan uji statistik yang lain.

2. Metode chi-square

Berikut ini terdapat rincian data tunggal mengenai data mahasiswa yang mendapat nilai ujian matematika sebanyak 30, sebagai berikut:

75	74	74	73	76	77	87	67	56	78
78	67	76	66	65	67	67	76	78	77
77	77	80	87	89	89	89	89	91	85

Ujilah apakah data tersebut berdistribusi normal atau tidak dengan $\alpha = 0{,}05$?

Langkah-langkah uji normalitas dengan metode chi-square:

(1) Merumuskan hipotesis

H_0 : data berdistribusi normal

H_1 : data tidak berdistribusi normal

(2) Membuat tabel bantu untuk penyajian data

(3) Menentukan taraf nyata (α)

Untuk mendapatkan nilai chi-square tabel, diperoleh dengan:

χ^2 tabel = $\chi^2_{df,\alpha}$

df = derajat kebebasan

$df = k - 3$

k = banyak kelas interval

α = level signifikansi = 5% = 0,05

(4) Menentukan nilai uji statistic:

Mencari nilai Z score dan Chi-square

Persamaan Z score: $Z = \frac{\bar{X}-\mu}{\sigma}$

Ket: $\bar{x}$ = nilai rata-rata yang diamati

μ = rata-rata populasi

σ = standar deviasi populasi

Z = Z score (nilai baku)

Persamaan chi-square: $\chi^2 = \sum \frac{(O_i-E_i)^2}{E_i}$

(5) Menentukan kriteria pengujian hipotesis

H_0 ditolak jika χ^2 hitung $\geq \chi^2$ tabel
H_0 diterima jika χ^2 hitung $< \chi^2$ tabel

Penyelesaian:

(1) Merumuskan hipotesis
H_0 : data berdistribusi normal
H_1 : data tidak berdistribusi normal

(2) Membuat tabel bantu untuk penyajian data

No	Kelas interval	f	x_i	x_i^2	fx_i	fx_i^2
1	56-61	1	58,5	3422,25	58,5	3422,25
2	62-67	6	64,5	4160,25	387	24961,5
3	68-73	1	70,5	4970,25	70,5	4970,25
4	74-79	13	76,5	5852,25	994,5	76079,25
5	80-85	2	82,5	6806,25	165	13612,5
6	86-91	7	88,5	7832,25	619,5	54825,75
	$\sum$				2295	177871,5

(3) Menentukan chi-square dengan taraf nyata (α)
Untuk mendapatkan nilai chi-square tabel:
Rumus chi-square (χ^2) tabel:
df = derajat kebebasan
$df = 6 - 3 = 3$
k = banyak kelas interval
α = level signifikansi = 5% = 0,05
χ^2 tabel = $\chi^2{}_{df,\alpha} = \chi^2{}_{3,0,05} = 7,81$

(4) Menentukan nilai uji statistic:
Nilai Z score dan chi-square

$$\bar{x} = \frac{\sum f.x_i}{n} = \frac{2295}{30} = 76,5$$

Simpangan baku (s) diperoleh sebagai:

$$s = \sqrt{\frac{n.\sum fx_i^2 - (\sum fx_i)^2}{n.(n-1)}} = \sqrt{\frac{30 \times 177871,5 - (2295)^2}{30.(30-1)}} = 8,91$$

Kelas	Frek (f_i)	Tepi kelas	Nilai Z	Luas 0-Z	Luas Kelas Interval	Frek harapan (E_i)	$\frac{(O_i - E_i)^2}{E_i}$
56-61	1	55,5	-2,3569	0,491	0,0374	1,122	0,0133
62-67	6	61,5	-1,6835	0,454	0,1097	3,291	2,2299
68-73	1	67,5	-1,0101	0,344	0,2107	6,321	4,4792
74-79	13	73,5	-0,3367	0,133	0,2662	7,986	3,1480
80-85	2	79,5	0,3367	0,133	0,2107	6,321	2,9538
86-91	7	85,5	1,0101	0,344	0,1097	3,291	4,1801
		91,5		0,454		χ^2	17,0043

Luas interval adalah harga mutlak, luas interval kelas 1 adalah 0,491-0,454=0,0374;

Luas interval kelas ke-2 adalah 0,454-0,344=0,1097; dan seterusnya

(5) Menentukan kriteria pengujian hipotesis

H_0 ditolak jika χ^2 hitung $\geq \chi^2$ tabel

H_0 diterima jika χ^2 hitung $< \chi^2$ tabel

χ^2 hitung = 17,0043

χ^2 tabel = 7,81

(6) Memberikan kesimpulan

Karena χ^2 hitung > χ^2 tabel, yaitu 12,017 > 7,81,

Maka kesimpulannya tidak cukup bukti untuk tolak H_0, data ujian matematika tidak berdistribusi normal.

3. Metode Liliefors

Berikut ini terdapat rincian data tunggal yang akan dibuktikan apakah data berdistribusi normal atau sebaliknya. Berikut ini data nilai matematika siswa didik Sekolah Dasar sebanyak 12 orang, dengan rincian sebagai berikut:

Dina	Adi	Bela	Daffi	Gilang	Ananda	Putri	Sekar	Bayu	Ayu	Desi	Bima
23	27	33	40	48	48	57	59	62	68	69	70

Berdasarkan data di atas apakah dapat ditarik kesimpulan bahwa populasi asal sampel nilai matematika siswa didik Sekolah Dasar adalah normal?

Proses melakukan uji normalitas:

(1) Mengurutkan data dari nilai terkecil sampai terbesar

(2) Menentukan besaran rata-rata dari sampel

(3) Menentukan besaran standar deviasi dari sampel

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

$$= \sqrt{\frac{33{,}414 - \frac{604^2}{12}}{11}} = \sqrt{\frac{3012}{11}} = 16{,}55$$

(4) Menentukan nilai standarisasi dari tiap anggota dalam sampel

$$z_i = \frac{X_i - \bar{X}}{s}$$

Ex:

$$z_1 = \frac{23-50{,}3}{16{,}55} = -1{,}65$$

(5) Menentukan besaran nilai probabitas distribusi norma (z), berdasarkan tabel distribusi normal

(6) Menentukan nilai F(z) peluang dari distribusi normal

$$F(z_1) = 0{,}5 - 0{,}4505 = 0{,}0495$$

(7) Menentukan rasio urutan dari tiap sampel yang ada

Ex:

$$s(z_1) = \frac{1}{12} = 0{,}0833$$

(8) Menentukan nilai absolut atau nilai mutlak selisih antara $F(z_1)$ dengan $s(z_1)$, sehingga berlaku $|F(z_1) - s(z_1)|$

Perolehan dalam bentuk tabel, dapat dilihat berikut ini:

X_i	z_i	$F(z_i)$	$s(z_i)$	$\lvert F(z_i) - s(z_i)\rvert$
23	-1,65	0,0495	0,0833	0,0338
27	-1,41	0,0793	0,1677	0,0874
33	-1,05	0,1469	0,2500	0,103
40	-0,62	0,2675	0,3333	0,0657
48	-0,14	0,4443	0,5000	0,0557
48	-0,14	0,4443	0,5000	0,0577
57	0,40	0,6554	0,5833	0,0721
59	0,53	0,7019	0,6667	0,0352
62	0,71	0,7612	0,7500	0,0112
68	1,07	0,8577	0,8333	0,0244
69	1,13	0,8708	0,9167	0,0459
70	1,19	0,8830	1	0,1170
Σ=604 $\bar{X} = 50,3$				

Kesimpulan:

Nilai tertinggi dari $|F(z_i) - s(z_i)|$ atau L_o adalah 0,1170 (tidak selalu ditunjukkan nilai terendah pada urutan baris terakhir). $L_t = 0{,}242$. Karena L_o lebih rendah dari L_t atau L_o (0,1170) < L_t(0,242), dapat disimpulkan bahwa sampel berdistribusi normal. Selain itu, jika L_o lebih besar dibandingkan L_t atau $L_o \geq L_t$, dapat disimpulkan bahwa sampel tidak berdistribusi normal.

4. Metode Kolmogorov-Smirnov

Berikut ini data tunggal yang akan dibuktikan data berdistribusi normal atau sebaliknya. Hasil dari skor perolehan penilaian dari suatu lomba adalah sebagai berikut:

1348	1140	1086	1039	920
1233	1146	1002	1012	904
1255	1168	1016	1001	973

Berdasarkan data di atas apakah dapat ditarik kesimpulan bahwa populasi asal sampel adalah normal?

Langkah-langkah melakukan uji normalitas dengan metode Kolmogorov-Smirnov:

(1) Susun hipotesis
(2) Susun frekuensi-frekuensi dari tiap nilai teramati, berurutan dari nilai terkecil sampai nilai terbesar. Kemudian susun frekuensi kumulatif dari nilai-nilai teramati.
(3) Konversikan frekuensi kumulatif itu kedalam probabilitas, yaitu ke dalam fungsi distribusi frekuensi kumulatif $[S(x)]$. Distribusi frekuensi teramati harus merupakan hasil pengukuran variable paling sedikit dalam skala ordinal (tidak dalam skala nominal)
(4) Hitung nilai z untuk masing-masing nilai teramati di atas dengan rumus $z = {(x_I - \bar{x})}/{_S}$, dengan mengacu kepada tabel distribusi normal baku, carilah peluang (luas area) kumulatif untuk setiap nilai teramati. Hasilnya ialah sebagai $F_t(x_i)$.
(5) Susun $F_s(x)$ berdampingan dengan $F_t(x)$. Hitung selisih nilai absolut antara $S(x)$ dan $F_t(x)$ pada masing-masing nilai teramati.
(6) Statistik uji Kolmogorov-Smirnov ialah selisih absolut terbesar $F_s(x_i)$ dan $F_t(x_i)$ yang juga disebut deviasi maksimum D.
(7) Dengan mengacu pada distribusi pencuplikan bisa diketahui apakah perbedaan sebesar itu (yaitu nilai D maksimum teramati) terjadi hanya karena kebetulan. Dengan mengacu pada tabel D, dilihat berapa peluang (dua sisi) kejadian untuk menemukan nilai-nilai teramati sebesar D, bila H_0 benar. Jika peluang itu sama atau lebih kecil dari nilai tabel Kolmogorov-smirnov, maka H_0 ditolak.

Penyelesaian:

(1) Susun hipotesis:
$H_0: F(x) = F_t(x), \forall x$
$H_1: F(x) \neq F_t(x), paling\ sedikit\ satu\ x$
(2) Urutkan data dari nilai terkecil ke nilai terbesar

904	920	973	1001	1002
1012	1016	1039	1086	1140
1146	1168	1233	1255	1348

(3) Hitung distribusi $F_s(x_i)$ dengan rata-rata 1083 dan simpangan baku 129

x_i	f	f_{kum}	$F_s(x)$
904	1	1	0,067
920	1	2	0,133
973	1	3	0,2
1001	1	4	0,267
1002	1	5	0,333
1012	1	6	0,4
1016	1	7	0,467
1039	1	8	0,533
1086	1	9	0,6
1140	1	10	0,667
1146	1	11	0,733
1168	1	12	0,8
1233	1	13	0,867
1255	1	14	0,933
1348	1	15	1

(4) Hitung $F_t(x_i)$ dibantu tabel distribusi normal baku z, $z = {(x_i - \bar{x})}/{s}$
Diperoleh $\bar{x} = 1082{,}87$ dan $s = 128{,}792$

x_i	$x_i - \bar{x}$	${(x_i - \bar{x})}/{s}$	$F_t(x_i)$
904	-178,867	-1,389	0,082
920	-162,867	-1,265	0,104
973	-109,867	-0,853	0,198
1001	-81,867	-0,636	0,261
1002	-80,867	-0,628	0,264
1012	-70,867	-0,550	0,291
1016	-66,867	-0,519	0,302
1039	-43,867	-0,341	0,367
1086	3,133	0,024	0,508
1140	57,133	0,444	0,67
1146	63,133	0,490	0,688
1168	85,133	0,661	0,745

1233	150,133	1,166	0,877
1255	172,133	1,337	0,908
1348	265,133	2,059	0,980

(5) Hitung D, tentukan D_{max}

$F_s(x)$	$F_t(x_i)$	$\lvert F_s(x) - F_t(x_i)\rvert$
0,067	0,082	0,015
0,133	0,104	0,029
0,2	0,198	0,002
0,267	0,261	0,006
0,333	0,264	0,069
0,4	0,291	0,109
0,467	0,302	0,165
0,533	0,367	0,166
0,6	0,508	0,092
0,667	0,67	0,003
0,733	0,688	0,045
0,8	0,745	0,055
0,867	0,877	0,01
0,933	0,908	0,025
1	0,980	0,02

(6) Lihat tabel Kolmogorov Smirnov, dengan $\alpha = 0{,}05, n = 15$ diperoleh $k = 0{,}338$

(7) Keputusan uji:
Karena $D = 0{,}166 < K = 0{,}338$
Maka tidak cukup bukti untuk menolak H_0
Ini berarti sampel skor perolehan penilaian dari suatu lomba berasal dari distribusi normal.

5. Metode Shapiro Wilk

Berikut ini data tunggal yang akan dibuktikan data berdistribusi normal atau sebaliknya.

Berdasarkan data usia sebagian balita diambil sampel secara random sebanyak 24 balita, diperoleh data usia balita berikut ini (bulan):

58	23	58	56	46	37	18	32
36	19	34	33	41	36	55	30
24	36	33	26	40	35	48	27

Selidikilah data usai balita, apakah data tersebut berdistribusi normal pada $\alpha = 5\%$?

Langkah-langkah uji normalitas dengan metode Shapiro Wilk:

(1) Merumuskan hipotesis

H_0 : data berdistribusi normal

H_1 : data tidak berdistribusi normal

(2) Menentukan taraf nyata (α)

α = level signifikansi = 5% = 0,05

(3) Data diurutkan dari yang terkecil hingga terbesar, serta dibagi menjadi dua kelompok untuk dikonversi dalam Shapiro Wilk

(4) Menentukan nilai uji statistic:

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

Ket: $\bar{x}$ = nilai rata-rata yang diamati

x_i = angka ke-i pada data yang diamati

$$T_3 = \frac{1}{D}\left[\sum_{i=1}^{k} a_i (x_{n-i+1} - x_i)\right]^2$$

Ket:

T_3 = konversi statistic Shapiro-wilk pendekatan distribusi normal

a_i = koefisien test Shapiro wilk

x_{n-i+1} = data ke $n - i + 1$

(5) Menentukan signifikansi uji

Dalam menentukan signifikansi uji menggunakan tabel Shapiro Wilk dilihat dari posisi nilai peluangnya (p). Jika $p \geq \alpha$ maka terima hipotesis nol H_0. Sebaliknya jika $p < \alpha$ maka tolak hipotesis nol H_0.

(6) Selanjutnya dapat dilakukan transformasi dalam nilai Z untuk dapat dihitung luasan kurva normal. Signifikansi uji kemudian ditentukan berdasarkan nilai kritis dari kurva normal. Berikut adalah transformasi yang digunakan:

$$G = b_n + c_n + ln\left[\frac{T_3 - d_n}{1 - T_3}\right]$$

Dimana

G = identic dengan nilai Z distribusi normal

b_n, c_n, ln = konversi statistic Shapiro Wilk pendekatan distribusi normal

Penyelesaian:

(1) Merumuskan hipotesis

H_0 : data berdistribusi normal

H_1 : data tidak berdistribusi normal

(2) Menentukan taraf nyata (α)

α = level signifikansi = 5% = 0,05

(3) Data diurutkan dari yang terkecil hinga terbesar, dan menghitung nilai uji Shapiro Wilk

No	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	18	-18,71	350,06
2	19	-17,71	313,64
3	23	-13,71	187,96
4	24	-12,71	161,54
5	26	-10,71	114,70
6	27	-9,71	94,28
7	30	-6,71	45,02
8	32	-4,71	22,18
9	33	-3,71	13,76
10	33	-3,71	13,76
11	34	-2,71	7,34
12	35	-1,71	2,92
13	36	-0,71	0,50
14	36	-0,71	0,50
15	36	-0,71	0,50

16	37	0,29	0,08
17	40	3,29	10,82
18	41	4,29	18,40
19	46	9,29	86,30
20	48	11,29	127,46
21	55	18,29	334,52
22	56	19,29	372,10
23	58	21,29	453,26
24	58	21,29	453,26
Jumlah	881		
$\bar{x}$	36,71		

No	a_i	$(x_{n-i+1} - x_i)$	$a_i(x_{n-i+1})$
1	0,4493	58-18=40	17.97
2	0,3098	58-19=39	12,08
3	0,2554	56-23=33	8,43
4	0,2145	55-24=31	6,65
5	0,1807	48-26=22	3,98
6	0,1512	46-27=19	2,87
7	0,1245	41-30=11	1,37
8	0,0997	40-32=8	0,80
9	0,0764	37-33=4	0,31
10	0,0539	36-33=3	0,16
11	0,0321	36-34=2	0,06
12	0,0107	36-35=1	0,01
		Jumlah	54,69

$$T_3 = \frac{1}{D}\left[\sum_{i=1}^{k} a_i(x_{n-i+1} - x_i)\right]^2 = \frac{1}{3184{,}96}(54{,}69)^2 = 0{,}9391$$

Nilai a_i (koefisien test Shapiro wilk) diperoleh dari tabel Shapiro Wilk dengan n=24. Karena nilai $T_3 = 0{,}9391$ terletak antara nilai $\alpha(0{,}10) = 0{,}930$ dan $\alpha(0{,}50) = 0{,}963$, ini berarti bahwa nilai p terletak antara 0,10 dam 0,50.

(4) Menentukan signifikansi uji
Dengan demikian karena nilai $p > \alpha = 0{,}05$ maka cukup bukti untuk terima hipotesis nol H_0 dan dapat dinyatakan bahwa data usia 24 balita berdistribusi normal.

DAFTAR PUSTAKA

Anwar Hidayat. (2022). Tutorial Rumus Chi Square dan Metode Hitung, https:// https://www.statistikian.com/2012/11/rumus-chi-square.html. Diakses tanggal 20 April 2022.

Tri Cahyono. (2015). Statistik Uji Normalitas. Purwokerto: Yayasan Sanitarian Banyumas (YASAMAS).

Tju Ji Long. (2022). Uji Normalitas Shapiro Wilk. https://jagostat.com/metode-statistika-2/uji-shapiro-wilks. Diakses tanggal 25 April 2022.

Tju Ji Long. (2022). Uji Liliefors. https://jagostat.com/metode-statistika-2/uji-lilliefors. Diakses tanggal 23 April 2022.

Uji Kolmogorov Smirnov. https://fe.unisma.ac.id. Diakses tanggal 25 April 2022.

PROFIL PENULIS

Rahma Faelasofi, lahir di Tanjung Karang Barat pada tanggal 2 Februari 1985, merupakan anak keempat dari lima bersaudara, dari pasangan H. Sunanto (alm) dan Hj. Siti Chotijah. Semasa kecil melewati jenjang Pendidikan di SD N 8 Gedong Air pada tahun 1991 – 1997, SMP N 16 Bandar Lampung 1998 – 2000, SMA N 3 Bandar Lampung 2001 – 2003, selanjutnya melanjutkan S1 di Universitas Lampung tahun 2004 – 2007 dan melanjutkan S2 di Universitas Gadjah Mada pada tahun 2008 – 2010. Riwayat pekerjaan adalah dosen tetap di Universitas Muhammadiyah Pringsewu (UMPRI) pada tahun 2010 sampai sekarang. Amanah yang pernah diemban adalah sebagai staf Pusat Penjaminan Mutu 2014 - 2018, kepala Pusat Penjaminan Mutu 2018 – 2019, dan sebagai Dekan FKIP UMPRI tahun 2019 sampai dengan sekarang. Tugas lain yang diamanahi adalah sebagai sekretaris senat akademik UMPRI, anggota senat akademik FKIP UMPRI, dan asesor BAN-SM Provinsi Lampung. Penulis sekarang sedang melanjutkan program S3 Doktoral di Universitas Lampung, Provinsi Lampung.

BAB 10
UJI HOMOGENITAS

Tri Widyasari, M.Pd
IKIP PGRI Kalimantan Timur

Uji homogenitas digunakan untuk melihat apakah dua kelompok data atau lebih memiliki varians yang sama (homogen) atau tidak. Uji homogenitas termasuk dalam salah satu dari uji persyaratan analisis. Pengujian homogenitas dapat dilakukan secara perhitungan manual dan menggunakan bantuan aplikasi komputer (SPSS).

Pada bab ini, akan dijelaskan tentang pengujian homogenitas secara manual menggunakan dua cara. Cara yang pertama yaitu dengan membandingkan varians terbesar dan varians terkecil menggunakan uji F. Cara ini digunakan untuk menguji homogenitas dari dua kelompok data. Sedangkan untuk menguji homogenitas dari tiga kelompok data atau lebih, digunakan cara yang kedua yaitu dengan uji Bartlett. Pengujian homogenitas menggunakan bantuan aplikasi SPPS juga akan dijelaskan pada bab ini dengan menggunakan uji *levene*. Lebih lanjut, pengujian homogenitas yang telah disebutkan di atas akan diuraikan sebagai berikut:

A. UJI F

Seperti yang sudah diungkapkan sebelumnya, uji F ini digunakan untuk menguji homogenitas dari dua kelompok data. Cara kerjanya adalah dengan membandingkan varians yang terbesar dengan varians yang terkecil. Sehingga, sebelum menggunakan uji F ini untuk menguji homogenitas data, terlebih dahulu harus dihitung varians dari masing-masing kelompok, agar dapat ditentukan varians mana yang terbesar, dan varians mana yg terkecil. Secara ringkas, rumus yang digunakan adalah:

$$\boldsymbol{F = \frac{varians\ terbesar}{varians\ terkecil}}$$

Berikut ini adalah langkah-langkah yang digunakan untuk menguji homogenitas menggunakan uji F:

1) Merumuskan H_0
2) Merumuskan H_1
3) Menentukan taraf signifikan ($\boldsymbol{\alpha}$)
4) Menentukan kriteria penolakan H_0, yakni tolak H_0 jika $F_{hitung} > F_{tabel}$, dimana

   $$\boldsymbol{F_{tabel} = F_{\frac{1}{2}\alpha(dk\ varians\ terbesar-1,\ \ dk\ varians\ terkecil-1)}}$$

5) Menghitung nilai varians kelompok ke-i menggunakan rumus:

$$s^2{}_i = \frac{n_i \sum X_i{}^2 - (\sum X_i)^2}{n_i(n_i - 1)}$$

6) Menentukan nilai F_{hitung} dengan menggunakan rumus: $\boldsymbol{F = \frac{varians\ terbesar}{varians\ terkecil}}$
7) Membandingkan nilai F_{hitung} dengan F_{tabel} dan selanjutnya membuat simpulan

Contoh:
Sebuah penelitian ingin membandingkan kinerja karyawan di dua kantor cabang Perusahaan, yaitu kantor cabang A dan kantor cabang B. Adapun data yang diperoleh sebagai berikut:

No.	Kantor A	Kantor B
1	78	54
2	52	73
3	53	56
4	86	62
5	69	77
6	58	67
7	48	61
8	47	55
9	64	75
10	52	61
11	63	57
12	73	82
13	70	52
14	62	48
15	57	61
16	-	72
17	-	80

Dengan taraf signifikan $\boldsymbol{\alpha = 0,05}$, apakah kedua kelompok data memiliki varians yang homogen?

Jawab

Untuk menjawab soal di atas, maka digunakan langkah-langkah pengujian homogenitas menggunakan uji F berikut:

1) H_0 : Tidak ada perbedaan varians A dan varians B (kedua varians homogen)
   H_0 : $\sigma^2{}_A = \sigma^2{}_B$
2) H_1 : Ada perbedaan varians A dan varians B (kedua varians tidak homogen)
   H_1 : $\sigma^2{}_A \neq \sigma^2{}_B$
3) Taraf signifikan $\alpha = 0{,}05$
4) Kriteria penolakan H_0: tolak H_0 jika $F_{hitung} > F_{tabel}$.
5) Menentukan nilai varians masing-masing kelompok.
   Tabel bantu untuk menghitung varians.

No.	X_A	$X_A{}^2$	X_B	$X_B{}^2$
1	78	6084	54	2916
2	52	2704	73	5329
3	53	2809	56	3136
4	86	7396	62	3844
5	69	4761	77	5929
6	58	3364	67	4489
7	48	2304	61	3721
8	47	2209	55	3025
9	64	4096	75	5625
10	52	2704	61	3721
11	63	3969	57	3249
12	73	5329	82	6724
13	70	4900	52	2704
14	62	3844	48	2304
15	57	3249	61	3721
16	-	-	72	5184
17	-	-	80	6400
$\sum$	932	59722	1093	72021

Menghitung varians kelompok A:

$$s^2{}_A = \frac{n_A \sum {X_A}^2 - (\sum X_A)^2}{n_A(n_A - 1)}$$
$$= \frac{15 \times 59.722 - (932)^2}{15 \times (15 - 1)}$$
$$= \frac{895.830 - 868.624}{210}$$
$$= \frac{27.206}{210}$$
$$= 129{,}552$$

Menghitung varians kelompok B:

$$s^2{}_B = \frac{n_B \sum {X_B}^2 - (\sum X_B)^2}{n_B(n_B - 1)}$$
$$= \frac{17 \times 72.021 - (1.093)^2}{17 \times (17 - 1)}$$
$$= \frac{1.224.357 - 1.194.649}{272}$$
$$= \frac{29.708}{272}$$
$$= 109{,}221$$

6) Menentukan nilai F_{hitung}:
Untuk menentukan nilai F_{hitung}, terlebih dahulu perlu dilihat kelompok mana yang memiliki varians terbesar dan varians terkecil. Berdasarkan perhitungan varians pada langkah sebelumnya, ternyata kelompok A memiliki varians terbesar (129,552) dan kelompok B memiliki varians terkecil (109,221). Sehingga nilai F_{hitung} menjadi:
$$F = \frac{varians\ terbesar}{varians\ terkecil} = \frac{129{,}552}{109{,}221} = 1{,}186$$
7) Setelah melihat di tabel F, ternyata nilai $\boldsymbol{F_{tabel} = F_{0{,}025(14,16)} = 2{,}373}$. Jika dibandingkan dengan F_{hitung} = 1,186, ternyata 1,186 ≤ 2,373, yang artinya $F_{hitung} \leq F_{tabel}$, sehingga H_0 diterima.
Jadi, dapat disimpulkan bahwa tidak ada perbedaan varians A dan varians B, atau dengan kata lain, kedua varians homogen.

B. UJI BARTLETT

Rumus yang digunakan untuk menguji homogenitas varians menggunakan uji Bartlett adalah

$$\boldsymbol{\chi^2 = \ln 10 \left(B - \sum(n_i - 1) \log s^2{}_i\right)}$$

dimana $\boldsymbol{B = \log s^2{}_i \sum(n_i - 1)}$ dan

$$\boldsymbol{s^2 = \frac{\sum(n_i - 1)s^2{}_i}{\sum(n_i - 1)}}$$

Keterangan:

$s^2{}_i$: varians kelompok ke-i

n_i : banyak anggota kelompok ke-i

s^2 : varians gabungan

Adapun langkah-langkah yang digunakan untuk menguji homogenitas menggunakan uji Bartlett adalah sebagai berikut:

1. Merumuskan H_0
2. Merumuskan H_1
3. Menentukan taraf signifikan (α)
4. Menentukan kriteria penolakan H_0, yakni tolak H_0 jika $\chi^2{}_{hitung} \geq \chi^2{}_{tabel}$ dimana
   $\chi^2{}_{tabel} = \chi^2{}_{(1-\alpha)(dk)}$
5. Menghitung nilai varians kelompok ke-i menggunakan rumus:

$$s^2{}_i = \frac{n_i \sum X_i{}^2 - (\sum X_i)^2}{n_i(n_i - 1)}$$

6. Membuat tabel bantu untuk uji Bartlett:

Kelompok ke-	$\boldsymbol{dk}$	$\boldsymbol{s^2{}_i}$	$\boldsymbol{\log s^2{}_i}$	$\boldsymbol{(dk) \log s^2{}_i}$
1	$n_1 - 1$	$s^2{}_1$	$\log s^2{}_1$	$(n_1 - 1) \log s^2{}_1$
2	$n_2 - 1$	$s^2{}_2$	$\log s^2{}_2$	$(n_2 - 1) \log s^2{}_2$
3	$n_3 - 1$	$s^2{}_3$	$\log s^2{}_3$	$(n_3 - 1) \log s^2{}_3$
k	$n_k - 1$	$s^2{}_k$	$\log s^2{}_k$	$(n_k - 1) \log s^2{}_k$
$\sum$	$\boldsymbol{\sum(n_i - 1)}$	-	-	$\boldsymbol{\sum(n_i - 1) \log s^2{}_i}$

7. Menghitung varians gabungan dengan menggunakan rumus:
   $s^2 = \frac{\sum(n_i - 1)s^2{}_i}{\sum(n_i - 1)}$
8. Menghitung nilai B dengan menggunakan rumus:
   $B = \log s^2 \sum(n_i - 1)$
9. Menghitung nilai $\chi^2{}_{hitung}$ dengan menggunakan rumus:
   $\chi^2 = \ln 10 \, (B - \sum(n_i - 1) \log s^2{}_i)$
10. Membandingkan nilai $\chi^2{}_{hitung}$ dengan $\chi^2{}_{tabel}$ dan selanjutnya membuat simpulan.

Contoh:
Seorang peneliti ingin membandingkan hasil belajar siswa berdasarkan gaya belajar *Kolb*, yaitu gaya belajar konvergen (X_1), divergen (X_2), asimilatif (X_3), dan akomodatif (X_4). Untuk keperluan penelitian tersebut, peneliti mengumpulkan data berupa hasil belajar dari kelompok gaya belajar masing-masing dan ditampilkan seperti tabel berikut.

No.	X_1	X_2	X_3	X_4
1	76	53	56	68
2	82	64	74	77
3	59	62	72	72
4	73	75	88	75
5	61	86	84	83
6	64	82	79	85
7	57	69	58	89
8	71	72	65	63
9	74	73	75	58
10	88	69	57	72
11	72	57	62	67
12	69	72	73	73
13	72	74	89	54
14	71	67	82	-
15	80	-	74	-

Dengan menggunakan taraf signifikan $\alpha = 0{,}05$, apakah data dari keempat kelompok tersebut memiliki varians yang homogen?

Jawab:

Berdasarkan soal tersebut, data yang ingin diuji homogenitasnya terdiri dari 4 (empat) kelompok data. Jika banyak kelompok data yang ingin diuji homogenitasnya lebih dari 2 (dua) maka digunakan uji Bartlett.

Langkah-langkah pengujian homogenitas menggunakan *Uji Bartlett*:

1) H_0 : Tidak ada perbedaan varians (varians homogen)
   H_0 : $\sigma^2{}_1 = \sigma^2{}_2 = \sigma^2{}_3 = \sigma^2{}_4$
2) H_1 : Ada perbedaan varians (varians tidak homogen)
   H_1 : paling sedikit ada satu tanda "=" tidak berlaku (ada tanda yang ≠)
3) Taraf signifikan $\alpha = 0{,}05$
4) Tolak H_0 jika $\chi^2{}_{hitung} \geq \chi^2{}_{tabel}$
5) Menghitung nilai varians kelompok ke-i menggunakan rumus:

$$s^2{}_i = \frac{n_i \sum X_i{}^2 - (\sum X_i)^2}{n_i(n_i - 1)}$$

Tabel bantu untuk menghitung varians

No.	X_1	$X_1{}^2$	X_2	$X_2{}^2$	X_3	$X_3{}^2$	X_4	$X_4{}^2$
1	76	5776	53	2809	56	3136	68	4624
2	82	6724	64	4096	74	5476	77	5929
3	59	3481	62	3844	72	5184	72	5184
4	73	5329	75	5625	88	7744	75	5625
5	61	3721	86	7396	84	7056	83	6889
6	64	4096	82	6724	79	6241	85	7225
7	57	3249	69	4761	58	3364	89	7921
8	71	5041	72	5184	65	4225	63	3969
9	74	5476	73	5329	75	5625	58	3364
10	88	7744	69	4761	57	3249	72	5184
11	72	5184	57	3249	62	3844	67	4489
12	69	4761	72	5184	73	5329	73	5329
13	72	5184	74	5476	89	7921	54	2916

No.	X_1	X_1^2	X_2	X_2^2	X_3	X_3^2	X_4	X_4^2
14	71	5041	67	4489	82	6724	-	-
15	80	6400	-	-	74	5476	-	-
∑	1069	77207	975	68927	1088	80594	936	68648

- Menghitung varians kelompok 1:

$$s^2{}_1 = \frac{n_1 \sum {X_1}^2 - (\sum X_1)^2}{n_1(n_1 - 1)}$$
$$= \frac{15 \times 77.207 - (1.069)^2}{15 \times (15 - 1)}$$
$$= \frac{1.158.105 - 1.142.761}{210}$$
$$= \frac{15.344}{210}$$
$$= 73{,}067$$

- Menghitung varians kelompok 2:

$$s^2{}_2 = \frac{n_2 \sum {X_2}^2 - (\sum X_2)^2}{n_2(n_2 - 1)}$$
$$= \frac{14 \times 68.927 - (975)^2}{14 \times (14 - 1)}$$
$$= \frac{964.978 - 950.625}{182}$$
$$= \frac{14.353}{182}$$
$$= 78{,}863$$

- Menghitung varians kelompok 3:

$$s^2{}_3 = \frac{n_3 \sum {X_3}^2 - (\sum X_3)^2}{n_3(n_3 - 1)}$$
$$= \frac{15 \times 80.954 - (1.088)^2}{15 \times (15 - 1)}$$
$$= \frac{1.208.910 - 1.183.744}{210}$$
$$= \frac{25.166}{210}$$
$$= 119{,}838$$

- Menghitung varians kelompok 4:

$$s^2{}_4 = \frac{n_4 \sum {X_4}^2 - (\sum X_4)^2}{n_4(n_4 - 1)}$$
$$= \frac{13 \times 68.648 - (936)^2}{13 \times (13 - 1)}$$
$$= \frac{892.424 - 876.096}{156}$$
$$= \frac{16.328}{156}$$
$$= 104{,}667$$

6) Membuat tabel bantu untuk Uji Bartlett:
Tabel bantu untuk Uji Bartlett

Kelompok ke-	$\boldsymbol{dk}$	$\boldsymbol{s^2{}_i}$	$\boldsymbol{\log s^2{}_i}$	$\boldsymbol{(dk) \log s^2{}_i}$
1	14	73,067	1,8637	26,0918
2	13	78,863	1,8969	24,6597
3	14	119,838	2,0786	29,1004
4	12	104,667	2,0198	24,2376
$\sum$	**53**	-	-	$\mathbf{104,0895}$

7) Menghitung varians gabungan:

$$s^2 = \frac{\sum (n_i - 1) {s^2}_i}{\sum (n_i - 1)}$$

$$= \frac{14 \times 73{,}067 + 13 \times 78{,}863 + 14 \times 119{,}838 + 12 \times 104{,}667}{14 + 13 + 14 + 12}$$
$$= \frac{1.022{,}938 + 1.025{,}219 + 1.677{,}732 + 1.256{,}004}{53}$$
$$= \frac{4.981{,}893}{53}$$
$$= 93{,}998$$

8) Menghitung nilai B:
$B = (\log s^2) \sum(n_i - 1) = \log(93{,}998) \times 53 = 104{,}5753$
9) Menghitung nilai $\chi^2{}_{hitung}$:
$\chi^2 = \ln 10 \, (B - \sum(n_i - 1) \log s^2{}_i) = \ln 10 \times (104{,}5753 - 104{,}0895) = 1{,}1186$
10) Setelah melihat di tabel Chi-kuadrat (χ^2), ternyata nilai $\chi^2{}_{tabel} = \chi^2{}_{0{,}95(3)} = 7{,}815$. Jika dibandingkan dengan $\chi^2{}_{\text{hitung}}$ = 1,1186, ternyata 1,1186 < 7,815, yang artinya $\chi^2{}_{\text{hitung}} < \chi^2{}_{\text{tabel}}$, sehingga H_0 diterima.

Jadi, dapat disimpulkan bahwa tidak ada perbedaan varians, atau dengan kata lain, keempat kelompok tersebut memiliki varians yang homogen.

C. UJI LEVENE

Saat ini banyak bermunculan aplikasi yang membantu memudahkan dalam analisis data statistik. Salah satu aplikasi yang paling sering digunakan adalah SPSS. Pengujian homogenitas dengan SPSS dapat menggunakan uji Levene. Adapun langkah-langkah pengujiannya sebagai berikut.

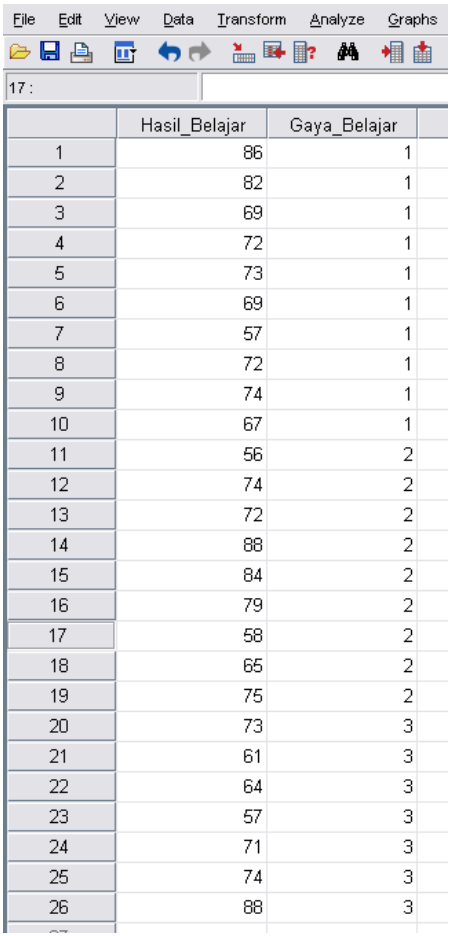
1. Buka aplikasi SPSS kemudian pilih/buka file data yang akan dianalisis.
2. Pilih menu *Analyze → Compare Means → One Way Anova*
3. Masukkan variabel yang diujikan ke kolom *Dependent List*
4. Masukkan variabel yang membedakan ke kolom *Factor*
5. Klik tombol *Option → Homogeneity of variance test → Continue*
6. Selanjutnya klik *OK*

Untuk menafsirkan hasil pengujian homogenitas menggunakan SPSS, digunakan kriteria berikut:

1. Nilai Sig. (p) ≥ α maka data memiliki varians yang homogen.

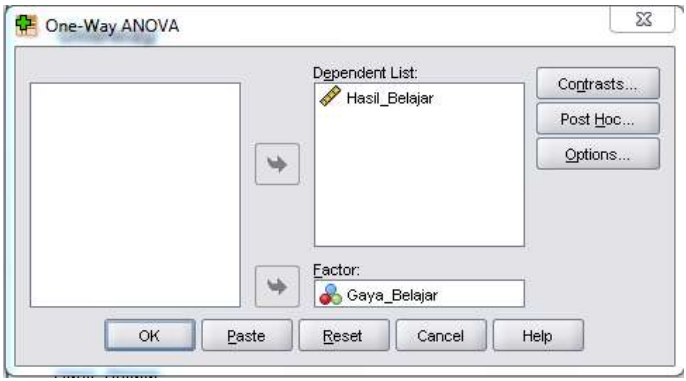
2. Nilai Sig. (p) < α maka data memiliki varians yang tidak homogen.

Contoh:
Diberikan data hasil belajar 26 siswa dengan gaya belajar yang berbeda. Data gaya belajar diberikan label "1 = gaya belajar visual", "2 = gaya belajar auditori", dan "3 = gaya belajar kinestetik". Untuk lebih jelas, data disajikan pada gambar berikut.

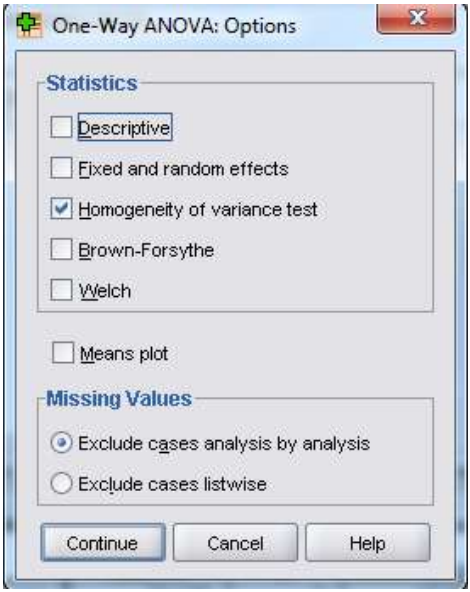


	Hasil_Belajar	Gaya_Belajar
1	86	1
2	82	1
3	69	1
4	72	1
5	73	1
6	69	1
7	57	1
8	72	1
9	74	1
10	67	1
11	56	2
12	74	2
13	72	2
14	88	2
15	84	2
16	79	2
17	58	2
18	65	2
19	75	2
20	73	3
21	61	3
22	64	3
23	57	3
24	71	3
25	74	3
26	88	3

Untuk menguji homogenitas data tersebut, pilih menu *Analyze → Compare Means → One Way Anova*. Selanjutnya, masukkan variabel Hasil_Belajar ke kolom *Dependent list*, dan variabel Gaya_Belajar ke kolom *Factor*.



Kemudian klik *Option → Homogeneity of variance test → Continue.*



Langkah terakhir adalah klik *OK.*
Setelah itu akan muncul hasil pengujian homogenitas seperti gambar berikut ini.

Test of Homogeneity of Variances

Hasil_Belajar

Levene Statistic	df1	df2	Sig.
.750	2	23	.483

Berdasarkan output di atas, nilai Sig. (p) = 0,483. Jika tingkat signifikan (α) yang digunakan adalah 0,05, maka nilai $\boldsymbol{p \geq \alpha}$, yang artinya data tersebut memiliki varians yang homogen.

DAFTAR PUSTAKA

Ananda, R., & Fadhli, M. 2018. *Statistik Pendidikan (Teori dan Praktik dalam Pendidikan)*. Medan: CV. Widya Puspita.

Bustami, Abdullah, D., & Fadlisyah. 2014. *Statistika: Terapannya pada Bidang Informatika*. Yogyakarta: Graha Ilmu.

Nuryadi, dkk. 2017. *Dasar-dasar Statistik Penelitian.* Yogyakarta: Sibuku Media.

Sudjana. 2002. *Metoda Statistika Edisi ke-6*. Bandung: Tarsito.

Sumanto. 2014. Statistika Terapan. Yogyakarta: CAPS (Center of Academic Publishing Service).

Usman, H., & Akbar, R.P.S. 2011. *Pengantar Statistika Edisi ke-2*. Jakarta: Bumi Aksara.

PROFIL PENULIS

Tri Widyasari, dilahirkan di Samarinda pada tanggal 6 April 1989. Pada tahun 2009 memperoleh gelar Sarjana Pendidikan (S.Pd) program studi Pendidikan Matematika Universitas Mulawarman. Kemudian tahun 2011 melanjutkan pendidikan S2 pada Universitas Negeri Surabaya program studi Pendidikan Matematika dan memperoleh gelar Magister Pendidikan (M.Pd) di tahun 2013. Sejak 2015 menjadi pengajar di IKIP PGRI Kaltim program studi Pendidikan Ekonomi sampai sekarang. Mata kuliah yang diampu adalah Matematika Ekonomi dan Statistika.

BAB 11
UJI RATA-RATA DAN PROPORSI

Risy Mawardati, M.Pd
Universitas Iskandar Muda
Banda Aceh

A. PENGUJIAN HIPOTESIS

Menurut Sudjana (2002:219), hipotesis adalah asumsi atau dugaan mengenai sesuatu hal yang dibuat untuk menjelaskan hal itu benar atau salah untuk pengambilan keputusan. Setiap hipotesis bisa benar atau salah,oleh karena itu perlu diadakan peneltian untuk memutuskan hipotesis itu diterima atau ditolak. Langkah atau prosedur untuk menentukan apakah menerima atau menolak hipotesis dinamakan dengan pengujian hipotesis.

Dalam statistika dikenal dua jenis hipotesis, yaitu: hipotesis nol dan hipotesis alternatif. Hipotesis nol (H_0), berupa suatu pernyataan tidak adanya perbedaan karakteristik atau parameter populasi. Perumusan H_0 selalu ditandai oleh tanda sama dengan (=).

Hipotesi alternatif (H_1), berupa suatu pernyataan yang bertentangan dengan H_0. Perumusan H_1 ditandai oleh tanda ($<$, $>$, dan $\neq$).

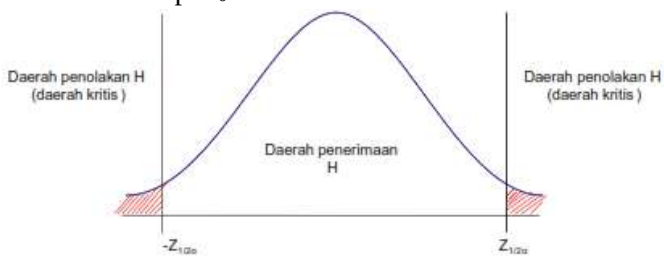
Contoh:

H_0 : $\mu = 65$ (Rata-rata nilai statistika mahasiswa semester IV adalah sama dengan 65)

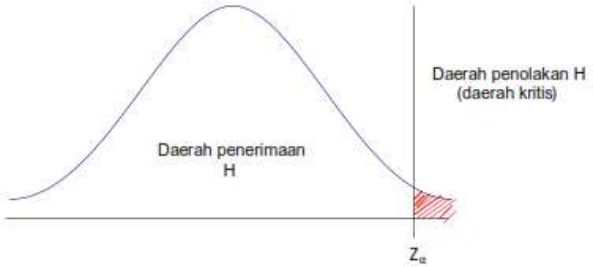
H_1 : $\mu \neq 65$ (Rata-rata nilai statistika mahasiswa semester IV adalah tidak sama dengan 65)

Adapun langkah-langkah pengujian hipotesis adalah sebagai berikut:

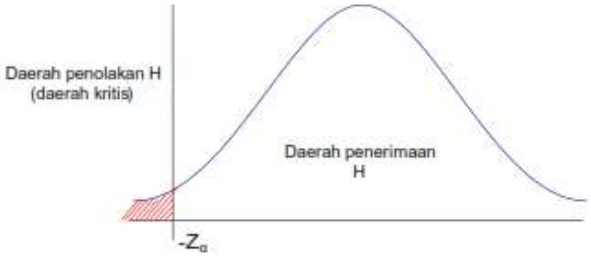
1. Rumuskan H_0 dan H_1 yang sesuai
2. Pilih taraf nyata (α) yang akan digunakan. Umumnya 1% atau 5%.
3. Pilih uji statistik yang akan digunakan
4. Tentukan arah pengujian (satu pihak atau dua pihak) dan daerah kritis atau penolakan terhadap H_0



Gambar 1. Uji dua pihak



Gambar 2. Uji satu pihak, pihak kanan



Gambar 3. Uji satu pihak, pihak kiri

5. Hitung statistik uji
6. Bandingkan hasil statistik uji dengan daerah kritis
7. Buat kesimpulan, berupa terima H_0 atau tolak H_0.

B. UJI RATA-RATA

Pada bagian ini kita akan membahas uji rata-rata (μ) dengan mengumpamakan kita mempunyai sebuah populasi yang berdistribusi normal dengan rata-rata (μ) dan simpangan baku (σ), yang akan diuji adalah mengenai parameter rata-rata (μ).

1. Uji Rata-Rata dengan (σ) diketahui

   Adapun langkah-langkah pengujian hipotesis pada uji rata-rata dengan (σ) diketahui adalah sebagai berikut:

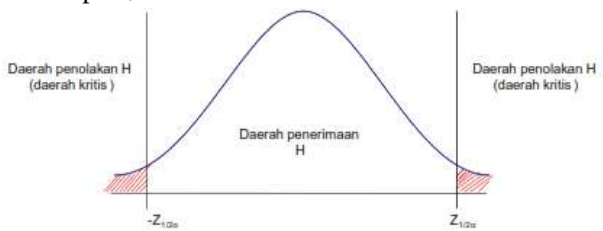
   a. Rumuskan H_0 dan H_1 yang sesuai, yaitu:

      1) $H_0 : \mu = \mu_0$

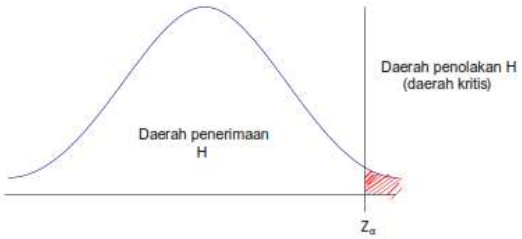
         $H_1 : \mu \neq \mu_0$

2) $H_0 : \mu = \mu_0$
   $H_1 : \mu > \mu_0$
3) $H_0 : \mu = \mu_0$
   $H_1 : \mu < \mu_0$

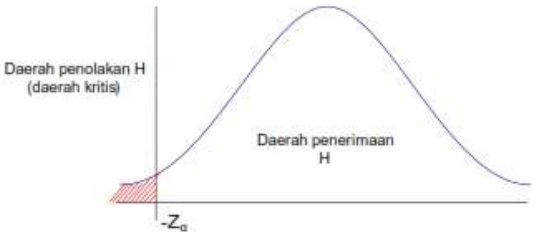
b. Pilih taraf nyata (α) yang akan digunakan. Umumnya 1% atau 5%.
c. Pilih uji statistik yang akan digunakan
d. Tentukan arah pengujian (satu pihak atau dua pihak) dan daerah kritis atau penolakan terhadap H_0



Gambar 1. Uji dua pihak



Gambar 2. Uji satu pihak, pihak kanan



Gambar 3. Uji satu pihak, pihak kiri

e. Hitung statistik uji dengan rumus $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$

f. Bandingkan hasil statistik uji dengan daerah kritis

g. Buat kesimpulan, berupa terima H_0 atau tolak H_0.
Dengan kriteria sebagai berikut:

1) $H_0 : \mu = \mu_0$ dan $H_1 : \mu \neq \mu_0$:
   a). H_0 diterima jika $-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}$,
   b). H_0 ditolak jika $Z > Z_{\alpha/2}$ atau $Z < -Z_{\alpha/2}$.
2) $H_0 : \mu = \mu_0$ dan $H_1 : \mu > \mu_0$:
   a). H_0 diterima jika $Z \leq Z_{\alpha}$
   b). H_0 ditolak jika $Z > Z_{\alpha}$
3) $H_0 : \mu = \mu_0$ dan $H_1 : \mu < \mu_0$:
   a). H_0 diterima jika $Z \geq Z_{\alpha}$
   b). H_0 ditolak jika $Z < -Z_{\alpha}$

Contoh soal:
Seorang peneliti menduga bahwa rata-rata nilai statistika mahasiswa semester IV adalah 65. Untuk menyelidiki dugaan ini, peneliti mengambil sampel 36 mahasiswa semester IV dan didapat rata-rata nilai mereka adalah 64 dengan simpangan baku 12. Tentukan apakah hipotesis peeliti tersebut dapat diterima atau ditolak pada taraf signifikan 0,05!

Penyelesaian:

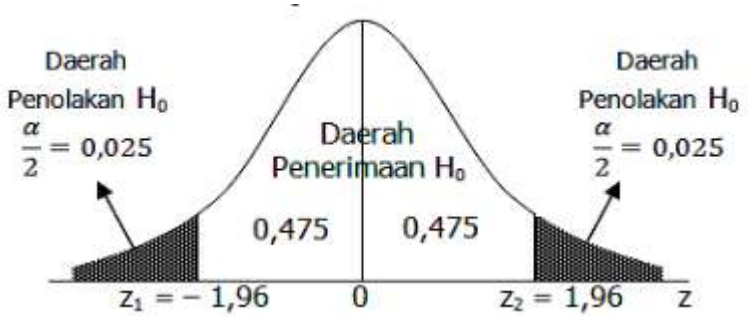
1) Rumusan H_0 dan H_1 yang sesuai
   Hipotesis: $H_0 : \mu = 65$
   $H_1 : \mu \neq 65$
2) Taraf nyata (α) yang akan digunakan 0,05 atau 5%.
3) Uji statistik yang akan digunakan adalah uji Z karena sampel besar.
4) Arah pengujian dua pihak
5) Menghitung statistik uji
   Diketahui $\mu_0 = 65$; $n = 36$; $\bar{x} = 64$; $\alpha = 0{,}05$; $\sigma = 12$.
   $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$

$$z = \frac{64 - 65}{12/\sqrt{36}}$$

z = - 0,5

6) Hasil statistik uji dengan daerah kritis

Dari daftar distribusi normal untuk uji dua pihak dengan α = 0,05 diperoleh nilai $z_{0,475}$ = 1,96. Uji dua pihak



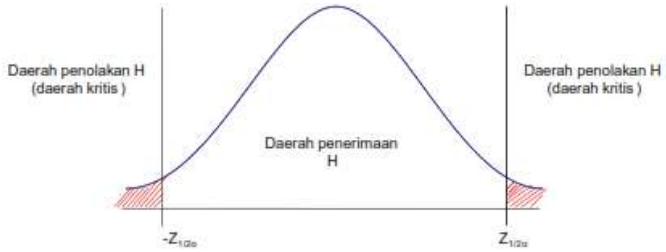
7) Kesimpulan:

Jika z hitung terletak antara -1,96 dan 1,96 maka H_0 diterima, jika tidak maka H_0 ditolak. Dari hasil z hitung diperoleh z = - 0,5 yang artinya berada pada interval -1,96 dan 1,96. Jadi dapat disimpulkan bahwa H_0 diterima. Artinya rata-rata nilai statistika mahasiswa semester IV adalah 65.

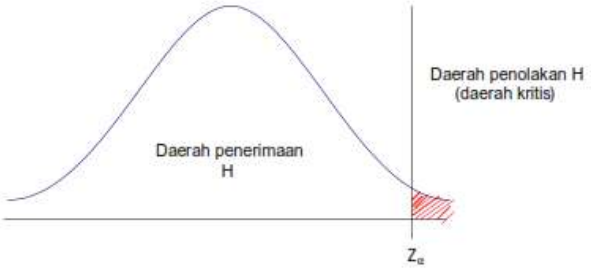
2. Uji Rata-Rata dengan (σ) tidak diketahui

Dalam aplikasi sehari-hari, seringkali (σ) tidak diketahui sehingga nilai (σ) akan diestimasi dengan (s). Dengan demikian pengujian hipotesis tidak lagi menggunakan uji z melainkan uji t. Adapun langkah-langkah pengujian hipotesis pada uji rata-rata dengan (σ) tidak diketahui adalah sebagai berikut:

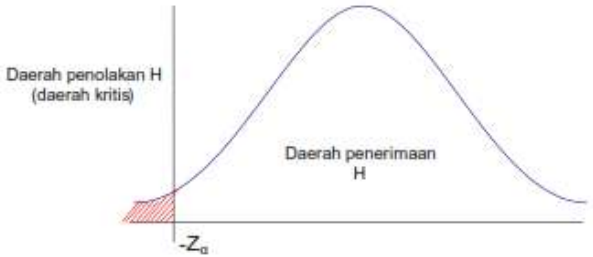
a. Rumuskan H_0 dan H_1 yang sesuai
b. Pilih taraf nyata (α) yang akan digunakan. Umumnya 1% atau 5%.
c. Pilih uji statistik yang akan digunakan
d. Tentukan arah pengujian (satu pihak atau dua pihak) dan daerah kritis atau penolakan terhadap H_0



Gambar 1. Uji dua pihak



Gambar 2. Uji satu pihak, pihak kanan



Gambar 3. Uji satu pihak, pihak kiri

e. Hitung statistik uji dengan rumus $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$
f. Bandingkan hasil statistik uji dengan daerah kritis
g. Buat kesimpulan, berupa terima H_0 atau tolak H_0.

Contoh soal:

Seorang peneliti ingin melakukan penelitian mengenai tinggi badan mahasiswa yang mengikuti mata kuliah Statistika. Untuk itu dilakukan penelitian terhadap sepuluh mahasiswa yang mengikuti mata kuliah tersebut dan diperoleh hasil sebagai berikut:

Mahasiswa ke-	1	2	3	4	5	6	7	8	9	10
Tinggi Badan (cm)	150	160	160	150	185	156	171	165	171	166

Ujilah hipotesis:

a. Apakah tinggi badan mahasiswa tersebut adalah 155 cm?
b. Apakah tinggi badan mahasiswa tersebut adalah diatas 155 cm?
c. Apakah tinggi badan mahasiswa tersebut adalah dibawah 155 cm?

Penyelesaian:

a. 1. Rumusan H_0 dan H_1 yang sesuai

Hipotesis: $H_0 : \mu = 155$

$H_1: \mu \neq 155$

2. Taraf nyata (α) yang akan digunakan 0,05 atau 5%.
3. Uji statistik yang akan digunakan adalah uji t
4. Arah pengujian dua pihak
5. Menghitung statistik uji

Diketahui $\mu_0 = 155$; n = 10; $\bar{x} = 163{,}40$; $\alpha = 0{,}05$; $s = 10{,}69$

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

$$t = \frac{163{,}40 - 155}{10{,}69/\sqrt{10}}$$

t = 2,48

6. Hasil statistik uji dengan daerah kritis

Dari daftar distribusi normal untuk uji dua pihak dengan $\alpha = 0{,}05$ diperoleh nilai $t_{0{,}025(9)} = 2{,}262$.

7. Kesimpulan:

Jika t hitung tidak terletak antara -2,26 dan 2,26 maka H_0 ditolak dan H_1 diterima. Artinya rata-rata tinggi badan mahasiswa tersebut bukan 155 cm.

b. 1. Rumusan H_0 dan H_1 yang sesuai

Hipotesis: $H_0 : \mu = 155$

$H_1: \mu > 155$

2. Taraf nyata (α) yang akan digunakan 0,05 atau 5%.
3. Uji statistik yang akan digunakan adalah uji t
4. Arah pengujian satu pihak yaitu pihak kanan
5. Menghitung statistik uji

Diketahui $\mu_0 = 155$; n = 10; $\bar{x} = 163{,}40$; α = 0,05; $s = 10{,}69$

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

$$t = \frac{163{,}40 - 155}{10{,}69/\sqrt{10}}$$

t = 2,48

6. Hasil statistik uji dengan daerah kritis

Dari daftar distribusi normal untuk uji dua pihak dengan α = 0,05 diperoleh nilai $z_{0,05(9)} = 1{,}833$.

7. Kesimpulan:

Jika t hitung lebih besar dari t tabel maka H_0 ditolak dan H_1 diterima. Artinya rata-rata tinggi badan mahasiswa tersebut bukan 155 cm.

c. 1. Rumusan H_0 dan H_1 yang sesuai

Hipotesis: $H_0 : \mu = 155$

$H_1: \mu < 155$

2. Taraf nyata (α) yang akan digunakan 0,05 atau 5%.
3. Uji statistik yang akan digunakan adalah uji t
4. Arah pengujian dua pihak
5. Menghitung statistik uji

Diketahui $\mu_0 = 155$; n = 10; $\bar{x} = 163{,}40$; α = 0,05; $s = 10{,}69$

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

$$t = \frac{163{,}40 - 155}{10{,}69/\sqrt{10}}$$

t = 2,48

6. Hasil statistik uji dengan daerah kritis

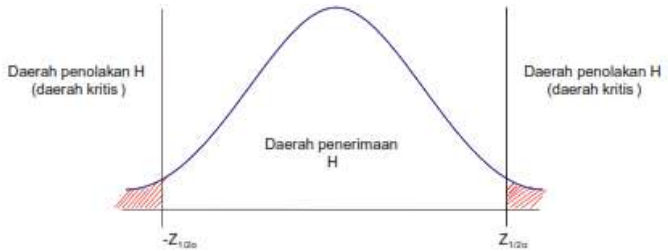
Dari daftar distribusi normal untuk uji dua pihak dengan α = 0,05 diperoleh nilai $t_{0,05(9)} = -1.833$.

7. Kesimpulan:
   Jika t hitung lebih dari t tabel maka H_0 diterima dan H_1 ditolak. Artinya rata-rata tinggi badan mahasiswa tersebut adalah 155 cm.

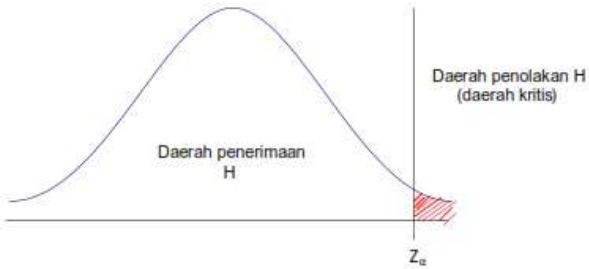
C. UJI PROPORSI

Pada bagian ini kita akan membahas mengenai uji proporsi (P), dimana pengujian hipotesis proporsi berhubungan erat dengan proporsi atau persentase kejadian sukses dalam suatu percobaan binomial (Sutarto,2021). Pada pengujian hipotesis proporsi, langkah pengujiannya sama dengan hipotesis umumnya hanya yang membedakan statistik pengujiannya adalah variabel random binomial X dengan P merupakan parameter distribusi binomial. Secara ringkas langkah pengujian proporsi adalah sebagai berikut:

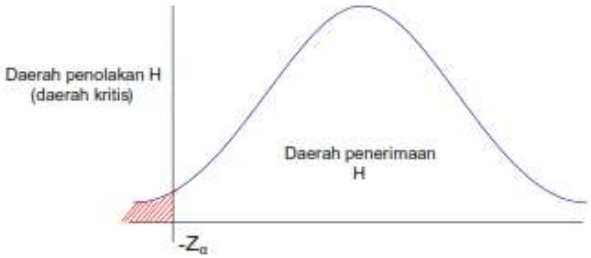
1. Rumuskan H_0 dan H_1 yang sesuai
2. Pilih taraf nyata (α) yang akan digunakan. Umumnya 1% atau 5%.
3. Pilih uji statistik yang akan digunakan
4. Tentukan arah pengujian (satu pihak atau dua pihak) dan daerah kritis atau penolakan terhadap H_0



Gambar 1. Uji dua pihak



Gambar 2. Uji satu pihak, pihak kanan



Gambar 3. Uji satu pihak, pihak kiri

5. Hitung statistik uji dengan rumus $z = \frac{X - nP_o}{\sqrt{nP_o(1-P_o)}}$ atau $z = \frac{\frac{X}{n} - P_o}{\sqrt{\frac{P_o(1-P_o)}{n}}}$, dengan n menyatakan banyaknya ukuran sampel dan X menyatakan banyaknya ukuran sampel dengan karakteristik tertentu.
6. Bandingkan hasil statistik uji dengan daerah kritis
7. Buat kesimpulan, berupa terima H_0 atau tolak H_0.

Contoh soal:
Seorang peneliti mengatakan bahwa paling banyak 60% orang tua di kampus A berprofesi sebagai PNS. Sebuah sampel acak sudah diambil yang terdiri dari 8.500 orang dan ternyata 5.426 berprofesi sebagai PNS. Apakah anda setuju dengan pernyataan tersebut? Gunakan taraf nyata 0,01.
Penyelesaian:
1. Rumusan H_0 dan H_1 yang sesuai
   Hipotesis: $H_0 : P = 0,6$
   $H_1: P > 0,6$
2. Taraf nyata (α) yang akan digunakan 0,01.
3. Uji statistik yang akan digunakan adalah uji z
4. Arah pengujian satu pihak yaitu pihak kanan
5. Menghitung statistik uji
   Diketahui $P_0 = 6\% = 0,6$; n = 8.500; X = 5.426,40; α = 0,01.
   $$z = \frac{X - nP_o}{\sqrt{nP_o(1-P_o)}}$$

$$z = \frac{5.426 - 8.500(0,6)}{\sqrt{8.500(0,6)(1-0,6)}}$$

z = 7,22

6. Hasil statistik uji dengan daerah kritis
   Dari daftar distribusi normal untuk uji pihak kanan dengan α = 0,01 diperoleh nilai $z_{0,49}$ = 2,33.
7. Kesimpulan:
   Jika z hitung lebih besar dari z tabel maka H_0 ditolak dan H_1 diterima. Artinya persentase orang tua di kampus A yang berprofesi sebagai PNS sudah mencapai 60%.

DAFTAR PUSTAKA

Hartati, Neneng. 2017. *Satatistika untuk Analisis Data Penelitian*. Bandung: CV.Putaka Setia.

Hasan, Iqbal. 2008. *Pokok-Pokok Materi Statistik 2 (Statistik Inferensif).* Jakarta: Bumi Aksara.

Pusponegoro, N.H.2013. *Pengujian Hipotesis*. Jakarta: Badan Pusat Statistik.

Sudjana. 2002. *Metoda Statistika.* Bandung: Tarsito.

Sutarto,A.P. 2021. *Probabilitas dan Statistika Dasar untuk Sains*. Yogyakarta: PT. Pustaka Baru.

Walpole, R.E. 1995. *Pengantar Statistika*. Jakarta: PT. Gramedia Pustaka Utama.

PROFIL PENULIS

Risy Mawardati lahir di Indrapuri pada 12 Agustus 1988. Sulung dari tiga bersaudara ini merupakan putri dari pasangan bapak Ridwan dan ibu Syamsiah. Lulusan magister pendidikan matematika ini sekarang aktif mengajar di Universitas Iskandar Muda. Yang membuatnya semangat memulai tulisan ini adalah sesuai dengan pesan Ali bin Abi Thalib yang mengatakan bahwa semua penulis akan mati, hanya karyanyalah yang akan abadi, maka tulislah sesuatu yang membahagiakan dirimu di akhirat nanti. Penulis merupakan pengelola e-jurnal DikMas: Jurnal Pendidikan Matematika dan Sains dengan alamat https://ejournal.unida-aceh.ac.id/index.php/dikmas/index. Pembaca bisa menghubungi penulis melalui email: risymawardati@gmail.com.

BAB 12
ANALISIS REGRESI LINEAR SEDERHANA

Dr. Lian G. Otaya, M.Pd
IAIN Sultan Amai Gorontalo

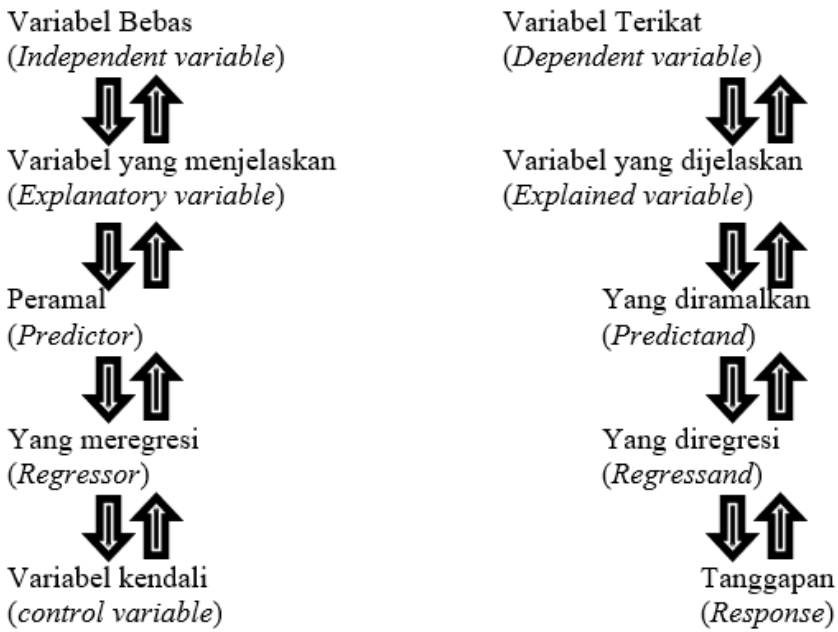
A. APA ITU REGRESI LINEAR SEDERHANA?

Istilah regresi pertama dipergunakan sebagai konsep statistika pada tahun 1877 oleh Sir Francis Galton yang melakukan studi tentang kecenderungan tinggi badan anak. Hasil studi tersebut merupakan suatu kesimpulan bahwa kecenderungan tinggi badan anak yang lahir terhadap orangtuanya adalah menurun (*regress*) mengarah pada tinggi badan rata-rata penduduk. Oleh karena itu, istilah regresi awalnya bertujuan untuk membuat perkiraan nilai satu variabel (tinggi badan anak) terhadap satu variabel yang lain (tinggi badan orangtua). Selanjutnya berkembang menjadi alat untuk membuat perkiraan nilai suatu variabel dengan variabel lain yang berhubungan dengan variabel tersebut. Sehingga dalam ilmu statistika, teknik yang umum digunakan untuk menganalisis hubungan antara dua atau lebih variabel adalah analisis regresi.

Analisis/uji regresi (*regression analysis*) merupakan suatu kajian dari hubungan antara satu variabel, yaitu *the explained variabel* sebagai variabel yang diterangkan dengan satu atau lebih variabel, yaitu *the explanatory* sebagai variabel yang menerangkan (Yuliara, 2016). Analisis regresi merupakan suatu teknik untuk membangun persamaan garis lurus dan menggunakan persamaan tersebut untuk membuat perkiraan (*prediction*). Analisis ini merupakan uji statistika yang memanfaatkan hubungan antara dua atau lebih peubah kuantitatif sehingga salah satu peubah bisa diramalkan dari peubah-peubah lainnya. Misalnya, jika ingin menguji dan membuktikan pengaruh motivasi belajar terhadap hasil belajar mahasiswa pada mata kuliah Statistika Pendidikan, maka dapat meramal hasil belajar mahasiswa tersebut melalui analisis regresi jika motivasi belajar mahasiswa telah diketahui. Konklusinya, analisis regresi dapat mengukur seberapa besar suatu variabel mempengaruhi variabel lain, dan dapat digunakan untuk melakukan peramalan nilai suatu variabel berdasarkan variabel lain, sehingga analisis ini dapat menyatakan hubungan sebab akibat

Analisis/uji regresi banyak digunakan dalam perhitungan hasil akhir untuk penulisan karya ilmiah/penelitian. Kegunaan analisis regresi dalam penelitian salah satunya adalah untuk meramalkan atau memprediksi variabel terikat apabila variabel bebas diketahui. Variabel bebas (*independent variable*) adalah variabel yang nilai-nilai tidak tergantung pada variabel lainnya, biasanya disimbolkan dengan ”X”. Variabel ini digunakan untuk

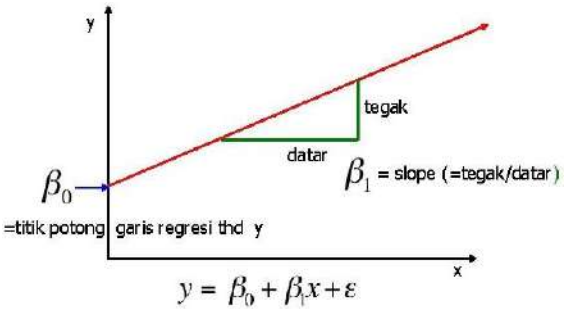
meramalkan atau menerangkan nilai variabel yang lain. Sementara variabel terikat (*dependent variable*) adalah variabel yang nilai-nilainya bergantung pada variabel lainnya, biasanya disimbolkan dengan ”Y”. Variabel ini merupakan variabel yang diramalkan atau diterangkan nilainya. Jika variabel bebas (variabel X) memiliki hubungan dengan variabel terikat (Variabel Y) maka nilai-nilai variabel X yang sudah diketahui dapat digunakan untuk menaksir atau memperkirakan nilai-nilai (Y). Berikut ini digambarkan berbagai istilah yang digunakan untuk variabel bebas dan variabel terikat.



Regresi sederhana dapat dianalisis karena didasari oleh hubungan fungsional atau hubungan sebab akibat (kausal) variabel bebas (X) terhadap variabel terikat (Y). Apabila variabel bebasnya hanya satu, maka analisis regresinya disebut dengan regresi sederhana. Uji regresi linier sederhana seperti uji signifikan dengan uji-t sangat membantu untuk mengetahui pengaruh secara kualitas dan kuantitas satu variabel bebas terhadap variabel tak bebas, maka perlu konsep dan teori yang mendasari kedua variabel tersebut (Riduwan & Sunarto, 2011). Data yang digunakan biasanya berskala interval atau rasio.

B. MODEL PERSAMAAN REGRESI LINEAR SEDERHANA

Persamaan regresi linier sederhana merupakan suatu model persamaan yang menggambarkan hubungan satu variabel bebas/ predictor (X) dengan satu variabel tak bebas/ response (Y), yang biasanya digambarkan dengan garis lurus, seperti disajikan pada Gambar 1.



Gambar 1. Ilustrasi Garis Regresi Linier

Ilustrasi garis regresi tersebut menunjukkan model regresi merupakan suatu cara formal untuk mengekspresikan dua unsur penting suatu hubungan statistik, yaitu: *Pertama*, suatu kecenderungan berubahnya peubah tidak bebas *Y* secara sistematis sejalan dengan berubahnya peubah besar *X*. Kedua, perpencaran titik-titik di sekitar kurva hubungan statistik itu. Kedua ciri ini disatukan dalam suatu model regresi dengan cara mempostulatkan bahwa :

1. Ada suatu rencana peluang peubah *Y* untuk setiap taraf (*level*) peubah *X*.
2. Rataan sebaran-sebaran peluang berubah secara sistematis sejalan dengan berubahnya nilai peubah *X*.

Menurut Gujarati (2006) model regresi dikatakan baik jika memenuhi beberapa kriteria yaitu:

1. Parsimoni. Artinya, suatu model tidak akan pernah dapat secara sempurna menangkap realitas, akibatnya kita akan melakukan sedikit abstraksi ataupun penyederhanaan dalam pembuatan model;
2. Mempunyai identifikasi tinggi. Arinya, dengan data yang ada parameter-parameter yang diestimasi harus mempunyai nilai-nilai yang unik atau dengan kata lain hanya akan ada satu parameter saja;
3. Keselarasan (*goodness of fit*). Tujuan analisis regresi menerangkan sebanyak mungkin variasi dalam variabel terikat dengan menggunakan variabel bebas dalam model. Suatu model dikatakan baik jika eksplanasi

diukur dengan menggunakan nilai adjusted (r^2) yang setinggi mungkin.

4. Konsistensi dalam teori. Model sebaiknya relevan dengan teori, karena pengukuran tanpa teori akan dapat menyesatkan hasilnya;
5. Kekuatan prediksi. Validitas suatu model berbanding lurus dengan kemampuan prediksi model tersebut. Oleh karena itu, pilihlah suatu model yang prediksi teoretisnya berasal dari pengalaman empiris;

Untuk menentukan hubungan fungsional antar X dan Y, yang diformulasikan dengan bentuk persamaan bagi populasinya, adalah :

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

Dimana X disebut sebagai variabel bebas dan Y merupakan variabel tidak bebas, β_0, β_1 adalah parameter koefisien regresi yang belum diketahui dan oleh karena itu nilainya akan ditaksir, ε adalah residu (*error*).

Selanjutnya untuk menaksir konstanta (β_0) dan koefisien model regresi (β_1), sebut saja nilai taksirannya : b_0 dan b_1, dapat digunakan Metode untuk menaksir koefisien regresi, antara lain dengan Metode Kuadrat Terkecil (MKT) atau *Method Of ordinary Least Squares* (OLS). Pada prinsipnya metode ini menentukan nilai taksiran untuk β_0, β_1 dari data sampel yang meminimumkan jumlah kuadrat kekeliruan.

Secara matematis, rumusan menaksir koefisien tersebut, masing-masing dinyatakan sebagai berikut :

$$b_1 = \{ n \sum X_i Y_i - \sum X_i \sum Y_i \} / \{ n \sum X_i^2 - (\sum X_i)^2 \} \ldots\ldots\ldots\ldots\ldots (2)$$

dan

$$b_0 = \{ \sum Y_i / n \} - b_1 \{ \sum X_i / n \} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

Oleh karena itu, untuk memudahkan menentukan nilai taksiran koefisien model regresi tersebut, dibuatkan tabel penolong hitungan dengan format berikut ini:

No	$\mathbf{X_i}$	$\mathbf{Y_i}$	$\mathbf{X_i Y_i}$	$\mathbf{X_i^2}$
1				
2				
..				
n				
Jumlah	$\mathbf{\sum X_i}$	$\mathbf{\sum Y_i}$	$\mathbf{\sum X_i Y_i}$	$\mathbf{\sum X_i^2}$

Kedua parameter β_0 dan β_1 dalam model regresi (1) dinamakan *koefisien regresi*. β_1 adalah kemiringan (*slope*) garis regresi. Kemiringan menunjukkan perubahan rataan sebaran peluang bagi Y untuk stiap kenaikan X satu satuan. Parameter β_0 adalah nilai intersep Y garis regresi tersebut. Bila cakupan model tidak mencakup X = 0, maka β_0 tidak mempunyai makna.

C. CONTOH ANALISIS REGRESI LINEAR SEDERHANA

Guna memberikan pemahaman yang lebih jelas mengenai regresi linier sederhana, adapaun langkah-langkah yang perlu dilakukan untuk melakukan analisis dan uji regresi linier sederhana adalah sebagai berikut:

Langkah 1. Menentukan tujuan dari analisis Regresi Linear Sederhana untuk mengidentifikasi variabel *predictor* (X) dan variabel *response* (Y)

Langkah 2. Membuat hipotesis dalam bentuk hipotesis deskriptif dan statistik

Langkah 3. Melakukan pengumpulan data dan mentabulasi dalam bentuk tabel dengan menggunakan bantuan Exel

Langkah 4. Membuat tabel penolong untuk menghitung angka statistik Menghitung X^2, XY dan total dari masing-masingnya

Langkah 5. Masukkan angka-angka statistik dari tabel penolong untuk menghitung a dan b menggunakan rumus yang telah ditentukan

Langkah 6. Mencari Jumlah Kuadrat Regresi ($JK_{Reg\ [a]}$) dengan rumus:

$$JK_{\text{Reg (a)}} = \frac{(\sum Y)^2}{n}$$

Langkah 7. Mencari Jumlah Kuadrat Regresi ($JK_{Reg\ [b|a]}$) dengan rumus:

$$JK_{\text{Reg [b|a]}} = b.\ \{\sum XY - \frac{(\sum X)(\sum Y)}{n}$$

Langkah 8. Mencari Jumlah Kuadrat Residu (JK_{Res}) dengan rumus:

$$JK_{\text{Res}} = \sum Y^2 - JK_{Reg[b|a]} - JK_{Reg[a]}$$

Langkah 9. Mencari Rata-rata Jumlah Kuadrat Regresi ($RJK_{Reg\ [a]}$) dengan rumus: $RJK_{\ Reg[a]} = JK_{Reg[a]}$

Langkah 10. Mencari Rata-rata Jumlah Kuadrat Regresi ($RJK_{Reg\ [b|a]}$ dengan rumus: $RJK_{\ Reg[b|a]} = JK_{Reg[b|a]}$

Langkah 11. Mencari Rata-rata Jumlah Kuadrat Residu (JK_{Res}) dengan rumus:

$$RJK_{\ Res} = \frac{JK_{Res}}{n-2}$$

Langkah 12. Menguji signifikansi pada uji regresi seperti uji-t, uji-F. Dalam pengujian signifikansi yang dicontohkan menggunakan uji-F dengan rumus:

$$F_{hitung} = \frac{RJK_{Reg(b|a)}}{RJK_{Res}}$$

Kaidah penguji signifikansi:

Jika $F_{hitung} \geq F_{\ tabel}$, maka tolak H_0 artinya signifikan dan

$F_{hitung} \leq F_{\ tabel}$, terima H_0 artinya tidak signifikan

Dengan taraf signifikan : a = 0,01 atau a = 0,05

Carilah nilai F tabel menggunakan Tabel F dengan rumus:

$F_{\ tabel} = F_{\{1-\alpha)\ (dk\ Reg\ [b|a],\ (dk\ Res)\}}$

Langkah 13. Membuat kesimpulan

Langkah 14. Menghitung kadar hubungan antara X dan Y atau sumbangan X terhadap Y atau disebut dengan koefisien determinasi. Koefisien determinasi (*coefficicient of determination*) dilambangkan dengan r^2 dan umumnya dinyatakan dalam persentase (%). Koefisien determinasi adalah nilai yang digunakan untuk mengukur besarnya konstribusi variabel independent (x) terhadp variasi (naik/turunnhya) variabel dependent (y). Dengan kata lain, variabel y dapat dijelaskan oleh varabel x sebesar r^2% dan sisanya dijelaskan oleh variabel lain. Variasi y lainnya (sisanya) disebabkan oleh faktor lain yang juga memengaruhi y dan sudah

termasuk dalam kesalahan pengganggu (*disturbance error*) dengan formula rumus:

$$r^2 = \frac{JK(TD) - JK(S)}{JK(TD)}$$

Guna memberikan pemahaman yang lebih jelas mengenai langkah-langkah analisis regresi linier sederhana, diberikan suatu contoh kasus, yaitu : Suatu data penelitian tentang kemampuan analisis statistika inferensial mahasiswa pada mata kuliah Statistika Pendidikan yang diprediksi dipengaruhi oleh motivasi belajarnya. Bagaimana menganalisis kasus ini ? Untuk menganalisis kasus ini, hal-hal dilakukan adalah:

Contoh: Judul Penelitian

”PENGARUH MOTIVASI BELAJAR TERHADAP KEMAMPUAN ANALISIS STATISTIKA MAHASISWA”

Data dianggap memenuhi asumsi dan persyaratan analisis; data dipilih secara random; berdistribusi normal; berpola linier; data sudah homogen. Datanya sebagai berikut:

No. Responden	X	Y
1	34	32
2	38	35
3	34	31
4	40	38
5	30	29
6	40	35
7	40	33
8	34	30
9	35	32
10	39	36
11	33	31
12	32	31
13	42	36
14	40	37
15	42	35
16	42	38
17	41	37

18	32	30
19	34	30
20	36	30
21	37	33
22	36	32
23	37	34
24	39	35
25	40	36
26	33	32
27	34	32
28	36	34
29	37	32
30	38	34
Jumlah	1105	1000

Pertanyaan :

a. Bagaimana persamaan regresinya?
b. Gambarkan arah garis regresi!
c. Apakah motivasi belajar (X) terhadap kemampuan analisis statistika inferensial (Y) mahasiswa pada mata kuliah Statistika Pendidikan?
d. Buktikan apakah data tersebut berpola linier?

Langkah-langkah analisis:

Langkah 1.

Hipotesis deksriptif:

H_0 : Tidak ada pengaruh motivasi belajar (X) terhadap kemampuan analisis statistika inferensial (Y) mahasiswa pada mata kuliah Statistika Pendidikan

H_1 : Ada pengaruh motivasi belajar (X) terhadap kemampuan analisis statistika inferensial (Y) mahasiswa pada mata kuliah Statistika Pendidikan

Langkah 2.

Hipotesis statistik:

H_1: $\beta \neq 0$

H_0: $\beta = 0$

Langkah 3.

Membuat tabel penolong untuk mencari nilai-nilai dalam persamaan regresi sebagai berikut:

$\sum Xi = 1105$

$\sum Yi = 1000$

$\sum XiYi = 37056$

$\sum Y^2 = 41029$

$\sum Xi^2 = 33528$

No. Responden	X_i	Y_i	$X_i Y$	X_i^2	Y_i	
1	34	32	1088	1156	1024	
2	38	35	1330	1444	1225	
3	34	31	1054	1156	961	
4	40	38	1520	1600	1444	
5	30	29	870	900	841	
6	40	35	1400	1600	1225	
7	40	33	1320	1600	1089	
8	34	30	1020	1156	900	
9	35	32	1120	1225	1024	
10	39	36	1404	1521	1296	
11	33	31	1023	1089	961	
12	32	31	992	1024	961	
13	42	36	1512	1764	1296	
14	40	37	1480	1600	1369	
15	42	35	1470	1764	1225	
16	42	38	1596	1764	1444	
17	41	37	1517	1681	1369	
18	32	30	960	1024	900	
19	34	30	1020	1156	900	
20	36	30	1080	1296	900	
21	37	33	1221	1369	1089	
22	36	32	1152	1296	1024	
23	37	34	1258	1369	1156	

24	39	35	1365	1521	1225	
25	40	36	1440	1600	1296	
26	33	32	1056	1089	1024	
27	34	32	1088	1156	1024	
28	36	34	1224	1296	1156	
29	37	32	1184	1369	1024	
30	38	34	1292	1444	1156	
$\sum$	1105	1000	37056	41029	33528	

Langkah 4.

Masukkan angka-angka statistik dari tabel penolong dengan menggunakan rumus:

Menghitung b_1:

$b_1 = \{ n \sum X_i Y_i - \sum X_i \sum Y_i \} / \{ n \sum X_i^2 - (\sum X_i)^2 \}$

$\{ 30 \sum 37056 - \sum 1105 \sum 1000 \} / \{ 30 \sum 4109 - (\sum 1105)^2 \}$

$\{1111680 - 1105000\}/\{ 1230870 - 1221025\}$

$=6680 / 9845= 0.678517$ dibulatkan 0.68

Menghitung b_0:

$b_0 = \{ \sum Y_i / n \} - b_1 \{ \sum X_i / n \}$

$\{ 1000 / 30 \} - 0.68 \{ 1105 / 30 \}$

$=33.33 - 0.68 \{ 36.83 \}$

$=33.33 - 25.04 = 8,29$

Persamaan regresinya:

$Y = 8.29 + 0.68 X$

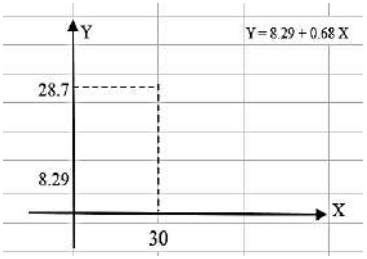
Penafsiran persamaan:

Koefisien garis sebesar 0.68 menunjukkan bahwa setiap kenaikan X sebanyak satu satuan, prediksi rata Y akan naik sebesar 0.68 satuan. Artinya jika skor motivasi belajar naik sebesar 1 point akan mengakibatkan kenaikan prediksi skor kemam puan analisis statistika mahasiswa sebesar 0.68

Langkah 5.

Membuat garis persamaan regresi:

Jika X=30, maka Y=8.29 + 0.68 (30)=28.70



Gambar 2: Persamaan Garis Regresi

Langkah 6.

Menguji kelinieran dan keberartian Regresi

Hipotesis yang diuji adalah:

H_0: Harga F regresi tidak signifikan/tidak bermakna/tidak berarti

H_1: Harga F regresi signifikan/ bermakna/ berarti

Langkah-langkah pengujian hipotesis:

Langkah (1). Mengurutkan data X dari yang terkecil sampai terpesar, diikuti oleh data Y

Setelah data diurutkan, dikelompokkan data X dan Y seperti disajikan pada tabel berikut:

X	Kelompok	ni	Y
30	1	1	29
32	2	2	31
32			30
33	3	2	31
33			32
34	4	5	32
34			31
34			30
34			30
34			32
35	5	1	32
36	6	3	30
36			32
36			34
37	7	3	33
37			34
37			32
38	8	2	35
38			34
39	9	2	36
39			35
40	10	5	38
40			35
40			33
40			37
40			36
41	11	1	37
42	12	3	36
42			35
42			38

jumlah kelompok ke i

Dari tabel di atas dapat terlihat terdapat 12 kelompok

Langkah (2). Menghitung jumlah kuadrat:

$\boldsymbol{JK}_{(total)}$ = 33528

$\boldsymbol{JK}_{Reg\ (a)}$ =33333.33

$\boldsymbol{JK}_{Reg\ (b/a)}$ = 151.41

$\boldsymbol{JK}_{Res}$ = 43.26

Mencari $\boldsymbol{JK}_{(G)}$ dengan rumus:

$$. JK(G) = \sum\left\{\Sigma Y^2 - \frac{(\Sigma Y)^2}{n}\right\}$$

$$=\left\{29^2 - \frac{(29)^2}{1}\right\} + \left\{31^2 + 30^2 - \frac{(31+30)^2}{2}\right\} + \left\{31^2 + 32^2 - \frac{(31+32)^2}{2}\right\} + \left\{32^2 + 31^2 + 30^2 + 30^2 + 32^2 - \frac{(32+31+30+30+32)^2}{5}\right\} + \left\{32^2 - \frac{(32)^2}{1}\right\} + \left\{30^2 + 32^2 + 34^2 - \frac{(30+32+34)^2}{3}\right\} + \left\{33^2 + 34^2 + 32^2 - \frac{(33+34+32)^2}{3}\right\} + \left\{36^2 + 34^2 - \frac{(36+34)^2}{2}\right\} + \left\{36^2 + 35^2 - \frac{(36+35)^2}{2}\right\} + \left\{38^2 + 35^2 + 33^2 + 37^2 + 36^2 - \frac{(38+35+33+37+36)^2}{5}\right\} + \left\{37^2 - \frac{(37)^2}{1}\right\} + \left\{36^2 + 35^2 + 38^2 - \frac{(36+35+38)^2}{3}\right\} = 37.67$$

$\boldsymbol{JK}_{(TC)}$ = JK(S) −JK(G) =43.26−37.67 = 5.59

Langkah (3). Menghitung derajat kebebasan :

dk(a)=1 dk=derajat kebebasan=*degree of freedom* (df)

$dk_{[b|a]=1}$ jumlah prediktor=1

dk sisa= n−2=30-2=28

dk tuna cocok =k-2=12−2=10 (jumlah pengelompokkan data X=12)

dk galat = n−k=30-12= 18

Langkah (4). Menghitung rata-rata jumlah kuadrat (RJK)

RJK(T)=JK(T)/n=33528/30=1117.6

$\boldsymbol{RJK}_{(reg\ a)}$ = $\boldsymbol{JK}_{reg\ a}$ /$dk_{(reg\ a)}$ =33333.33

$\boldsymbol{JK}_{(reg\ b|a)}$ = $\boldsymbol{JK}_{(reg\ b|a)}$ /$dk_{(reg\ b|a)}$=151.41/1=151.41

$\boldsymbol{JK}_{(res)}$ = $\boldsymbol{JK}_{(res)}$ /$dk_{(res)}$=43.26/2=1.54

$\boldsymbol{JK}_{(TC)}$ = $\boldsymbol{JK}_{(TC)}$ /$dk_{(TC)}$=5.59/10=0.56

$\boldsymbol{JK}_{(E)}$ = $\boldsymbol{JK}_{(E)}$ /$dk_{(E)}$=37.67/18=2.09

Langkah (5). Memasukan perhitungan kedalam tabel ANOVA untuk regresi linear

$F_{(sig)}$ = R$JK_{(reg\ b|a)}$ /RJK $_{(res)}$=151.41/1.54=98.01

$F_{(line)}$ = R$JK_{(TC)}$ /RJK $_{(E)}$=0.56/2.09=0.27

Tabel ringkasan ANOVA untuk menguji keberartian dan linieritas regresi

Sumber Variasi	dk	JK	RJK	F_{hitung}	F_{tabel}
Total	30	33528	1117.6	-	-
Regresi (a)	1	33333.33	33333.33	-	-
Regresi (alb)	1	151.41	151.41	98.01	4.2
Residu	28	43.26	1.54		
Tuna cocok	10	5.59	0.56	0.27	2.42
Error (Galat)	18	37.67	2.09		

Langkah (6). Membuat kesimpulan

Aturan pengambilan keputusan:

Jika F_{hitung} (regresi) > F_{tabel} maka harga F_{hitung} signifikan berarti koefisien regresi adalah berarti (bermakna)

Dalam perhitungan diperoleh:

F_{hitung} =98.01 sedangkan F_{tabel} untuk dk 1 : 28 (pembilang=1; penyebut =28) pada taraf signifikansi 5%=0.05= 4.20

Ini berarti harga F_{hitung} > F_{tabel}, sehingga H_0 ditola dan H_1 diterima, karena F regresi adalah signifikan. Dengan demikian dapat disimpulkan terdapat hubungan yang signifikan antara motivasi belajar dan kemampuan analisis statistika inferensial mahasiswa pada mata kuliah Statistika Pendidikan.

Menentukan keputusan menguji linieritas:

Jika harga F_{hitung} (tuna cocok) < harga F_{tabel}, regresi Y atas X adalah berpola linear

Dari perhitungan diperoleh:

F_{hitung} (tuna cocok) = 0.27, sedangkan harga F_{tabel} untuk taraf signifikansi 5%=2.42

Ternyata $F_{hitung} < F_{tabel}$ atau $0.27 < 2.42$, maka tolak Ho artinya data berpola linier. Kesimpulan variabel motivasi belajar terhadap kemampuan analisis statistika inferensial mahasiswa berpola linier.

Langkah 7. Menghitung kadar hubungan antara X dan Y atau sumbangan X terhadap Y

Koefisien korelasi (r) dapat dihitung dengan rumus berikut:

$$r^2 = \frac{JK(TD) - JK(S)}{JK(TD)}$$

Dimana: $JK_{(TD)}$=Jumlah kuadrat total dikoreksi

$JK_{(TD)} = JK_{(TD)} - JK_{(reg\ a)} = 33528 - 33333.33 = 194.67$

$$Jadi\ r^2 = \frac{194.67 - 43.26}{194.67} = 0.778$$

Koefisien korelasinya (r) = $\sqrt{0.778} = 0.881$

Interpretasi:

Sumbangan motivasi belajar terhadap kemampuan analisis statistika inferensial mahasiswa adalah sebesar 77,8% sedangkan sisanya (residunya) sebesar 22,2% dijelaskan oleh variabel lain yang tidak diteliti. Dengan kata lain motivasi belajar dapat memprediksi kemampuan analisis statistika inferensial mahasiswa sebesar 77,8%. Sedangkan sisanya sebesar 22,2% dijelaskan atau dipengaruhi oleh variabel lain yang tidak diteliti.

D. ANALISIS REGRESI LINEAR SEDERHANA DENGAN BANTUAN PROGRAM KOMPUTER

Analisis regresi linear sederhana dengan menggunakan bantuan komputer maksudnya adalah menganalisis termasuk mengolah dan menelaah data secara kuantitatif yang perhitungannya tidak dilakukan secara manual tetapi menggunakan bantuan program komputer. Keuntungannya adalah analisis data yang dilakukan memiliki tingkat akurasi hitungan yang lebih tinggi jika dibandingkan dengan analisis secara manual. Hanya dengan hitungan menit bahkan detik dapat diperoleh output analisis datanya sehingga sangat simpel, fleksibel, ringkas dan tepat.

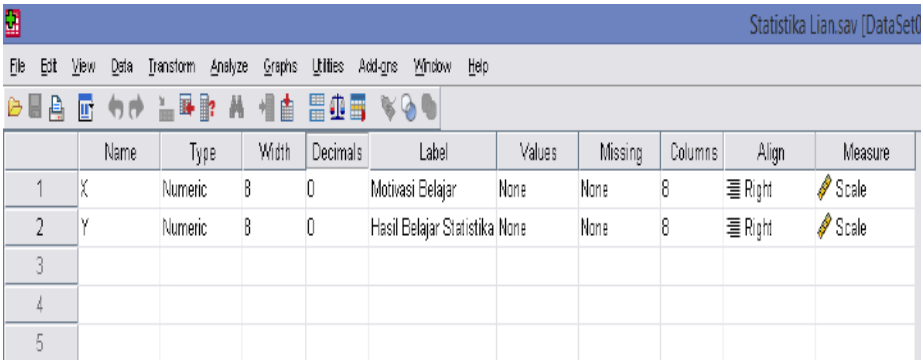
Program komputer yang dipergunakan untuk menganalisis data modelnya bermacam-macam tergantung tujuan dan maksud analisis yang diperlukan. Program komputer yang dikenal secara umum dan banyak digunakan salah satunya adalah *Statistical Program for Social Science*

(SPSS). Penggunaannya sangat mudah dipahami oleh penggunanya baik mahasiswa, guru, dosen, dan peneliti, karena semua data diinput dalam format SPSS (*Variabel view*, dan *Data view*) kemudian tinggal memilih uji statistika yang akan dipergunakan pada menu *Analize*. SPSS menyediakan menu khusus *Regression* yang meliputi banyak perhitungan model regresi, seperti regresi linier, curve estimation, regresi non-linier dan lainnya.

Tahapan analisis regresi linear sederhana diawali dengan pemenuhan uji asumsi klasik. Uji asumsi ini merupakan syarat-syarat yang harus dipenuhi pada model regresi linear OLS (*ordinary least square*) agar model tersebut menjadi valid sebagai alat penduga. Uji Asumsi Klasik pada regresi linear sederhana antara lain: data interval atau rasio, linearitas, normalitas, heteroskedastisitas, outlier, autokorelasi (Hanya untuk data time series atau runtut waktu).

Berikut ini merupakan contoh analisis data regresi linear sederhana dengan menggunakan bantuan SPSS.

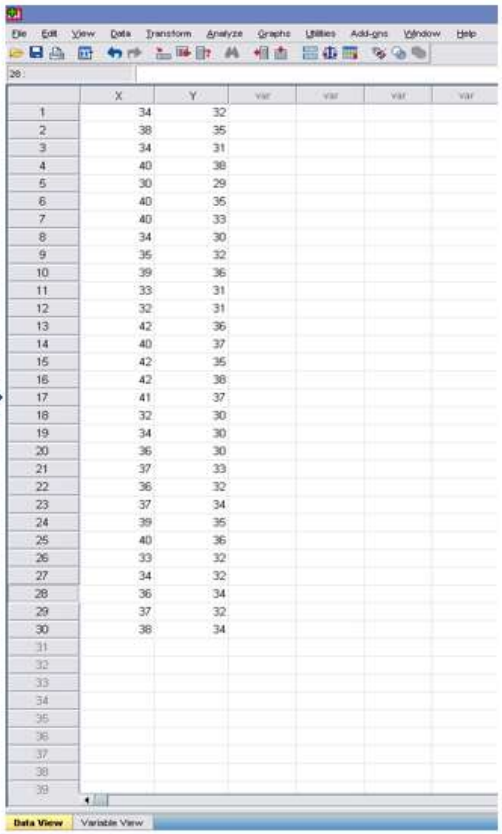
Buka Program SPSS, Klik Variable View, pada bagian Name ketik nama variabel X dan Y, pada Decimals ubah semua menjadi angka 0, pada bagian Label tuliskan nama masing-masing variabel, sebagaimana tampilan berikut.



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	X	Numeric	8	0	Motivasi Belajar	None	None	8	Right	Scale
2	Y	Numeric	8	0	Hasil Belajar Statistika	None	None	8	Right	Scale
3										
4										
5										

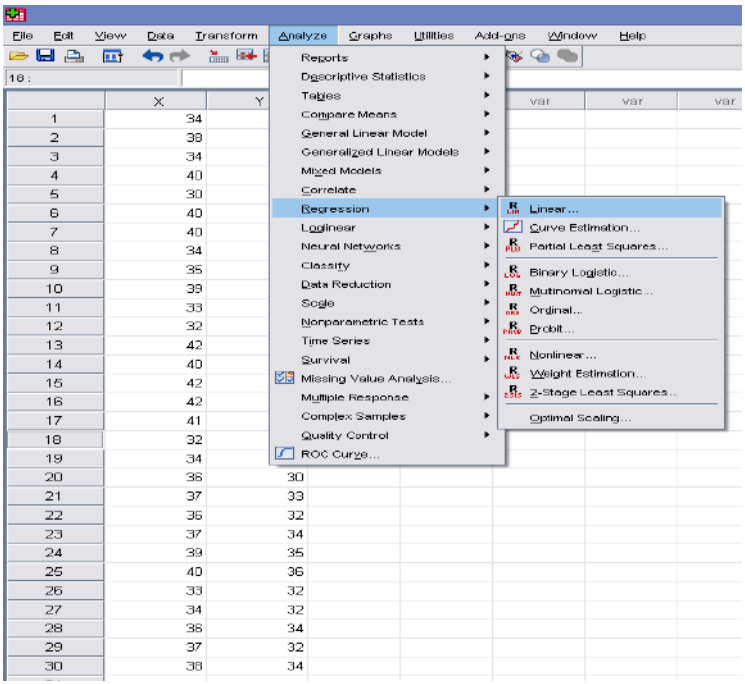
Selanjutnya Klik *Data View*, masukkan data hasil penelitian masing-masing variabel dengan cara copy paste data dari excel ke SPSS Data View, sebagaimana tampilan berikut.

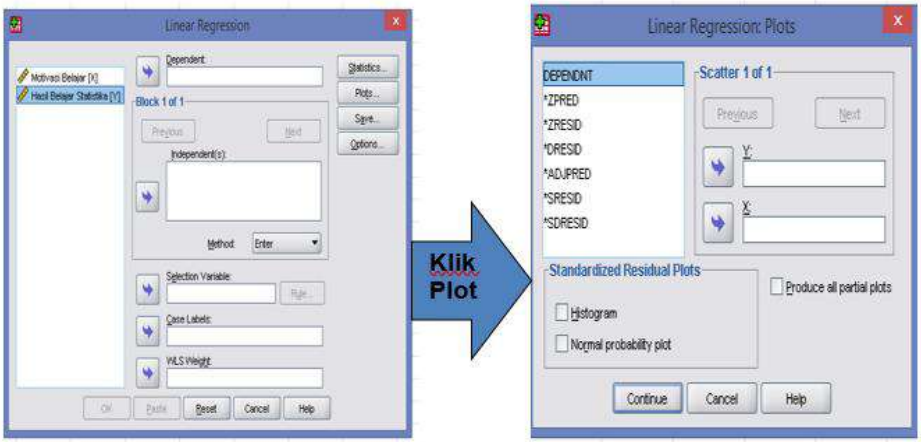
No. Responden	X	Y
1	34	32
2	38	35
3	34	31
4	40	38
5	30	29
6	40	35
7	40	33
8	34	30
9	35	32
10	39	36
11	33	31
12	32	31
13	42	36
14	40	37
15	42	35
16	42	38
17	41	37
18	32	30
19	34	30
20	36	30
21	37	33
22	36	32
23	37	34
24	39	35
25	40	36
26	33	32
27	34	32
28	36	34
29	37	32
30	38	34
Jumlah	1105	1000

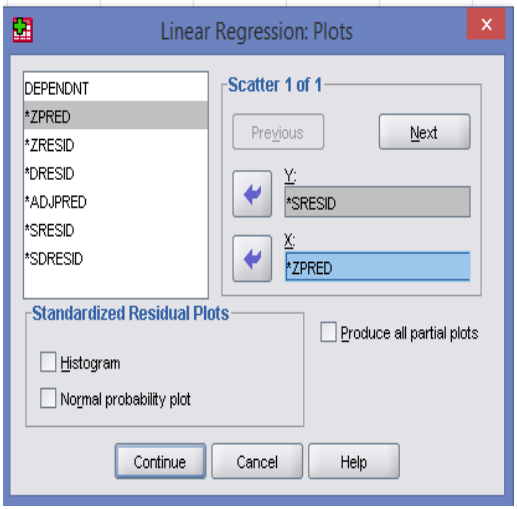


Langkah-langkah analisis regresi linear sederhana pada SPSS sebagai berikut:

1) Untuk analisis data, klik menu **Analyze >> Regression >> Linear**
2) Pada kotak dialog Linear Regression, masukkan variabel Y ke kotak Dependent, kemudian masukkan variabel X ke kotak Independent(s).
3) Klik tombol Plots, maka akan terbuka kotak dialog 'Linear Regression: Plots'.
4) Klik *SRESID (Studentized Residual) lalu masukkan ke kotak Y dengan klik tanda penunjuk. Kemudian klik *ZPRED (Standardized Predicted Value) lalu masukkan ke kotak X. Jika sudah klik tombol Continue. Akan terbuka kotak dialog sebelumnya, klik tombol OK, dengan tampilan sebagai berikut.


File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Neural Networks
Classify
Data Reduction
Scale
Nonparametric Tests
Time Series
Survival
Missing Value Analysis...
Multiple Response
Complex Samples
Quality Control
ROC Curve...
Linear...
Curve Estimation...
Partial Least Squares...
Binary Logistic...
Multinomial Logistic...
Ordinal...
Probit...
Nonlinear...
Weight Estimation...
2-Stage Least Squares...
Optimal Scaling...
X
Y
var


Linear Regression
Dependent:
Block 1 of 1
Independent(s):
Method: Enter
Selection Variable:
Case Labels:
WLS Weight:
Statistics...
Plots...
Save...
Options...
OK
Paste
Reset
Cancel
Help
Klik Plot
Linear Regression: Plots
DEPENDNT
*ZPRED
*ZRESID
*DRESID
*ADJPRED
*SRESID
*SDRESID
Scatter 1 of 1
Previous
Next
Y:
X:
Standardized Residual Plots
Histogram
Normal probability plot
Produce all partial plots
Continue
Cancel
Help



Output untuk analisis regresi dengan langkah-langkah analisis SPSS di atas dapat diinterpretasi sebagai berikut.

Regression

```
[DataSet0] C:\Users\ASUS\Documents\Statistika Lian.sav
```

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	Motivasi Belajar[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: Hasil Belajar Statistika

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.881[a]	.776	.768	1.248

a. Predictors: (Constant), Motivasi Belajar

b. Dependent Variable: Hasil Belajar Statistika

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	151.083	1	151.083	97.062	.000[a]
	Residual	43.584	28	1.557		
	Total	194.667	29			

a. Predictors: (Constant), Motivasi Belajar

b. Dependent Variable: Hasil Belajar Statistika

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	8.341	2.547		3.275	.003
	Motivasi Belajar	.679	.069	.881	9.852	.000

a. Dependent Variable: Hasil Belajar Statistika

Output Model Summary dapat diinterpretasi sebagai berikut:

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.881[a]	.776	.768	1.248

a. Predictors: (Constant), Motivasi Belajar

b. Dependent Variable: Hasil Belajar Statistika

Tabel di atas menjelaskan tentang besarnya nilai korelasi hubungan yang dilambangkan dengan (R), yaitu sebesar 0,881. Sedangkan pada kolom R Square sebesar 0,776 menjelaskan besarnya persentase (%) pengaruh variabel Independent (X) yaitu motivasi belajar terhadap variabel Dependent (Y) Hasil belajar statistika yang disebut dengan koefisien determinasi. Nilai koefisien determinasi R Square artinya bahwa pengaruh variabel motivasi belajar terhadap variabel hasil belajar statistika adalah sebesar 77,6%, sedangkan sisanya dipengaruhi oleh variabel lain yang tidak diteliti atau dikontrol dalam penelitian ini yaitu sebesar 22,4%.

Output *Coefficients* dapat diinterpretasi sebagai berikut :

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	8.341	2.547		3.275	.003
	Motivasi Belajar	.679	.069	.881	9.852	.000

a. Dependent Variable: Hasil Belajar Statistika

Pada tabel *Coefficients* pada kolom B nilai Constants (a) adalah 8.341, sedangkan nilai Motivasi Belajar (b) adalah 0.679, sehingga persamaan regresi dapat ditulis Y= a+bx = 8.341+ 0.679x. Koefisien b dinamakan koefisien arah regresi yang menyatakan perubahan rata-rata variabel Y untuk setiap perubahan variabel X sebesar satu satuan. Perubahan ini merupakan pertambahan bila b bertanda positif (+) dan penurunan bila b bertanda negatif (-). Output *Coefficients* dapat diinterpretasi sebagai berikut.

Penjelasan Persamaan Regresi Y= 8.341+ 0.679x

Nilai konstanta (a)=8.341. Artinya apabila motivasi belajar sama dengan nol atau tidak ada perubahan, maka hasil belajar statistika sebesar 8.341

Koefisien regresi motivasi belajar (b)= + 0.679 koefisien regresi positif searah 0.679, artinya jika motivasi belajar sebesar 1 satuan maka hasil belajar statistika akan meningkat sebesar 0.679. Artinya jika motivasi belajar meningkat sebesar 0.679 maka hasil belajar statistika akan meningkat sebesar 0.679

Persamaan regresi ini menampilkan uji signifikansi dengan **uji t** yaitu untuk mengetahui apakah ada pengaruh yang signifikan antara variabel X terhadap variabel Y. Dari output diatas tabel *Coefficients* diketahui nilai t_{hitung} 9.852 dengan nilai signifikansi 0.000 < 0.05 maka H_0 ditolak, artinya ada pengaruh yang signifikan variabel motivasi belajar terhadap hasil belajar statisika.

Cara 1: Membandingkan nilai t hitung dan t tabel

Jika nilai $t_{hitung} < t_{tabel}$, maka H_0 diterima artinya tidak ada pengaruh yang signifikan*Jika nilai $t_{hitung} > t_{tabel}$, maka H_0 ditolak artinya ada pengaruh yang signifikan*

Dari contoh penelitian di atas dapat dilihat dari output tabel c*oefficients** diketahui nilai t_{hitung} adalah= 9.852 dan nilai t tabelnya adalah 2.025 yang berarti nilai $t_{hitung} > t_{tabel}$*, maka H_0 ditolak. Artinya H_0 ditolak, motivasi belajar berpengaruh signifikan terhadap hasil belajar statistika*.

Tingkat signifikansi 5%

Df=jumlah sampel - jumlah variabel) df* n-k =30-2=28

n=jumlah sampel

K=jumlah variabel

Uji dua arah/dua sisi

karena pengujian 2 arah maka =5%: 2= 2,5%:100=0,025

*. Sehingga nilai t tabel dari 28 pada kolom 0,025 =adalah sebesar 2.025 Darimana bisa melihat nilai t tabel 2.025 cara melihatnya melalui tabel **Tabel Harga Kritik Untuk t** (Tabel distribusi t)

Tabel Harga Kritik Untuk t

	Level of significance for one-tailed test					
	.10	.05	.025	.01	.005	.0005
	Level of significance for one-tailed test					
df	.20	.10	.05	.02	.01	.001
1	3,078	6,314	12,706	31,821	63,657	636,619
2	1,886	2,920	4,303	6,965	9,925	31,598
3	1,638	2,353	3,182	4,541	5,841	12,941
4	1,533	2,132	2,770	3,747	4,604	8,613
5	1,476	2,015	2,571	3,365	4,032	6,859
6	1,440	1,943	2,447	3,143	3,707	5,959
7	1,415	1,895	2,365	2,998	3,499	5,405
8	1,397	1,860	2,306	2,896	3,355	5,041
9	1,383	1,833	2,262	2,821	3,250	4,781
10	1,372	1,812	2,228	2,764	3,169	4,587
11	1,363	1,796	2,201	2,718	3,106	4,437
12	1,356	1,782	2,179	2,681	3,055	4,318
13	1,350	1,771	2,160	2,650	3,012	4,221
14	1,345	1,761	2,145	2,624	2,977	4,140
15	1,341	1,753	2,131	2,602	2,947	4,073
16	1,337	1,746	2,120	2,583	2,921	4,015
17	1,333	1,740	2,110	2,567	2,898	3,965
18	1,330	1,734	2,101	2,552	2,878	3,922
19	1,328	1,729	2,093	2,539	2,861	3,883
20	1,325	1,725	2,086	2,528	2,845	3,850
21	1,323	1,721	2,080	2,518	2,831	3,819
22	1,321	1,717	2,074	2,508	2,819	3,792
23	1,319	1,714	2,069	2,500	2,807	3,767
24	1,318	1,711	2,064	2,492	2,797	3,745
25	1,316	1,708	2,060	2,485	2,787	3,725
26	1.315	1,706	2,056	2,479	2,779	3,707
27	1,314	1,703	2,052	2,473	2,771	3,690
28	1,313	1,701	2,052	2,467	2,763	3,674
29	1,311	1,699	2,048	2,462	2,756	3,659
30	1,310	1,697	2,045	2,457	2,750	3,646
40	1,303	1,684	2,021	2,423	2,704	3,551
60	1,296	**1,671**	2,000	2,390	2,660	3,460
120	1,289	1,658	1,980	2,358	2,617	3,373
∞	1,282	1,645	1,960	2,326	2,576	3,291

Cara 2: membandingkan nilai Signifikansi dengan probabilitas 0,05=Jika nilai Sig > 0,05= maka Ho diterima)artinya tidak ada pengaruh yang signifikan *Jika nilai Sig < 0,05 = maka Ho ditolak artinya ada pengaruh yan Nilai Sig. adalah *Sig*. 0.000 yang berarti lebih kecil dari 0.05 (0.000 < 0.05 maka H_0 ditolak* Artinya motivasi belajar ada pengaruh yang signifikan terhadap hasil belajar statistika.

DAFTAR PUSTAKA

Casella, G., and R. L. Berger. (2002). *Statistical Inference*, 2d ed. Pacific Grove, Calif.: Duxbury.

Davis, C. S. (2002). *Statistical methods for the analysis of repeated measurements*. Springer Science & Business Media.

Gujarati, D. N. (2006). *Dasar-Dasar Ekonometrika*, Jilid 1. Jakarta: Erlangga.

Kumaidi., & Manfaat, B. (2013). Pengantar Metode Statistika (Teori dan Terapannya dalam Penelitian Bidang Pendidikan dan Psikologi). Cirebon: Eduvision.

Kurniawan, R. (2016). *Analisis regresi*. Jakarta: Prenada Media.

Meeker, W. Q., & Escobar, L. A. (2014). *Statistical methods for reliability data*. John Wiley & Sons.

Ott, R. L., & Longnecker, M. T. (2015). *An introduction to statistical methods and data analysis*. Nelson Education.

Riduwan., & Sunarto. (2011). *Pengantar Statistika (Untuk Penelitian Pendidikan, Sosial, Ekonomi, Komunikasi dan Bisnis*. Bandung: Alfabeta.

Yuliara, I. M. (2016). Modul Regresi Linier Sederhana. In *Universitas Udayana*. Retrieved from ttps://simdos.unud.ac.id/uploads/file_pendidikan_1_dir/3218126438990fa0771ddb555f70be42.pdf

PROFIL PENULIS

A. Identitas Diri:

Nama	: Lian G. Otaya, M.Pd
NIP	: 198203202007102001
NIDN	: 2020038201
ID Scopus	: 57205351639
Tempat / Tanggal lahir	: Limehe Barat, 20 Maret 1982
Jenis Kelamin	: Perempuan
Pekerjaan	: Dosen Tetap pada Fakultas Ilmu Tarbiyah dan Keguruan IAIN Sultan Amai Gorontalo
Jabatan Fungsional	: Lektor Kepala, Pembina/Iva
Nomor Telp/HP	: 082189933334
Alamat	: Toto Selatan Kecamatan Kabila Kabupaten Bone Bolango Provinsi Gorontalo

B. Riwayat Pendidikan

Jenjang	Nama Lembaga	Jurusan/ Program Studi	Tahun lulus
SD	SD Muhammadiyah Limehe Barat	-	1994
SLTP	SLTP Negeri 2 Batudaa	-	1997
SLTA	SMK Negeri 1 Gorontalo	Akuntansi	2000
SI	Universitas Gorontalo	Ilmu Manajemen	2004
S2	Universitas Negeri Makassar	Penelitian dan Evaluasi Pendidikan	2012
S3	Universitas Negeri Yogyakarta	Penelitian dan Evaluasi Pendidikan	2020

Artikel yang pernah ditulis oleh penulis bisa dilihat melalui: https://scholar.google.com/citations?user=jkG55DwAAAAJ&hl=id
Email: lianotaya82@iaingorontalo.ac.id

BAB 13

KORELASI DAN ANALISIS VARIANS SATU ARAH

Siti Rahmatina, M.Pd
Universitas Iskandar Muda
Banda Aceh

A. KORELASI

Teknik korelasi product moment merupakan salah satu teknik mencari korelasi antara dua variabel atau lebih. Disebut product moment correlation karena koefisien korelasinya diperoleh dengan cara mencari hasil perkalian dari momen-momen variabel yang dikorelasikan sehingga dapat diketahui bagaimana kuat hubungan suatu variabel dengan variabel lain dengan tidak mempersoalkan apakah suatu variabel tertentu tergantung kepada variabel lain.

Koefisien korelasi merupakan indeks atau bilangan yang digunakan untuk mengukur keeratan (kuat, lemah, atau tidak ada) hubungan antarvariabel, dan memiliki nilai antara -1 dan +1 ($-1 \leq KK \leq +1$) (Hasan, 2008), dengan ketentuan sebagai berikut:

1. Jika KK bernilai positif, maka variable-variabel berkorelasi positif. Semakin dekat nilai KK ini ke +1 semakin kuat korelasinya, demikian pula sebaliknya
2. Jika KK bernilai negatif, maka variable-variabel berkorelasi negatif. Semakin dekat nilai KK ini ke -1 semakin kuat korelasinya, demikian pula sebaliknya
3. Jika KK bernilai 0, maka variable-variabel tidak menunjukkan korelasi
4. Jika KK bernilai +1 atau -1, maka variable menunjukkan korelasi positif atau negatif yang sempurna.

Teknik analisis korelasi digunakan untuk mengetahui ada atau tidak adanya kecenderungan hubungan antara dua variabel atau lebih.Dalam menggunakan teknik analisis korelasi, paling sedikit harus ada dua variabel yang dikorelasikan.Teknik analisis korelasi terutama digunakan untuk mengetahui kecenderungan hubungan antara variabel yang satu dengan variabel lainnya. Hasil analisis korelasi akan diperoleh koefisien korelasi yang menunjukkan besarnya hubungan antar variabel. Hubungan antara variabel-variabel yang dikorelasikan tersebut tidak mempermasalahkan apakah ada hubungan sebab akibat atau tidak ada hubungan sebab akibat (Budiwanto, 2017). Ada berbagai macam teknik korelasi di dalam statistika yaitu Pearson product moment (PPM), Rank Spearman, koefisien penentu (KP) atau Koefisien Determinasi (R), Tan Kendall, Biserial, Biserial Widespread, Point-Biserial, Tentrachoris, Phi, Contingensi, dan Rasio otomatis (Usman, 2009). Tentunya penggunaan rumus korelasi tersebut disesuaikan dengan jenis

variabel yang akan diukur, diantaranya jenis variabel tersebut yaitu: keduanya berskala interval, keduanya berskala ordinal, satu berskala dikotomi dan satu berskala interval, atau keduanya berskala nominal.

Selain korelasi sederhana juga terdapat korelasi linier berganda yang merupakan alat ukur untuk mengetahui hubungan antara dua variabel bebas (X) dan satu variabel terikat (Y). Beberapa teknik statistik yang digunakan antara lain koefisien korelasi linier berganda, koefisien penentu berganda (KPB) atau koefisien determinasi berganda (KDB), dan koefisien korelasi parsial (Hasan, 2009). Berikut ini akan dibahas salah satu dari beberapa teknik analisis korelasi yaitu analisis korelasi product moment dari Pearson dan teknik koefisien korelasi linier berganda.

1. Analisis korelasi product moment

Teknik analisis korelasi product moment ini diciptakan oleh Pearson, digunakan untuk menentukan kecenderungan hubungan antara dua variabel interval atau rasio. Ada empat cara menghitung koefisien korelasi product moment, yaitu menggunakan skor kasar, skor deviasi, standar deviasi, dan menggunakan scatter diagram (Budiwanto: 2017). Menghitung koefisien korelasi product moment menggunakan skor mentah, rumusnya (Hasan 2008) adalah:

$$r_{xy} = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{[N\sum X^2 - (\sum X)^2][N\sum Y^2 - (\sum Y)^2]}}$$

Korelasi Pearson menghasilkan koefesien korelasi yang berfungsi untuk mengukur kekuatan hubungan linier antara dua variabel.Jika hubungan dua variabel tidak linier, maka koefesien korelasi Pearson tersebut tidak mencerminkan kekuatan hubungan dua variabel yang sedang diteliti, meski kedua variabel mempunyai hubungan kuat.Koefisien korelasi ini disebut koefisien korelasi Pearson karena diperkenalkan pertama kali oleh Karl Pearson tahun 1990 (Firdaus, 2009). Nilai koefesien korelasi berada di antara -1 < 0 < 1yaitu apabila r = -1 korelasi negatif sempurna, artinya taraf signifikansi dari pengaruh variabel X terhadap variabel Y sangat lemah dan apabila r = 1 korelasi positif sempurna, artinya taraf signifikansi dari pengaruh variabel X terhadap variabel Y sangat kuat (Sudjana, 2005). Korelasi product moment ini digunakan jika variabel

yang dikorelasikan berbentuk gejala atau datanya bersifat continue dan sampel yang diteliti mempunya sifat homogen.

Syarat-syarat data yang digunakan dalam Korelasi Pearson, diantaranya: Bersekala interval/ rasio. variabel X dan Y harus bersifat independen satu dengan lainnya, dan variabel harus kuatitatif simetris. Asumsi dalam Korelasi Pearson diantaranya (Safitri, 2016) ialah:

a. Terdapat hubungan linier antara X dan Y
b. Data yang berdistribusi normal
c. Variabel X dan Y simetris, artinya variabel X tidak berfungsi sebagai variabel bebas dan Y sebagai variabel tergantung
d. Sampling representative
e. Varian kedua variabel sama

Menurut Johnston ciri-ciri data yang mempunyai distribusi normal (Safitri, 2016) ialah sebagai berikut:

a. Kurva frekuensi normal menunjukkan frekuensi tertinggi berada di tengah, yaitu berada pada rata-rata (mean) nilai distribusi dengan kurva sejajar dan tepat sama pada bagian sisi kiri dan kanannya. Kesimpulannya, nilai yang paling sering muncul dalam distribusi normal adalah rata-rata (average), dengan setengahnya berada dibawah rata-rata dan setengahnya yang lain berada di atas rata
b. Kurva normal, sering juga disebut sebagai kurva bel, berbentuk simetris sempurna.
c. Karena dua bagian sisi dari tengah-tengah benar-benar simetris, maka frekuensi nilai-nilai diatasrata-rata (mean) akan benar-benar cocok dengan frekuensi nilainilai di bawah rata-rata
d. Frekuensi total semua nilai dalam populasi akan berada dalam area dibawah kurva. Perlu diketahuibahwa area total dibawah kurva mewakili kemungkinan munculnya karakteristik tersebut.
e. Kurva normal dapat mempunyai bentuk yang berbeda-beda. Yang menentukan bentuk-bentuktersebut adalah nilai rata-rata dan simpangan baku (standard deviation) populasi.

Langkah-langkah menggunakan Korelasi Pearson Product Moment (r):

a. Menentukan Hipotesis pengujian :
   H_0 : r = 0 (tidak terdapat korelasi atau hubungan yang signifikan antara X dengan Y)

Ha: r ≠ 0 (terdapat korelasi atau hubungan yang signifikan antara X dengan Y)

Jika $t_{hitung} \geq t_{tabel}$, maka tolak H_0 dan korelasi signifikan

Jika $t_{hitung} \leq t_{tabel}$, maka terima H_0 dan korelasi tidak signifikan

b. Menentukan tingkat signifikan (α) Dalam menguji korelasi ini, menggunakan tingkat signifikansi (α) = 5 %

c. Uji statistik yang digunakan adalah Korelasi pearson (r), selanjutnya menghitung nilai r (Hasan 2008):

$$r_{xy} = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{[N\sum X^2 - (\sum X)^2][N\sum Y^2 - (\sum Y)^2]}}$$

Keterangan:

r_{xy} = Angka Indeks korelasi r product moment

$\sum$ XY = Jumlah hasil perkalian antara skor X dan Y

$\sum$ X = Jumlah seluruh skor X

$\sum$ Y = Jumlah seluruh Y

N = Banyaknya data

d. Menentukan interpretasi koefisien dari nilai r berdasarkan tabel berikut (Bustami, 2014):

Interval Koefisien	Tingkat Hubungan
0,01 – 0,199	Sangat rendah
0,20 – 0,399	Rendah
0,40 – 0,599	Cukup
0,60 – 0,799	Kuat
0,80 – 1,000	Sangat Kuat

Untuk menyatakan besar atau kecil sumbangan variabel (X) terhadap (Y) dapat ditentukan dengan rumus koefisien determinan:

$KP = r^2 \times 100\%$ atau $R^2 = r^2 \times 100\%$

e. Menghitung signifikan dengan rumus (Bustami, 2014)

$$t_{hitung} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

Berikut merupakan contoh pengolahan data mengunakan statistik analisis korelasi product moment:

Seseorang ingin mengetahui secara signifikan apakah terdapat korelasi antara gaji pokok (X) dengan motivasi kerja (Y) di suatu perusahaan A. Berikut datanya

No	X	Y
1	79	59
2	68	54
3	70	49
4	84	59
5	79	49
6	81	64
7	88	59
8	92	64
9	75	49
10	85	64
11	70	44
12	68	49

Berikut langkah-langkah korelasi linier sederhana

a. Membuat Hipotesis Statistik

   $Ho: r = 0$

   $Ha: r \neq 0$

b. Membuat Hipotesis penelitian

   Ho = Tidak terdapat hubungan yang signifikan antara gaji pokok (X) dengan motivasi kerja (Y) di suatu perusahaan A

   Ha = Terdapat hubungan yang signifikan antara gaji pokok (X) dengan motivasi kerja (Y) di suatu perusahaan A

c. Membuat tabel bantu statistik

No	X	Y	XY	X^2	Y^2
1	79	59	4661	6241	3481
2	68	54	3672	4624	2916
3	70	49	3430	4900	2401
4	84	59	4956	7056	3481
5	79	49	3871	6241	2401

6	81	64	5184	6561	4096
7	88	59	5192	7744	3481
8	92	64	5888	8464	4096
9	75	49	3675	5625	2401
10	85	64	5440	7225	4096
11	70	44	3080	4900	1936
12	68	49	3332	4624	2401
Jumlah	**939**	**663**	**52381**	**74205**	**37187**

d. Mencari r_{hitung} menggunakan rumus r_{xy} Product Moment.

$$r_{xy} = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{[N\sum X^2 - (\sum X)^2][N\sum Y^2 - (\sum Y)^2]}}$$

$$r_{xy} = \frac{(12)(52381) - (939)(663)}{\sqrt{[(12)(74205) - (939)^2][(12)(37187) - (663)^2]}}$$

$$r_{xy} = \frac{(628572) - (622557)}{\sqrt{[(890460) - (881721)][(446244) - (439569)]}}$$

$$r_{xy} = \frac{(6015)}{\sqrt{[8739][6675]}}$$

$$r_{xy} = \frac{(6015)}{7637{,}59}$$

$$r_{xy} = 0{,}787$$

e. Mencari besar kontribusi variabel X terhadap variabel Y

$$R^2 = r^2 \times 100\%$$

$$R^2 = 0{,}787^2 \times 100\%$$

$$R^2 = 0{,}619 \times 100\%$$

$$R^2 = 61{,}9\%$$

Ini artinya gaji pokok memberikan kontribusi terhadap motivasi kerja diperusahaan A sebesar 61%, dan sisanya ditentukan oleh faktor lainnya.

f. Menghitung nilai t_{hitung} untuk mengetahui signifikan atau tidak

$$t_{hitung} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

$$t_{hitung} = \frac{0{,}787\sqrt{12-2}}{\sqrt{1-0{,}787^2}}$$

$$t_{hitung} = \frac{0{,}787\sqrt{10}}{\sqrt{1-0{,}619}}$$

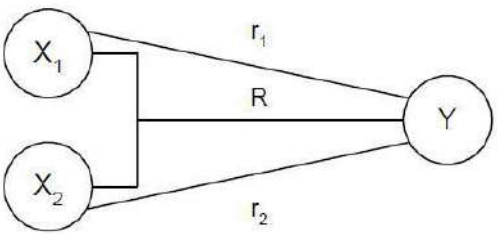
$$t_{hitung} = \frac{0{,}787(3{,}16)}{\sqrt{0{,}381}}$$

$$t_{hitung} = 4{,}03$$

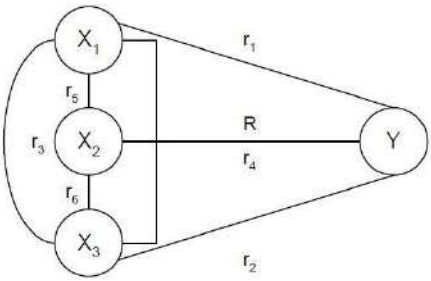
g. Membandingkan t_{hitung} dengan t_{tabel} serta penarikan kesimpulan
Menentukan taraf signifikan yang digunakan yaitu $\alpha = 0{,}05$
Degree of freedom $df = n - 2 = 12 - 2 = 10$
kemudian dengan $\alpha = 0{,}05$ dan df = 10, maka diperoleh t_{tabel} = 2,228 sehingga perbandingannya adalah $t_{hitung} > t_{tabel}$, yaitu 4,03 > 2,228, maka H_0 ditolak dan H_a diterima. Dengan demikian terdapat hubungan yang signifikan antara gaji pokok (X) dengan motivasi kerja (Y) di suatu perusahaan A.

2. Analisis korelasi ganda

Korelasi ganda merupakan angka yang menunjukkan arah dan kuatnya hubungan antara dua variabel secara bersama-sama atau lebih dengan variabel yang lain. Pemahaman tentang korelasi ganda dapat dilihat melalui gambar berikut



Gambar 1. Korelasi Ganda Dua Variabel Independen dan Satu Dependen
X_1= Kesejahteraan Karyawan
X_2= Kelengkapan sarana
Y = Kesuksesan kerja pegawai



Gambar 2. Korelasi Ganda Tiga Variabel Independen dan Satu Dependen

X_1= Kesejahteraan Karyawan

X_2 = Kelengkapan sarana

X_3 = Pengawasan

Y = Kesuksesan kerja pegawai

Berdasarkan dua gambar di atas, terlihat bahwa korelasi ganda R, bukan merupakan akumulasi dari beberapa korelasi sederhana dari setiap variabel ($r_{1,}r_2, r_{3,}$).Sehingga R ≠($r_1 + r_2 + r_{3,}$). Korelasi ganda merupakan hubungan secara bersamaan atau simultan antara X_1 dengan X_2 dan X_n dengan Y. Pada gambar 1 korelasi ganda merupakan hubungan secara bersama-sama antara variabel kesejahteraan, dan kelengkapan sarana dengan kesuksesan kerja pegawai. Langkah-langkah menghitung Koefisien ganda R:

a. Menentukan Hipotesis pengujian :

   H_0 : R_{yx1x2} = 0 (tidak terdapat korelasi atau hubungan yang signifikan antara X_1 dan X_2 dengan Y secara bersamaan atau simultan)

   Ha: $R_{yx1x2} \neq 0$ (terdapat korelasi atau hubungan yang signifikan antara X_1 dan X_2 dengan Y secara bersamaan atau simultan)

   Jika $F_{hitung} \leq F_{tabel}$, maka H_0 diterima, dalam hal lainnya H_0 ditolak

b. Menentukan tingkat signifikan (α) Dalam menguji korelasi ini, menggunakan tingkat signifikansi (α) = 5 %

c. Uji statistik yang digunakan adalah Korelasi ganda, selanjutnya menghitung nilai R(Usman, 2009):

$$R_{yx_1x_2} = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1}r_{yx_2}r_{x_1x_2}}{1 - r_{x_1x_2}^2}}$$

Keterangan:

$R_{yx_1x_2}$ = Koefisien korelasi ganda antara variabel X_1 dan X_2

$r^2_{yx_1}$ = Koefisien korelasi X_1 terhadap Y

$r^2_{yx_2}$ = Koefisien korelasi X_2 terhadap Y

$r^2_{x_1x_2}$ = Koefisien korelasi X_1 terhadap X_2

Jika harga r belum diketahui, maka hitunglah menggunakan rumus Korelasi pearson (r).

d. Menghitung signifikan dengan rumus (Usman, 2009):

$$F = \frac{\frac{R^2}{k}}{\frac{1-R^2}{n-k-1}}$$

e. Cari F_{tabel}

F_{tabel} $= \{(1-\alpha)(df = k)(df = n-k-1)\}$

$= \{(1-0{,}05)(df = 2)(df = n-2-1)\}$

$df = dk_{pembilang} = k$

$df = dk_{penyebut} = k$

keterangan:

k =banyaknya variabel bebas

n = banyaknya anggota sampel

f. Membandingkan F_{hitung} dengan F_{tabel} dengan kriteria pada langkah 1 di atas.

Berikut merupakan contoh pengolahan data mengunakan statistik analisis korelasi ganda:

Seseorang ingin mengetahui secara signifikan apakah terdapat korelasi antara Kesejahteraan Karyawan (X_1), Kelengkapan sarana (X_2), dan dengan Kesuksesan kerja pegawai (Y) secara simultan di suatu perusahaan A. Berikut datanya:

No	**X_1**	**X_2**	**Y**
1	48	44	60
2	34	53	43
3	54	37	50
4	55	61	45

5	34	23	34
6	60	46	45
7	45	60	54
8	45	53	40
9	60	55	54
10	45	34	45
11	30	45	40
12	60	65	62
13	65	51	58
14	45	52	56
15	36	50	48
16	23	18	52
17	46	35	44
18	30	40	45
19	55	74	70
20	58	50	60
21	20	35	50

Berikut langkah-langkah korelasi linier sederhana

1) Membuat Hipotesis Statistik

   Ho: $R_{yx1x2} = 0$

   Ha: $R_{yx1x2} \neq 0$

2) Membuat Hipotesis penelitian

   Ho = Tidak terdapat hubungan yang signifikan antara Kesejahteraan Karyawan (X_1), dan Kelengkapan sarana (X_2), dengan Kesuksesan kerja pegawai (Y) secara simultan di suatu perusahaan A

   Ha = Terdapat hubungan yang signifikan antara Kesejahteraan Karyawan (X_1) dan Kelengkapan sarana (X_2), dengan Kesuksesan kerja pegawai (Y) secara simultan di suatu perusahaan A

3) Membuat tabel bantu statistik

No	X_1	X_2	Y	X_1^2	X_2^2	Y^2	X_1Y	X_2Y	X_1X_2
1	48	44	60	2304	1936	3600	2880	2640	2112
2	34	53	43	1156	2809	1849	1462	2279	1802
3	54	37	50	2916	1369	2500	2700	1850	1998
4	55	61	45	3025	3721	2025	2475	2745	3355
5	34	23	34	1156	529	1156	1156	782	782
6	60	46	45	3600	2116	2025	2700	2070	2760
7	45	60	54	2025	3600	2916	2430	3240	2700
8	45	53	40	2025	2809	1600	1800	2120	2385
9	60	55	54	3600	3025	2916	3240	2970	3300
10	45	34	45	2025	1156	2025	2025	1530	1530
11	30	45	40	900	2025	1600	1200	1800	1350
12	60	65	62	3600	4225	3844	3720	4030	3900
13	65	51	58	4225	2601	3364	3770	2958	3315
14	45	52	56	2025	2704	3136	2520	2912	2340
15	36	50	48	1296	2500	2304	1728	2400	1800
16	23	18	52	529	324	2704	1196	936	414
17	46	35	44	2116	1225	1936	2024	1540	1610
18	30	40	45	900	1600	2025	1350	1800	1200
19	55	74	70	3025	5476	4900	3850	5180	4070
20	58	50	60	3364	2500	3600	3480	3000	2900
21	20	35	50	400	1225	2500	1000	1750	700
Jumlah	**948**	**981**	**1055**	**46212**	**49475**	**54525**	**48706**	**50532**	**46323**

4) Mencari r_{hitung}yaitu r_{x_1y}, r_{x_2y}, $r_{x_1x_2}$menggunakan rumusKorelasi Pearson Product Moment (r).

$$r_{x_1y} = \frac{N\sum X_1Y - (\sum X_1)(\sum Y)}{\sqrt{\left[N\sum {X_1}^2 - (\sum X_1)^2\right][N\sum Y^2 - (\sum Y)^2]}}$$

$$r_{x_1y} = \frac{(21)(48706) - (948)(1055)}{\sqrt{[(21)(46212) - (948)^2][(21)(54525) - (1055)^2]}}$$

$$r_{x_1y} = 0{,}473$$

$$r_{x_2y} = \frac{N\sum X_2Y - (\sum X_2)(\sum Y)}{\sqrt{\left[N\sum {X_2}^2 - (\sum X_2)^2\right]\left[N\sum Y^2 - (\sum Y)^2\right]}}$$

$$r_{x_2y} = \frac{(21)(50532) - (981)(1055)}{\sqrt{[(21)(49475) - (981)^2][(21)(54525) - (1055)^2]}}$$

$$r_{x_2y} = 0{,}529$$

$$r_{x_1x_2} = \frac{N\sum X_1X_2 - (\sum X_1)(\sum X_2)}{\sqrt{\left[N\sum {X_1}^2 - (\sum X_1)^2\right]\left[N\sum {X_2}^2 - (\sum X_2)^2\right]}}$$

$$r_{x_1x_2} = \frac{(21)(46323) - (948)(981)}{\sqrt{[(21)(46212) - (948)^2][(21)(49475) - (981)^2]}}$$

$$r_{x_1x_2} = 0{,}577$$

g. Uji statistik yang digunakan adalah Korelasi ganda, selanjutnya menghitung nilai R:

$$R_{yx_1x_2} = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1}r_{yx_2}r_{x_1x_2}}{1 - r_{x_1x_2}^2}}$$

$$R_{yx_1x_2} = \sqrt{\frac{(0{,}473)^2 + (0{,}529)^2 - 2(0{,}473)(0{,}529)(0{,}577)}{1 - (0{,}577)^2}}$$

$$R_{yx_1x_2} = 0{,}567$$

h. Menghitung signifikan:

$$F_{hitung} = \frac{\frac{R^2}{k}}{\frac{1 - R^2}{n - k - 1}} = \frac{\frac{(0{,}567)^2}{2}}{\frac{1 - (0{,}567)^2}{21 - 2 - 1}} = 4{,}285$$

i. Cari F_{tabel}

F tabel $= F\{(1-\alpha)(df=k)(df=n-k-1)\}$

$= F\{(1-0{,}05)(df=2)(df=n-2-1)\}$

$= \{(0{,}95),(2{,}18)\}$

Mencari F_{tabel}

Angka 2 sebagai pembilang dan angka 18 sebagai penyebut, dengan demikian $F_{tabel} = 3{,}98$

j. Penarikan kesimpulan

Perbandingannya adalah $F_{hitung} > F_{tabel}$, yaitu 4,285 > 3,98, maka H0 ditolak dan Ha diterima. Dengan demikian Terdapat hubungan yang signifikan antara Kesejahteraan Karyawan (X_1), Kelengkapan sarana (X_2), terhadap Kesuksesan kerja pegawai (Y) secara simultan di suatu perusahaan A.

B. ANALISIS VARIANS SATU ARAH

Diantara tujuan penelitian eksperimen yaitu untuk menguji perbedaan rata-rata dua data distribusi kelompok atau lebih secara bersamaan.Dua atau lebih data distribusi kelompok tersebut dapat juga dikatakan dengan dua variabel atau lebih dalam suatu penelitian eksperimen, yang kemudian diberikan perlakuan tertentu dalam kurun waktu yang telah ditentukan.Setelah mengumpulan data dari hasil tes yang diberikan, selanjutnya dilakukan pengolahan data serta analisis data untuk mendapatkan kesimpulan. Diantara cara untuk menganalisis perbedaan rata-rata dua kelompok data atau lebih tersebut salah satunya digunakan Uji statistik analisis varians yang disingkat dengan Anava atau Analysis of varians (ANOVA), yang juga disebut dengan uji-F.

Berikut akan dirangkum beberapa definisi Anava dari pakar. Anava atau Analysis of varians (Anova) adalah tergolong analisis komparatif lebih dari dua variable atau lebih dari dua rata-rata.Tujuannya ialah untuk membandingkan lebih dari dua ratarata.Gunanya untuk menguji kemampuan generalisasi artinya data sampel dianggap mewakili populasi (Riduwan, 2004).Menurut Thomas dan Nelson Analisis varians klasifikasi tunggal disebut juga analisis varians satu jalan atau analisis varians sederhana. Analisis varian klasifikasi tunggal digunakan untuk menguji perbedaan dua

mean kelompok atau lebih sampel bebas atau sampel mandiri (independent sample) (Budiwanto, 2017).Anova merupakan pengembangan lebih lanjut dari uji-t yang dikenal dengan t_{hitung}.Uji-t atau uji-z hanya dapat digunakan untuk melihat perbandingan dua data distribusi kelompok saja.Sedangkan anova satu jalur dapat melihat perbandingan lebih dari dua kelompok data.

Anava satu arah yaitu analisis yang melibatkan hanya satu peubah bebas. Secara rinci, Anava satu arah digunakan dalam suatu penelitian yang memiliki ciri-ciri berikut: Melibatkan hanya satu peubah bebas dengan dua kategori atau lebih yang dipilih dan ditentukan oleh peneliti secara tidak acak. Kategori yang dipilih disebut tidak acak karena peneliti tidak bermaksud menggeneralisasikan hasilnya ke kategori lain di luar yang diteliti pada peubah itu (Setiawan, 2019). Sebagai contoh, peneliti akan membandingkan hasil yang diperoleh dari penggunaan model A, B, dan C dalam meningkatkan motivasi belajar peserta didik serta ditidak menggeneralisasikan ke model lain selain dari pada ke tiga model yang dibandingkan tersebut.

Terdapat berapa syarat yang harus dipenuhi terlebih dahulu untuk bisa melakukan analisis uji perbedaan rata-rata menggunakan analisis varian satu arah. Syarat tersebut antara lain pengambilan data secara random dari populasi sehingga diperoleh sampel data yang representatif dari populasi penelitian. Selanjutnya data yang dianalisis haruslah berbentuk interval atau rasio, berdistribusi normal serta distribusi data dalam variabelnya bersifat homogen.

Hal tersebut sejalan dengan pendapat Sugiartodalam uji Anava, bukti sampel diambil dari setiap populasi yang sedang dikaji.Data-data yang diperoleh dari sampel tersebut digunakan untuk menghitung statistik sampel.Distribusi sampling yang digunakan untuk mengambil keputusan statistik, yakni menolak atau menerima hipotesis nol (H0), adalah Distribusi F (F Distribution).Dalam uji Anava diasumsikan bahwa semua populasi yang sedang dikaji memiliki keragaman atau varians (variance) sama tanpa mempertimbangkan apakah populasi-populasi tersebut memiliki rata-rata hitung (mean) sama atau berbeda (Sugiarto, 2009).

Distribusi sampling yang digunakan untuk pengambilan keputusan statistik untuk menolak atau menerima hipotesis nol (H0), adalah Distribusi F (F Distribution).Hipotesis nol dalam uji Anava adalah semua variabel dari populasi yang sedang dikaji memiliki rata-rata hitung (mean) sama, minimal

3 variabel atau data distribusi kelompok yang akan dilihat perbandingannya. Berikut Hipotesis dalam Anava adalah:

H_0 : 1 = 2 = 3 = … = n

H_a : Tidak semua populasi memiliki rata-rata hitung sama.

Huruf F pada uji-F diambil dari huruf pertama nama Fisher, pengembang analisis uji-F. Analisis varians (Analysis of Variance—ANOVA) merupakan metode statistika untuk mengkaji apakah rata-rata hitung (mean) dari 3 populasi atau lebih, sama atau tidak. Menurut Ardana dijelaskan bahwa jika hasil penghitungan uji beda mean F hitung yang diperoleh adalah negatif, maka tanda negatif tersebut diabaikan, yang digunakan adalah angka absolutnya. Berikut ini disajikan tiga teknik analisis varians yaitu teknik analisis varians klasifikasi tunggal, analisis varians klasifikasi ganda, dan analisis varians amatan ulangan (Budiwanto, 2017).

Anava menjadi bagian yang ada dalam teknik analisis statistik parametris.Anava (analisis varians) digunakan untuk menguji hipoteisis komparatif rata-rata K sampel bila datanya berbentuk interval atau rasio.Analisis varians klasifikasi tunggal (single classification).Anava jenis ini sering disebut juga dengan anava satu jalan (one way anova).Anava jenis ini digunakan untuk menguji hipotesis komparatif rata-rata K sampel secara serempak. Setiap sampel akan mempunyai mean (ratarata) dan varians (simpangan baku kuadrat). Ada dua mean (rata-rata) dalam anava ini yaitu mean dalam kelompok yaitu mean tiap-tiap kelompok sampel) dan mean total yaitu mean yang merupakan gabungan dari mean tiap-tiap kelompok. Pada anava satu jalan ini juga memiliki perhitungan deviasi yang dibagi menjadi tiga bentuk yaitu deviasi total, deviasi antar kelompok dan deviasi dalam kelompok.Jumlah deviasi yang kuadratkan (JK) yaitu variansi.Karena pengujian hipotesis melibatkan lebih dari dua kelompok sampel, maka akan terdapatbeberapa macam jumlah kuadrat (JK)(Sugiyono, 2007) yaitu :

1. Jumlah kuadrat total (JKtot) merupakan penjumlahan kuadrat deviasi nilai individual dengan Mtot (rata-rata total).

$$JK_{tot} = \sum X_{tot^2} - \frac{(\sum X_{tot})^2}{N}$$

2. Jumlah Kuadrat antara (JKant), merupakan jumlah selisih kuadrat mean total (Mtot) dengan Mean setiap kelompok (Mi) dikalikan dengan jumlah setiap kelompok sampel setiap kelompok.

$$JK_{ant} = \sum \frac{(\sum X_1)^2}{n_k} - \frac{(\sum X_{tot})^2}{N}$$

3. JK dalam kelompok (JKdal)

$$JK_{dal} = JK_{tot} - JK_{ant}$$

Setiap sumber variasi didampingi dengan dk (derajat kebebasan), dan dk untuk setiap sumber variasi tidak sama. Berikut pedoman penentuan dk untuk setiap varians:

- Untuk varians antar kelompok (dk = m – 1)
- Untuk variasn dalam kelompok (dk = N – m)
- Total (dk = N – 1)

Keterangan :

m = Jumlah kelompok sampel

N = Jumlah seluruh anggota sampel

Untuk dapat menghitung Fhitung, maka beberapa sumber variansi harus dihitung mean kelompoknya, yang meliputi:

- Mean antar kelompok $MK_{ant} = \frac{JK_{ant}}{m-1}$
- Mean dalam kelompok $MK_{dal} = \frac{JK_{dal}}{N-m}$

Hal yang perlu diperhatikan jika menggunakan anava ini yaitu setiap sampelnya hanya memiliki satu kategori.Misalnya bila ingin menguji hipotesis ada tidaknya perbedaan secara signifikan antara motivasi belajar peserta didik yang diajarkan dengan ModelA dan Model B, Model C, maka digunakan anava satu jalan (one way anova).

Dalam uji Anova kerap diminta untuk menempatkan ringkasan perhitungan yang dilakukan dalam bentuk table yangn berisi ringkasan nilai yang diperoleh dari proses uji statistic. Table tersebut diberi nama table Anova. Berikut table Anova satu arah (Harinaldi, 2005)

Tabel 1. Ringkasan Anova

Sumber Varian (SV)	**Derajat Bebas (db)**	**Jumlah Kuadrat (JK)**	**MK**	**F_{hitung}**	**F_{tabel}**	**keputusan**
Total	N – 1	$JK_{tot} = \sum X_{tot^2} - \frac{(\sum X_{tot})^2}{N}$		$\frac{MK_{ant}}{MK_{dal}}$	**F_{tabel}**	$F_{hitung} > F_{tabel}$, Ha diterima
Antar Kelompok	m – 1	$JK_{ant} = \sum \frac{(\sum X_1)^2}{n_k} - \frac{(\sum X_{tot})^2}{N}$	$\frac{JK_{ant}}{m-1}$			
Dalam Kelompok	N – m	$JK_{dal} = JK_{tot} - JK_{ant}$	$\frac{JK_{dal}}{N-m}$			

Berikut merupakan contoh pengolahan data mengunakan statistik Analysis of Variance:

Seseorang ingin mengetahui tingkat keberhasilan motivasi belajar dengan perbandingan tiga Model pembelajaran yang digunakannya yaitu Model A, Model B, dan Model C. Berikut datanya.

Resp.	**Model A**	**Model B**	**Model C**
1	55	45	55
2	85	55	75
3	75	55	45
4	65	65	65
5	75	45	65
6	85	45	55
7	55	45	55
8	55	55	75
9	85	45	65
10	75	55	55
11	75	60	75
12	60	68	75
13	55	75	65
14	75	65	0

Berikut langkah-langkahAnava Satu Arah

1. Membuat Hipotesis

   Hipotesis Statistik:

   $Ho: A_1 = A_2 = A_3$

   $Ha: A_1 \neq A_2 \neq A_3$

   Hipotesis penelitian:

   Ho= Tidak terdapat perbedaan yang signifikan antara Model A, Model B, dan Model C dalam meningkatkan motivasi belajar

   Ha = Terdapat perbedaan yang signifikan antara Model A, Model B, dan Model C dalam meningkatkan motivasi belajar

2. Membuat tabel bantu statistik

Resp.	Model A	Model B	Model C	Model A^2	Model B^2	Model C^2
1	55	45	55	3025	2025	3025
2	85	55	75	7225	3025	5625
3	75	55	45	5625	3025	2025
4	65	65	65	4225	4225	4225
5	75	45	65	5625	2025	4225
6	85	45	55	7225	2025	3025
7	55	45	55	3025	2025	3025
8	55	55	75	3025	3025	5625
9	85	45	65	7225	2025	4225
10	75	55	55	5625	3025	3025
11	75	60	75	5625	3600	5625
12	60	68	75	3600	4624	5625
13	55	75	65	3025	5625	4225
14	75	65	0	5625	4225	0
Jumlah	**975**	**778**	**825**	**69725**	**44524**	**53525**
Mean	69,64	55,57	58,93			
Varians	140,25	99,19	377,61			
SD	11,84	9,959	19,43			

3. Menghitung jumlah kuadrat total

$$JK_{tot} = \sum X_{tot^2} - \frac{(\sum X_{tot})^2}{N}$$

$$JK_{tot} = (69725 + 44524 + 53525) - \frac{(975 + 778 + 825)^2}{41}$$

$$JK_{tot} = (167774) - \frac{(2578)^2}{41}$$

$$JK_{tot} = (167774) - (162099{,}6)$$

$$JK_{tot} = 5674{,}39$$

4. Menghitung jumlah kuadrat antar baris

$$JK_{ant} = \sum \frac{(\sum X_1)^2}{n_k} - \frac{(\sum X_{tot})^2}{N}$$

Atau

$$JK_{ant} = \sum \frac{(\sum X_{kel})^2}{n_{kel}} - \frac{(\sum X_{ant})^2}{N}$$

$$JK_{ant} = \left(\frac{975^2}{14} + \frac{778^2}{14} + \frac{825^2}{13}\right) - \left(\frac{2578^2}{41}\right)$$

$$JK_{ant} = (67901{,}79 + 43234{,}47 + 52355{,}77) - (162099{,}6)$$

$$JK_{ant} = (163492{,}1) - (162099{,}6)$$

$$JK_{ant} = 1392{,}52$$

5. Menghitung jumlah kuadrat dalam kelompok

$$JK_{dal} = JK_{tot} - JK_{ant}$$

$$JK_{dal} = 5674{,}39 - 1392{,}52$$

$$JK_{dal} = 4281{,}87$$

6. Menghitung derajat bebas antar grup

$$dk = m - 1 = 3 - 1 = 2$$

7. Menghitung mean antar kelompok

$$MK_{ant} = \frac{JK_{ant}}{m - 1}$$

$$MK_{ant} = \frac{1392{,}52}{3 - 1}$$

$$MK_{ant} = 696{,}26$$

8. Menghitung derajat bebas dalam kelompok
   $dk = N - m = 41 - 3 = 39$
9. Menghitung mean dalam kelompok
   $$MK_{dal} = \frac{JK_{dal}}{N - m}$$
   $$MK_{dal} = \frac{4281,87}{41 - 3}$$
   $$MK_{dal} = 112,68$$
10. Menghitung F_{hitung}
    $$F_{hitung} = \frac{MK_{ant}}{MK_{dal}}$$
    $$F_{hitung} = \frac{696,26}{112,68}$$
    $$F_{hitung} = 6,179$$
11. Menentukan Taraf signifikan yang digunakan
    Taraf signifikan yang digunakan dalam analisis ini adalah $\alpha = 0,05$
12. Membandingkan F_{hitung} dengan F_{tabel} serta penarikan kesimpulan.
    Membandingkan F_{hitung} danF_{tabel}dengan dk pembilang *m*-1 dan penyebut N- *m*
    Cari F_{tabel} dengan rumus $F_{tabel} = F_{(1-\alpha)(db_A, db_D)}$
    $$F_{tabel} = F_{(0,95)(2\ ,38)}$$
    Berdasarkan F_{tabel} dengan pembilang (3-1 = 2) dan penyebut (41-3 = 38) diperoleh nilai 3, 24
    Membuat keputusan pengujian hipotesis bila F_{hitung} lebih besar daripada F_{tabel} maka Ha diterima dan Ho ditolak
    Karena $F_{hitung} > F_{tabel}$yaitu 6,179 > 3,24maka Ha diterima
13. Tabel Ringkasan Anava satu arah untuk menguji hipotesis k sampel

Sumber Varian (SV)	**Derajat Bebas (db)**	**Jumlah Kuadrat (JK)**	**MK**	**F_{hitung}**	**F_{tabel} (5%)**
Total	41	5674,39			
Antar Kelompok	2	1392,52	696,26	6,179	3,24
Dalam Kelompok	38	4281,87	112,68		

DAFTAR PUSTAKA

Budiwanto, Setyo, (2017). Metode Statistika. Malang: Universitas Negeri Malang.

Bustami, Dahlan. A, Fadlisyah. (2014). Statistika Terapannya pada Bidang Informatika. Yogyakarta: Graha Ilmu.

Firdaus, Zamal. (2009). Korelasi antara Pelatihan Teknis Perpajakan, Pengalaman dan Motivasi Pemeriksa Pajak dengan Kinerja Pemeriksa Pajak pada Kantor Pelayanan Pajak di Jakarta Barat. Fakultas Ekonomi dan Ilmu Sosial Universitas Islam Negeri Syarif Hidayatullah Jakarta.

Harinaldi. (2005). Prinsip-prinsip Statistik untuk Taknik dan Sains. Jakarta: Erlangga.

Hasan, M. Iqbal. (2009). Pokok-pokok Materi Statistik 1 (StatistikDeskriptif). Jakarta: PT Bumi Aksara.

Hasan, M. Iqbal. (2008). Pokok-pokok Materi Statistik 2 (Statistik Inferensif). Jakarta: PT Bumi Aksara.

Riduwan, (2004). Metode dan Teknik Menyusun Tesis. Bandung: Alfabeta.

Safitri, W. R. (2016). Analisis Korelasi Pearson Dalam Menentukan Hubungan Antara Kejadian Demam Berdarah Dengue Dengan Kepadatan Penduduk Di Kota Surabaya Pada Tahun 2012 - 2014: Pearson Correlation Analysis to Determine The Relationship Between City Population Density with Incident Dengue Fever of Surabaya in The Year 2012-2014. *Jurnal Ilmiah Keperawatan (Scientific Journal of Nursing)*, *2*(2),Issn:2477-4391, 21-29.

Sekaran, Uma & Bougie, R., (2010), Research Methods for Business: A Skill Building Approach, John Wiley and sons, inc: London.

Setiawan, Kukuh, (2019). Buku Ajar Metodologi Penelitian. Bandar Lampung: Universitas Lampung.

Sudjana. (2005). Metoda Statistika. Bandung: Tarsito.

Sugiharto, Toto, (2009). Bahan Ajar. Depok: Universitas Gunadarma.

Sugiyono, (2007). Statistika untuk penelitian. Bandung: Alfabeta.

Usman, H. & R. Purnomo Setiady Akbar. (2009). Pengantar Statistika. Jakarta: PT Bumi Aksara

PROFIL PENULIS

Siti Rahmatina, M.Pd. Lahir di Kopelma Darussalam tanggal 10 Oktober 1988. Lulus S1 Pendidikan Matematika IAIN Ar-Raniry Banda Aceh tahun 2011, dan lulus sekolah Pascasarjana Magister Pendidikan Matematika Universitas Syiah Kuala tahun 2014. Saat ini adalah Dosen tetap dan Ketua Program Studi Pendidikan Matematika FKIP Universitas Iskandar Muda Banda Aceh.

Penulis merupakan pengelola e-jurnal DikMas: Jurnal Pendidikan Matematika dan Sains dengan alamat https://ejournal.unida-aceh.ac.id/index.php/dikmas/index.

Artikel yang pernah ditulis oleh penulis dapat dilihat melalui https://scholar.google.co.id/citations?user=3E2t1y0AAAAJ&hl=id.

Email: siti.rahmatina@unida-aceh.ac.id

TAHTA MEDIA GROUP

IKAPI IKATAN PENERBIT INDONESIA

CV. Tahta Media Group
Surakarta, Jawa Tengah
Web : www.tahtamedia.com
Ig : tahtamediagroup
Telp/WA : +62 813 5346 4169

ISBN 978-623-5981-77-2