Editor:
Sary Shandy, S.T., M.T.
Dr. Yanneri Elfa Kiswara Rahmantya

STATISTIKA DASAR

Penulis:
Pandriadi, Vina N. Van Harling, Abdul Wahab, Sisca Vaulina, Sri Sutjiningtyas, Endang Kusdiah Ningsih, Bagus Dwi Hari Setyono, Vini Rizqi, Muhammad Iqbal Harisuddin, Syamsidar Gaffar, Trisna Yuniarti, Anisa Rahmawati, Firdhani Faujiyah, Siti Mudawanah

STATISTIKA DASAR

Penulis:
Pandriadi, Vina N. Van Harling, Abdul Wahab, Sisca Vaulina, Sri Sutjiningtyas, Endang Kusdiah Ningsih, Bagus Dwi Hari Setyono, Vini Rizqi, Muhammad Iqbal Harisuddin, Syamsidar Gaffar, Trisna Yuniarti, Anisa Rahmawati, Firdhani Faujiyah, Siti Mudawanah

Penerbit Widina

STATISTIKA DASAR

Penulis:
Pandriadi, Vina N. Van Harling, Abdul Wahab, Sisca Vaulina, Sri Sutjiningtyas, Endang Kusdiah Ningsih, Bagus Dwi Hari Setyono, Vini Rizqi, Muhammad Iqbal Harisuddin, Syamsidar Gaffar, Trisna Yuniarti, Anisa Rahmawati, Firdhani Faujiyah, Siti Mudawanah

Desain Cover:
Septian Maulana

Sumber Ilustrasi:
www.freepik.com

Tata Letak:
Handarini Rohana

Editor:
Sary Shandy, S.T., M.T.
Dr. Yanneri Elfa Kiswara Rahmantya

ISBN:
978-623-459-752-3

Cetakan Pertama:
Oktober, 2023

PENERBIT:
WIDINA MEDIA UTAMA
Komplek Puri Melia Asri Blok C3 No. 17 Desa Bojong Emas
Kec. Solokan Jeruk Kabupaten Bandung, Provinsi Jawa Barat

Anggota IKAPI No. 360/JBA/2020
Website: www.penerbitwidina.com
Instagram: @penerbitwidina
Telepon (022) 87355370

KATA PENGANTAR

Rasa syukur yang mendalam dan kami tidak bisa mengucapkan apa pun selain terima kasih. Atas karunia dan rahmat Tuhan Yang Maha Esa, buku berjudul Statistika Dasar ini dapat disusun dan diterbitkan. Kami berharap buku ini dapat memberikan kontribusi ilmiah dan memberikan wawasan bagi para peminat statistika.

Kegunaan statistik dalam penelitian sangat beragam, antara lain penggunaannya sebagai alat pengambilan sampel, untuk menguji validitas dan reliabilitas instrumen, untuk penyajian data, dan untuk analisis data. Analisis data lebih fokus pada menjawab rumusan masalah dan menguji hipotesis penelitian yang diajukan.

Banyak peneliti yang merasa khawatir sebelum melakukan penelitian karena kesulitan menggunakan statistik. Dalam praktiknya, hal ini tidak terjadi, selama Anda memahami jenis data dan jenis hipotesis yang akan diuji. Oleh karena itu, untuk memilih teknik statistik yang akan digunakan dalam pengujian hipotesis, pertama-tama perlu dipahami jenis data yang akan dianalisis dan jenis hipotesis yang diajukan.

Jenis data dalam penelitian ini meliputi data nominal, ordinal, interval, dan proporsional. Selain itu, bentuk hipotesis penelitian adalah deskriptif (hipotesis terhadap satu atau lebih variabel bebas), komparatif (perbandingan dua sampel atau k sampel) dan gabungan (hubungan dua variabel atau lebih). Dalam hipotesis komparatif terdapat sampel korelasional dan independen. Setelah tipe data dan asumsi terbentuk, yang tersisa hanyalah mendefinisikan teknik statistik yang digunakan. Statistik yang digunakan meliputi statistik parametrik dan *non* parametrik. Statistik parametrik digunakan untuk menganalisis sampel data yang besar, datanya berdistribusi normal dalam bentuk interval dan rasio, sedangkan statistik *non* parametrik digunakan untuk menganalisis sampel data kecil yang *non* parametrik, yang harus berdistribusi normal, serta data nominal dan ordinal.

Kami mengucapkan terima kasih kepada semua pihak yang telah mendukung dan berkontribusi selama seluruh proses penyusunan dan penerbitan buku ini sehingga kami dapat menyediakan kebutuhan pembaca dalam bidang statistik. Semoga buku ini bermanfaat bagi semua pihak dan dapat memberikan kontribusi bagi perkembangan ilmu pengetahuan di Indonesia.

Oktober, 2023

Tim Penulis

DAFTAR ISI

BAB

1

PENGERTIAN STATISTIKA, MACAM-MACAM STATISTIKA, PENGUKURAN, PENGGUNAAN NOTASI SIGMA, PENGUMPULAN, PENGOLAHAN SERTA PENYAJIAN DATA

A. PENGANTAR

Statistika memiliki peran yang cukup penting dalam kehidupan manusia serta pengembangan ilmu pengetahuan. Oleh karenanya tidak heran jika saat ini statistika telah berkembang dalam ranah ilmu yang berbeda-beda, misalnya Statistika Ekonomi, Statistika Bisnis, Statistika Pendidikan, Statistika Sosial, Statistika Kesehatan dan sebagainya.

B. PENGERTIAN STATISTIKA

Statistika berasal dari kata dasar statistik, sementara itu kata statistik secara etimologi berasal dari bahasa beberapa negara, misalnya dari kata *status* dalam bahasa latin, *state* dalam Bahasa Inggris serta *staat* dari Bahasa Belanda. Semua kata tersebut mengacu kepada kata "negara" dimana awalnya statistik dimaknai sebagai semua bentuk data dan informasi baik yang bersifat kualitatif maupun kuantitatif yang berguna bagi negara. Data-data yang berguna bagi negara tersebut misalnya adalah data jumlah penduduk, data nilai perdagangan baik ekspor maupun impor, data pengangguran dan kemiskinan serta data-data lainnya.

Berbagai pengertian statistik juga coba diungkapkan oleh berbagai ahli. Menurut Nuryadi *et al* (2017) statistik merupakan kumpulan fakta yang berisi angka-angka yang disusun dalam bentuk daftar atau tabel yang menggambarkan suatu persoalan. Sementara itu, Wahyuning (2021) mendefinisikan statistik sebagai alat yang digunakan untuk menyatakan ukuran sebagai wakil dari kumpulan data mengenai suatu hal. Lebih lanjut Sudjana dalam Sulisityowati dan Astuti (2017) menjelaskan statistik sebagai sebuah kumpulan data, bilangan maupun *non* bilangan yang disusun dalam bentuk tabel dan atau diagram yang melukiskan atau menggambarkan suatu persoalan. Hal yang sama juga dinyatakan oleh Wirawan (2016) dimana statistik adalah kumpulan keterangan yang berbentuk angka yang dapat memberikan informasi bagi pengguna, misalnya data jumlah kendaraan bermotor, data jumlah dosen dan jumlah mahasiswa, data jumlah nasabah suatu bank dan lain sebagainya.

Setelah memahami pengertian statistik maka selanjutnya akan dijelaskan mengenai pengertian statistika. Pada dasarnya statistika menunjukkan suatu pengetahuan yang berhubungan dengan cara-cara pengumpulan data, pengolahan, penganalisisan, dan penarikan kesimpulan serta pembuatan keputusan yang cukup beralasan berdasarkan hasil analisis data yang ada. Hal yang sama juga diungkapkan oleh Mason dalam Ananda dan Fadhli (2018) mendefinisikan statistika sebagai ilmu yang mempelajari bagaimana cara mengumpulkan data, penyusunan, penganalisisan dan penafsiran data dalam bentuk angka untuk tujuan pengambilan keputusan yang terbaik dalam sebuah ketidakpastian.

Sebagai sebuah ilmu, statistika memiliki beberapa ciri, diantaranya yaitu:

1. Statistika bekerja dengan angka-angka.

   Dalam penelitian yang menggunakan statistika sebagai alat analisisnya maka hasil penelitian hanya mungkin diperoleh jika tersedia data-data dalam bentuk angka. Dalam statistika, angka yang dimaksud ada dua pengertian yaitu angka sebagai jumlah dan angka sebagai suatu nilai. Angka sebagai suatu jumlah misalnya jumlah mahasiswa dan dosen di sebuah perguruan tinggi, jumlah buku yang ada di perpustakaan dan lain-lain. Angka sebagai sebuah nilai misalnya tinggi badan seorang mahasiswa, penghasilan seorang dosen dan lain-lain. Jika sebuah penelitian yang dilakukan tidak berkaitan dengan angka-angka namun ingin menggunakan metode statistik maka data tersebut harus diangkakan atau dikuantitatifkan terlebih dahulu, misalnya dalam penelitian yang berkaitan dengan jenis kelamin laki-laki dan perempuan maka dapat dilakukan penomoran perempuan dengan nilai 0 dan laki-laki dengan nilai 1. Dengan demikian yang awalnya data jenis kelamin adalah data kualitatif maka akan berubah menjadi data kuantitatif.

2. Statistika bersifat obyektif.
   Sebagai sebuah alat, statistika melakukan proses kegiatan mulai dari pengumpulan data sampai pada tahap publikasi data secara obyektif. Statistika bekerja apa adanya sesuai dengan data yang ada. Namun demikian, bisa saja hasil statistika digunakan untuk kepentingan tertentu yang tidak obyektif misalnya dalam kegiatan jajak pendapat pemilihan seorang pemimpin negara, tapi ini sudah di luar konteks statistika.
3. Statistika bersifat universal
   Karena statistik dapat digunakan dalam berbagai bidang penelitian maka statistika dianggap bersifat universal. Semua penelitian baik bidang eksakta maupun *non* eksakta (sosial humaniora) dapat menggunakan statistik sebagai alat analisisnya.

C. JENIS STATISTIKA

Sebagai sebuah ilmu, statistika terbagi dalam beberapa jenis. Jenis-jenis tersebut dikelompokkan menurut fungsi, ruang lingkup dan indikator yang diteliti.

Berdasarkan fungsinya, statistika dikelompokkan menjadi dua kategori yaitu (Hanafiah *et al,* 2020):

1. Statistik Deskriptif (*descriptive statistics*).
   Jenis statistik ini hanya menggambarkan atau menerangkan data tanpa melakukan pengujian data sehingga statistik deskriptif berfokus tentang bagaimana memperoleh, menyusun, menyajikan, menganalisis serta menginformasikannya kepada pihak-pihak yang berkepentingan.
   Contoh dari statistik deskriptif misalnya untuk mengetahui tingkat penghasilan *driver online* maka dikumpulkan data-data penghasilan mereka. Setelah data diperoleh maka data diolah, dianalisis, disajikan dan diinformasikan kepada pihak yang berkepentingan. Informasi yang diberikan misalnya terkait dengan penghasilan terendah, penghasilan tertinggi, penghasilan rata-rata dan sebagainya. Informasi bisa saja ditambahkan misalnya usia berapa yang

mendapatkan penghasilan yang lebih tinggi, pada waktu kapan saja mereka bisa mendapatkan penghasilan yang tinggi dan lain sebagainya.

2. Statistik Inferensial (*inferensial statistics*).

   Jika statistik deskriptif hanya berupa penggambaran dan penjelasan data-data maka dalam statistik inferensi ditandai dengan adanya tahapan pengujian data yang dimulai dari penentuan hipotesis sampai dengan penarikan kesimpulan. Kesimpulan yang diperoleh akan menjadi dasar pengambilan keputusan oleh pihak-pihak yang berkepentingan. Penarikan sampel dalam statistik inferensial bersifat general atau induksi sehingga statistik inferensial biasanya disebut juga dengan statistik induktif (*inductive statistics*). Selain penarikan kesimpulan, dalam statistik jenis ini juga dimungkinkan untuk memprediksi (meramal) ataupun mengestimasi (membuat penaksiran). Contoh statistik inferensi misalnya hendak menguji pernyataan seorang produsen mengenai daya tahan bola lampu buatan mereka. Untuk menguji pernyataan produsen tersebut maka akan dilakukan berbagai langkah yang dimulai dari penentuan hipotesis, melakukan pengujian sampel produk, menentukan nilai-nilai statistik sampai dengan penarikan kesimpulan. Kesimpulan akhir dari statistik inferensi adalah untuk menguji apakah pernyataan produsen bola lampu tersebut benar atau tidak. Setelah diketahui kesimpulannya maka dapat diambil keputusan yang tepat, misalnya bagi seorang konsumen jika pernyataan produsen tersebut tidak benar maka konsumen dapat saja memutuskan untuk tidak membeli produk bola lampu perusahaan tersebut dan memilih bola lampu produksi perusahaan lain.

Jika dilihat dari ruang lingkupnya, statistika dapat dibagi kedalam beberapa kelompok yaitu (Hamzah *et al,* 2016):

1. Statistik Pendidikan.

   Sesuai dengan namanya, statistik ini digunakan dalam bidang ilmu pendidikan. Penerapan statistik pendidikan misalnya pada saat dilakukan evaluasi hasil belajar siswa atau evaluasi kinerja guru. Untuk melakukan evaluasi hasil belajar siswa ataupun hasil kinerja guru tentu saja harus dikumpulkan data-data yang terkait dengan hal tersebut. Setelah data terkumpul kemudian diolah, dianalisis dan disajikan sehingga hasilnya dapat dijadikan sebagai tolok ukur dalam pengambilan kebijakan.

2. Statistik Sosial.
   Jenis statistik ini diterapkan pada bidang ilmu sosial yang dimanfaatkan untuk menganalisis gejala-gejala sosial yang terjadi di masyarakat. Salah satu contoh penerapan statistik sosial misalnya ketika ingin mengetahui dampak pandemi Covid-19 terhadap interaksi sosial di masyarakat. Pertanyaan yang mengemuka misalnya adalah apakah pandemi tersebut menyebabkan pergeseran pola interaksi masyarakat, misalnya masyarakat lebih senang berdiam di rumah daripada berkunjung dan mengobrol dengan tetangga. Untuk mengetahuinya maka dilakukan upaya penelitian dengan didukung alat analisis statistik. Karena diterapkan dalam ilmu sosial maka lebih dikenal dengan istilah statistik sosial.
3. Statistik Kesehatan.
   Statistik ini berkaitan erat dengan bidang kesehatan, misalnya dengan data sehat, sakit, kelahiran, kematian atau faktor-faktor lain yang terkait dengan populasi manusia.
4. Statistik Ekonomi.
   Jenis statistik ini tidak asing didengar karena bisa dirasakan setiap saat, baik dalam bidang ekonomi maupun bisnis. Statistik mengenai Indeks Harga Saham Gabungan (IHSG), Indeks Harga Konsumen (IHK) adalah contoh dari Statistik Ekonomi.
5. Statistik Pertanian.
   Jenis statistik ini diterapkan pada bidang atau disiplin ilmu Pertanian. Contoh dari statistik ini adalah yang rutin dilakukan oleh Badan Pusat Statistik (BPS) yaitu melakukan sensus pertanian serta statistik pertanian. Data-data yang dikumpulkan dalam sensus tersebut tentu akan menjadi potret sektor pertanian sehingga bisa dijadikan sebagai dasar dalam pengambilan kebijakan di sektor ini.
6. Statistik bidang ilmu/kajian lainnya.
   Selain bidang-bidang ilmu di atas, statistik juga diterapkan pada bidang ilmu lainnya misalnya di bidang fisika, antariksa dan lain sebagainya.

Jika dilihat dari indikator yang dianalisis, maka statistik dapat dikelompokkan dalam 2 kategori, yaitu:

1. Statistik parametrik.

   Data dalam statistik ini perlu memenuhi ketentuan tertentu, misalnya datanya dalam bentuk rasio atau interval, penentuan sampel dilakukan secara acak atau random, data terdistribusi secara normal, homogen dan lain sebagainya. Dengan demikian, yang dianalisis dalam statistik parametrik adalah parameter-parameter dari objek yang diteliti. Contoh statistik parametrik misalnya regresi, korelasi dan determinasi yang akan dipelajari pada bagian selanjutnya.
2. Statistik *non*parametrik.

   Jika statistik parametrik memerlukan persyaratan-persyaratan tertentu maka berbeda halnya dengan statistik *non*parametrik. Jenis statistik ini tidak membutuhkan persyaratan-persyaratan tersebut. Dalam statistik *non*parametrik, sisi lain dari parameter yang diteliti adalah indikator yang dianalisis. Contoh statistik parametrik misalnya *chi square test* dan uji tanda yang juga akan dijelaskan pada bagian berikutnya.

D. PENGUKURAN DATA

Dalam ilmu statistik, pengukuran data adalah hal yang penting untuk dilakukan. Menurut Pasaribu *et al* (2021) terdapat beberapa pengukuran data dimana jenis pengukuran data ini akan menjadi penentu metode atau rumus apa yang akan digunakan.

Sebelum dilakukan pengukuran terlebih dahulu harus ditentukan skala pengukurannya. Terdapat beberapa skala pengukuran yang dapat dijelaskan sebagai berikut:

1. Skala Nominal

   Skala ini merupakan skala yang paling sederhana yang tersusun menurut jenis tertentu. Skala ini dilakukan untuk membedakan antara satu karakteristik dengan karakteristik yang lain. Contoh skala nominal antara lain jenis kelamin (laki-laki dan perempuan), jenis mahasiswa (reguler, transfer atau yang lainnya), pendidikan terakhir (SMA, diploma, sarjana atau yang lainnya).
2. Skala Ordinal

   Skala ini mendasari pada perangkingan, misalnya hasil *test* penerimaan mahasiswa baru, jenjang kepangkatan Aparatur Sipil Negara (ASN), daftar orang terkaya di Indonesia dan lain sebagainya. Skala ordinal juga dapat digunakan dalam perangkingan preferensi seseorang misalnya dari sangat tidak setuju ke sangat setuju, sangat tidak puas sampai sangat puas, tidak pernah sampai sering sekali dan lain sebagainya.

3. Skala Interval

   Skala ini menunjukkan jarak antara satu data dengan data yang lain yang memiliki bobot yang sama, misalnya data pengalaman kerja dibuat interval kurang dari 5 tahun, 5-10 tahun, 10-15 tahun dan seterusnya.

4. Skala Rasio

   Kelebihan skala ini yaitu dapat dilakukan pada semua jenis perhitungan serta kesimpulan yang ditarik bisa lebih pasti. Selain itu, skala rasio juga memberikan kemudahan pada peneliti untuk menentukan perbedaan pada setiap variabel serta dalam hal pengurutan data. Perbedaan skala rasio dengan skala interval terletak pada nilainya. Jika skala interval dimungkinkan untuk nilai negatif maka skala rasio tidak ada nilai negatif. Contoh data rasio misalnya nilai ujian statistik mahasiswa di suatu kelas. Dalam kasus ini tentu saja tidak akan ada mahasiswa yang mendapatkan nilai negatif karena tentu saja dosen akan memberikan nilai yang terkecil adalah 0. Skala rasio memberikan hasil yang lebih rinci dibandingkan dengan skala lainnya.

E. PENGGUNAAN NOTASI SIGMA

Sigma dalam Bahasa sederhana dikatakan sebagai jumlah. Notasi sigma merupakan sebuah simbol untuk menjumlahkan bilangan berurut yang mengikuti suatu pola atau aturan tertentu. Rumus dengan notasi sigma adalah:

$\sum_{i=1}^{n} X_i$, dibaca sigma X_i , i dari 1 sampai *n.*

Aturan penjumlahan dengan notasi sigma adalah sebagai berikut (Wahyuning, 2021):

a. $\sum_{i=1}^{n}(X_i + Y_i + Z_i) = \sum_{i=1}^{n} X_i + \sum_{i=1}^{n} Y_i + \sum_{i=1}^{n} Z_i$

b. $\sum_{i=1}^{n} kX_i = k\sum_{i=1}^{n} X_i$, k = bilangan konstan

c. $\sum_{i=1}^{n} k = k + k + \cdots + k = nk$

d. $\sum_{i-1}^{n}(X_i - k)^2 = \sum_{i=1}^{n}(X_1^2 - 2kX_i + k^2)$

e. $\sum_{i-1}^{n}(Y_i - a - bX_i) = \sum_{i-1}^{n} Y_i - na - b\sum_{i-1}^{n} X_i$

F. PENGUMPULAN DATA

Terdapat beberapa cara yang digunakan untuk mengumpulkan data-data yang dibutuhkan dalam sebuah penelitian. Cara tersebut antara lain adalah sebagai berikut: (Wahyuning, 2021):

a. Wawancara

Wawancara merupakan teknik pengumpulan data dimana antara pencari data dengan penyedia data bertemu langsung atau bertatap muka. Namun demikian, dengan kemajuan teknologi wawancara bisa saja dilakukan meskipun jarak keduanya berjauhan karena wawancara dapat dilakukan melalui saluran telepon atau melalui saluran internet melalui berbagai fitur. Dalam wawancara biasanya peneliti telah menyiapkan daftar pertanyaan yang akan ditanyakan kepada sumber informasi. Hal ini penting untuk dilakukan agar wawancara dapat terarah dan sesuai dengan tujuan penelitian. Wawancara dapat dilakukan secara langsung maupun tidak langsung kepada sumber informasi. Dikatakan langsung jika data diperoleh dari sumbernya langsung, misalnya ketika ingin mengetahui kinerja karyawan maka ditanyakan langsung kepada karyawan yang bersangkutan, sedangkan dikatakan tidak langsung jika kinerja karyawan tersebut ditanyakan kepada atasan mereka.

b. Angket (kuesioner)

Angket atau kuesioner merupakan alternatif teknik pengumpulan data dimana pada cara ini peneliti telah menyiapkan daftar pertanyaan dan memberikan kesempatan kepada sumber informasi untuk menjawabnya sendiri. Angket atau kuesioner biasanya disebar di lokasi-lokasi dimana obyek penelitian berada, misalnya obyek penelitian adalah mahasiswa di suatu kota maka kuesioner disebar di kampus-kampus yang ada di kota tersebut. Namun demikian, seiring kemajuan zaman, saat ini lebih sering dijumpai penyebaran kuesioner secara *online* misalnya melalui *google form*. Cara ini dianggap lebih efektif karena selain menghemat waktu, tenaga dan biaya juga lebih efisien dalam pengolahan data.

c. Pengamatan (observasi)

Teknik observasi juga merupakan salah satu alternatif dalam mengumpulkan data. Pada teknik ini peneliti melakukan pengamatan terhadap obyek yang diteliti, baik secara langsung maupun tidak langsung. Penelitian-penelitian dalam bidang sejarah, budaya dan eksperimen biasanya banyak menggunakan teknik observasi dalam pengumpulan datanya, misalnya penelitian mengenai kehidupan sebuah suku yang tinggal di pedalaman, penelitian mengenai sejarah sebuah cagar budaya dan lain sebagainya.

G. PENGOLAHAN DATA

Data yang diperoleh baik melalui wawancara, kuesioner maupun observasi masih bersifat data mentah sehingga untuk dapat diinformasikan ke publik terlebih dahulu harus dilakukan pengolahan. Dengan demikian ada beberapa tujuan mengapa perlu dilakukan pengolahan data, yaitu: *pertama*, mengubah data mentah menjadi sebuah informasi; *kedua*, mempermudah pengambilan keputusan; *ketiga*, menyediakan data yang lebih valid dan akurat; *keempat*, mengurangi biaya dan waktu; *kelima*, membuat data mudah dikelola dan diolah kembali.

Terdapat dua tahapan yang biasanya dilakukan saat pengolahan data; *pertama*, melakukan proses tabulasi data, dimana data-data yang diperoleh dimasukkan ke dalam bentuk tabel. *Kedua*, melakukan pemeriksaan ulang apakah data yang sudah dimasukkan ke tabel sudah benar. Hal ini untuk mencegah kekeliruan data yang akan mengakibatkan hasil penelitian menjadi kurang baik (Pasaribu *et al,* 2021).

H. PENYAJIAN DATA

Data yang diperoleh dari sebuah hasil penelitian belum mampu "berbicara" jika tidak disajikan dengan tepat. Oleh karena itu, setelah dilakukan pengolahan data sebagaimana yang dijelaskan pada bagian sebelumnya maka data harus disajikan dengan baik, tepat, ringkas, menarik dan tentu saja harus mudah dipahami bukan saja oleh peneliti tetapi juga pihak-pihak lain.

Secara garis besar terdapat dua bentuk penyajian data yaitu penyajian dalam bentuk tabel dan penyajian dalam bentuk grafik atau diagram yang disesuaikan dengan peruntukannya. Pada hakikatnya tabel adalah kumpulan angka yang disusun sedemikian rupa berdasarkan kelompok/klasifikasi tertentu yang akan memudahkan saat proses analisa data. Sementara itu, grafik (diagram) merupakan visualisasi berupa gambar yang menampilkan angka. Penyajian dalam bentuk diagram biasanya lebih menarik dan lebih mudah dipahami karena selain memainkan warna-warna juga dapat ditampilkan dalam berbagai bentuk.

1. Tabel

Tabel merupakan teknik menyajikan data dimana angka-angka disusun sedemikian rupa dalam suatu baris dan kolom menurut kelompok datanya. Misalnya, jumlah mahasiswa yang dikelompokkan menurut usia, asal daerah dan pekerjaan orang tua. Demikian juga pegawai yang dikelompokkan berdasarkan jenis jabatan, masa kerja dan jumlah tanggungan.

Berdasarkan cara penyajiannya, terdapat dua jenis tabel, yaitu tabel referensi dan tabel ikhtisar. Tabel referensi menyediakan data yang lebih lengkap dan rinci sehingga dapat menjadi bahan referensi bagi pihak-pihak yang memerlukan. Tabel-tabel yang ada di laporan BPS misalnya data survei atau data sensus adalah contoh dari tabel referensi. Hal ini dikarenakan penyajian tabel dalam laporan-laporan tersebut sangat lengkap dan rinci dan sangat memungkinkan untuk diolah kembali.

Jika pada tabel referensi data-data disajikan secara lengkap dan rinci maka pada tabel ikhtisar penyajiannya lebih singkat, sederhana dan lebih mudah dimengerti. Tabel ikhtisar biasanya banyak disajikan dalam laporan-laporan perusahaan ataupun laporan-laporan penelitian. Misalnya laporan akhir tahun perusahaan yang menampilkan data perkembangan laba per bulan pada tahun tersebut dan membandingkannya dengan laba tahun sebelumnya.

Setelah dilakukan penyusunan tabel maka selanjutnya dilakukan penyajian tabel yang dapat diklasifikasikan ke dalam 4 bentuk tabel:

1) *Tabel Klasifikasi Acak/Tunggal*

Berikut disampaikan contoh tabel klasifikasi acak dengan mengambil contoh data dari BPS.

Tabel 1.1. Jumlah Penduduk Miskin Menurut Wilayah (Juta Jiwa) Tahun 2022

Wilayah	Jumlah
Kota	**11,98**
Desa	**14,38**
Kota+ Desa	**26,36**

Sumber: BPS (2023)

Tabel di atas memperlihatkan bahwa penyajian data dilakukan secara tunggal dan tidak berjenjang.

2) *Tabel Klasifikasi Berkelompok*

Berikut disampaikan contoh tabel klasifikasi berkelompok dengan mengambil contoh data hipotetis.

Tabel 1.2. Jumlah Dosen Perguruan Tinggi "A" Menurut Jenis Kelamin dan Jenjang Akademik Tahun 2022

Jenis Kelamin	Jenjang				Jumlah
	Non Jabatan	Asisten Ahli	Lektor	Lektor Kepala	
Laki-Laki	3	15	23	12	53
Perempuan	8	20	25	18	71
Jumlah	**11**	**35**	**48**	**30**	**124**

Sumber: Data Hipotetis

Tabel di atas memperlihatkan bahwa penyajian data dilakukan secara berkelompok dengan memberikan uraian pada data utama.

3) *Tabel Kontingensi*

Berikut disampaikan contoh tabel kontingensi dengan mengambil contoh data hipotetis.

Tabel 1.3. Hubungan Pekerjaan dengan Olahraga Yang Diminati

Pekerjaan	Jenis Olahraga			Jumlah
	Bulutangkis	Atletik	Renang	
TNI/Polri	23	25	36	84
Dosen	56	10	9	75
Dokter	20	14	32	66
Karyawan Swasta	46	18	8	72
Jumlah	**145**	**67**	**85**	**297**

Sumber: Data Hipotetis

Tabel di atas memperlihatkan bahwa penyajian data dilakukan dengan membuat perbandingan antar beberapa kelompok data.

4) *Tabel Frekuensi/Distribusi Frekuensi*

Berikut disampaikan contoh tabel frekuensi/distribusi frekuensi dengan mengambil contoh data hipotetis.

Tabel 1.4. Data Nilai Ujian Statistik Mahasiswa

Nilai Ujian	Frekuensi
1 – 20	15
21 – 40	20
41 – 60	23
61 – 80	10
81 – 100	5
Jumlah	**73**

Sumber: Data Hipotetis

Tabel di atas memperlihatkan bahwa penyajian data dilakukan dengan memberikan interval pada data utama yang disertai dengan frekuensi di setiap kelas atau barisnya.

2. Diagram/Grafik

Diagram/grafik merupakan bentuk penyajian data dalam bentuk lukisan garis, gambar maupun lambang tertentu. Penyajian data akan lebih menarik dan mudah dipahami jika disajikan dalam bentuk grafik/diagram, misalnya dalam menyajikan data 5 besar perguruan tinggi terbaik di Indonesia tentu akan lebih menarik jika disajikan dalam bentuk diagram/grafik daripada disajikan dalam bentuk tabel. Demikian juga misalnya daftar produk terlaris dari sebuah gerai makanan tentu akan lebih menarik jika disajikan dalam bentuk diagram/grafik.

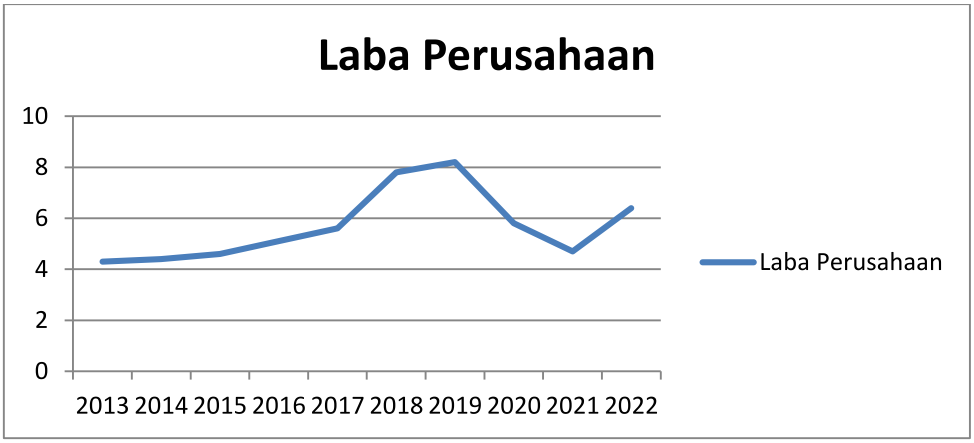
Namun demikian, biasanya data yang disajikan dalam diagram/grafik bersifat aprosimatif karena angkanya tidak serinci dengan angka yang ditampilkan pada tabel. Oleh karena itu, biasanya diagram/grafik beriringan dengan tabel dimana fungsi diagram adalah mendukung tabel yang sudah disajikan terlebih dahulu.

Diagram/grafik mempunyai beberapa keunggulan, diantaranya lebih menarik, lebih cepat dipahami, dan mampu menggambarkan data secara umum dan menyeluruh. Sebaliknya, diagram/grafik juga memiliki kekurangan, diantaranya yaitu lebih sukar membuatnya dan membutuhkan waktu yang lebih lama, penyajian terbatas karena jika terlalu banyak gambar maka akan membuat bingung dan sulit dimengerti.

Terdapat banyak bentuk penyajian diagram, namun yang paling sering digunakan adalah sebagai berikut (Siagian, 2021):

1) *Diagram Garis (Polygon)*

Berikut disampaikan contoh Diagram Garis dengan menggunakan contoh data hipotetis.

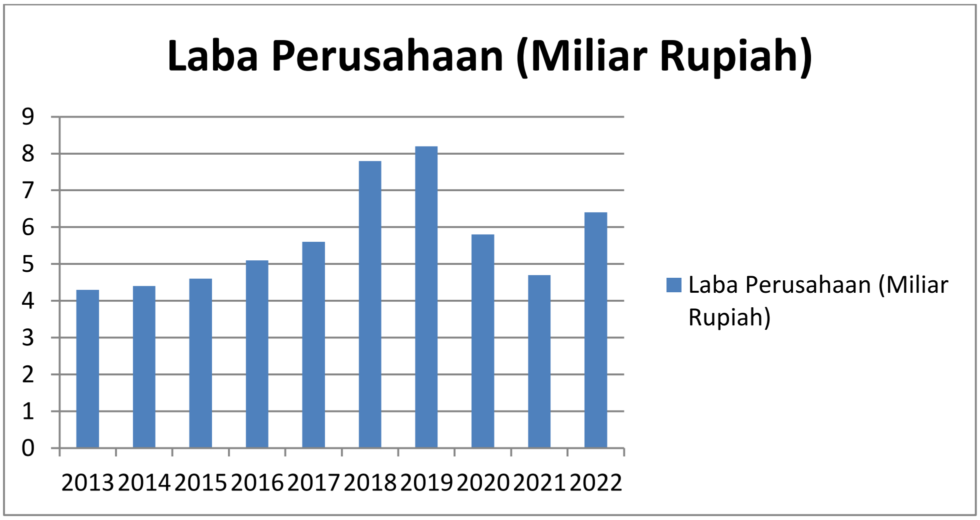


Gambar 1.1. Perkembangan laba Perusahaan X Tahun 2013-2022. *Sumber: Data Hipotetis*

Dalam diagram garis bisa dimungkinkan untuk menyajikan beberapa data sekaligus yang memiliki relevansi antar data, misalnya selain menyajikan data laba juga disajikan data deviden setiap tahunnya. Untuk membedakannya maka masing-masing garis dibedakan warnanya dan diberikan keterangan.

2) *Diagram Batang (Histogram)*

Berikut disampaikan contoh Diagram Batang dengan menggunakan contoh data hipotetis.

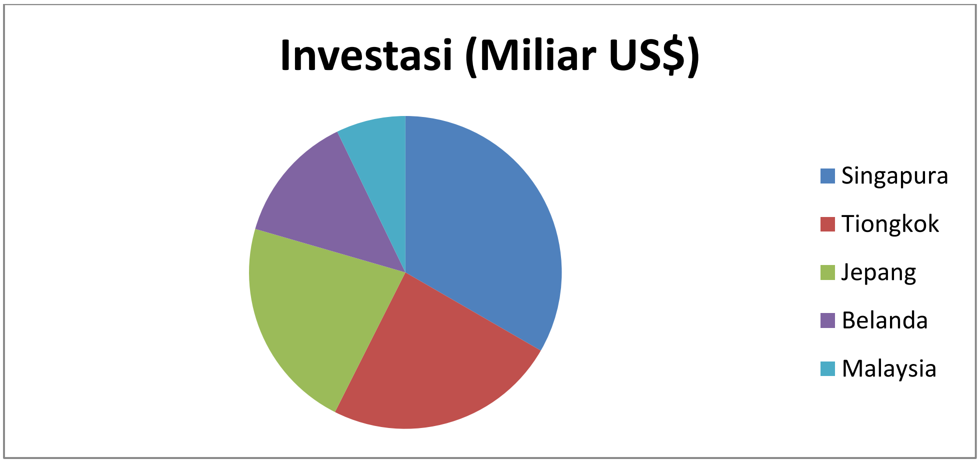


Gambar 1.2. Perkembangan laba Perusahaan X Tahun 2013-2022. *Sumber: Data Hipotetis*

Sama halnya dengan diagram garis, pada diagram batang juga bisa dimungkinkan untuk menyajikan beberapa data sekaligus yang memiliki relevansi antar data, misalnya selain menyajikan data laba juga disajikan data deviden setiap tahunnya. Untuk membedakannya maka masing-masing diagram dibedakan warnanya dan diberikan keterangan.

3) *Diagram Lingkaran (Pie Diagram)*

Berikut disampaikan contoh diagram lingkaran dengan menggunakan contoh data dari Bappenas.

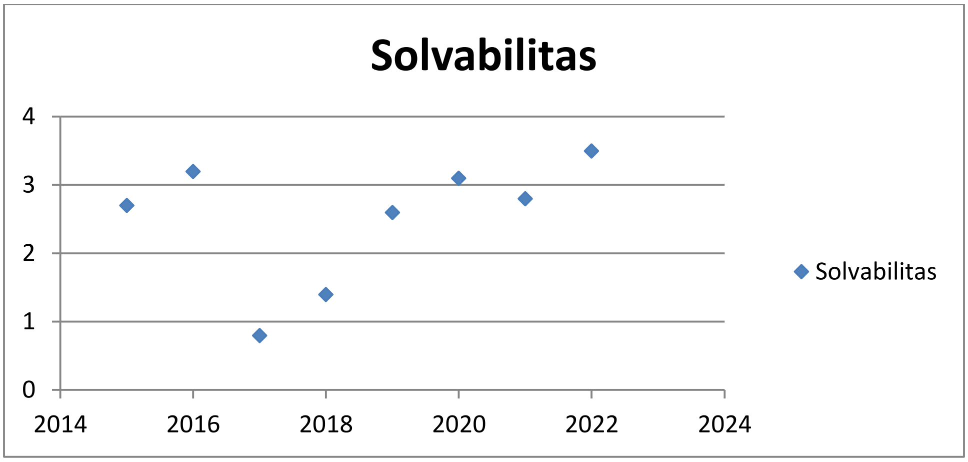


Gambar 1.3. Daftar 5 Besar Investor di Indonesia Tahun 2019. *Sumber: Bappenas (2023)*

Diagram lingkaran biasanya digunakan untuk menyajikan sebuah nilai komposisi suatu obyek yang dapat dimaknai sebagai perbandingan antar komposisi.

4) *Diagram Pencar*

Berikut disampaikan contoh diagram pencar dengan menggunakan contoh data hipotetis.



Gambar 1.4. Tingkat Solvabilitas Perusahaan XYZ.
Sumber: Data Hipotetis

Diagram pencar terlihat seperti sekumpulan titik-titik yang memberikan informasi tertentu sesuai dengan obyek yang sedang diteliti.

I. PENUTUP

Statistika telah digunakan pada berbagai bidang ilmu sebagai alat analisis. Data-data yang diperoleh selama penelitian akan menjadi informasi yang baik dan berkualitas bagi khalayak jika data-data tersebut dikumpulkan, diolah, dianalisis dan disajikan dengan tepat.

BAB

2

DISTRIBUSI FREKUENSI

Data-data statistik dan juga data hasil penelitian yang dikumpulkan sering digunakan sebagai informasi baik untuk laporan maupun analisis suatu penelitian. Untuk itu hendaknya data ini disajikan dalam bentuk tabel dan juga grafik sehingga lebih mudah untuk dipahami.

Data yang diperoleh dari hasil pengukuran umumnya berupa angka-angka ataupun skor-skor penilaian yang umumnya dikenal dengan data mentah (*raw data*) atau skor mentah (*raw score*). Sebagai contoh hasil suatu pengukuran diperoleh data sebagai berikut:

23	20	16	18	30	22	26	15	13	18	14	16
10	20	22	23	17	16	16	26	29	28	17	24
17	16	15	28	24	17	9	17	23	25	20	11
10	24	18	27	28	15	19	20	12	21	37	20
24	19	22	19	18	18	21	19	22	21	22	35

Sumber: Data fiktif

Apabila data disajikan seperti di atas, apakah kita mampu untuk memperoleh gambaran ataupun makna dari data tersebut? Penyajian data seperti ini sangatlah tidak menarik, lebih dari itu jika tampilan data seperti ini maka akan sangat sulit untuk dibaca. Andaikata seseorang mampu untuk menafsirkan data tersebut, maka tafsiran yang diberikan belum tentu sesuai dengan maksud yang sebenarnya karena banyak tafsiran yang dapat dijabarkan dari data tersebut.

Ada berbagai cara untuk dapat menyajikan data hasil pengukuran, penyajian data dapat disesuaikan dengan jenis data serta tujuan dari penyajiannya. Salah satu bentuk penyajian data yang mudah untuk dipahami adalah dalam bentuk distribusi frekuensi.

A. PENGERTIAN DISTRIBUSI FREKUENSI

Definisi distribusi frekuensi menurut Dr. Riduwan (2020) adalah penyusunan suatu data mulai dari terkecil sampai terbesar yang membagi banyaknya data ke dalam beberapa kelas. Sedikit berbeda dengan definisi di atas Prof. Furqon (2014) mengartikan distribusi frekuensi sebagai rincian nilai dari suatu *instrument* data bersama frekuensinya masing-masing didalam suatu pengukuran. Lain halnya dengan Sofyan Siregar (2015) yang mengatakan bahwa distribusi frekuensi adalah penyusunan suatu data mulai dari yang terkecil sampai yang terbesar yang membagi data ke dalam beberapa kelas. Dari beberapa definisi ini maka dapat katakan bahwa distribusi frekuensi adalah penyusunan data terkecil hingga terbesar beserta frekuensinya dalam suatu tabel.

Penyusunan data dalam tabel memudahkan kita untuk mengetahui skor terendah, skor tertinggi, dan juga perhitungan-perhitungan tendensi sentral yang lain seperti *mean*, median, modus, persentil, range, dan sebagainya. Selain itu juga penyusunan data dalam distribusi frekuensi memudahkan kita untuk membuat gambar *statistic* dalam berbagai bentuk penyajian data.

B. TABEL DISTRIBUSI FREKUENSI

Penggolongan distribusi frekuensi dalam kelas umumnya dapat dilakukan atas dasar kategori atau sifat dari data yang diteliti. Menurut jenis datanya terdiri atas distribusi frekuensi kualitatif/kategori dan distribusi frekuensi kuantitatif.

1. Tabel Distribusi Kualitatif/Kategori

Distribusi frekuensi kualitatif atau dikenal dengan distribusi frekuensi kategori adalah distribusi frekuensi yang data-datanya dikelompokkan dan disusun berdasarkan pada data kategori.

Contoh distribusi frekuensi kategori

Tabel 2.1 Distribusi Frekuensi Penggolongan Status Dalam Keluarga

Kategori	Frekuensi
Anak-anak	30
Remaja	35
Menikah	25
Cerai	10
Jumlah	100

Sumber: Data Fiktif

Tabel 2.2 Impor 5 Tahun Terakhir Menurut Golongan Penggunaan Barang (dalam ton)

Tahun	Frekuensi		
	Barang Konsumsi	Bahan Baku	Barang Modal
2018	7398,9	160 217,1	4004,4
2019	5255,6	153275,3	4097,8
2020	5204,9	142818,4	3856,7
2021	5878,8	167768,5	4112,0
2022	6465,0	171913,0	4875,5

Sumber: BPS Indonesia Tahun 2023

Tabel 2.1 adalah contoh daftar distribusi frekuensi kualitatif sederhana. Pada kolom pertama kolom kategori penggolongan tidak didasarkan atas besarnya angka atau bilangan, atau sebagai kelas interval melainkan penggolongan status dalam keluarga. Begitupun untuk Tabel 2.2. yang kolom pertama merupakan kolom tahun untuk pengisian banyaknya jumlah barang konsumsi, bahan baku dan barang modal yang diimpor.

2. Tabel Distribusi Frekuensi Kuantitatif

Tabel distribusi frekuensi kuantitatif dapat dibedakan menjadi tabel distribusi frekuensi kuantitatif tunggal, yang dapat dibedakan menjadi tunggal dengan frekuensi dan tunggal/tanpa frekuensi.

Dalam distribusi frekuensi tunggal, data yang diperoleh diurutkan terlebih dahulu dari nilai terendah hingga nilai tertinggi. Untuk distribusi frekuensi tunggal dengan frekuensi, data yang diperoleh diurutkan kemudian dihitung berdasarkan jumlah frekuensi data tersebut. Umumnya penggunaan distribusi ini digunakan untuk jumlah data yang sedikit sehingga masih mudah untuk dibentuk dalam tabel distribusi frekuensi tunggal. Berikut contoh penyajian data pengukuran kadar nikotin (dalam μg) dari 20 batang rokok Hijau dengan nilai sebagai berikut: (data fiktif)

21	17	15	8	19	16	17	18	19	20
17	17	11	13	15	14	16	19	18	19

Dari data yang diberikan kita belum memperoleh gambaran, sehingga data ini perlu dibuat dalam distribusi frekuensi agar data tersebut memiliki arti.

Tabel 2.3 Hasil Pengukuran Kadar Nikotin (dalam μg) dari 20 batang rokok Hijau

Kadar nikotin (dalam μg)	**Frekuensi**
8	1
11	1
13	1
14	1
15	2
16	2
17	4
18	2
19	4
20	1
21	1
Jumlah	20

Sumber: Data fiktif

Apabila data telah disajikan seperti Tabel 2.3 maka data tersebut dapat memberikan gambaran bahwa nilai kadar nikotin tertinggi pada rokok Hijau adalah 21 μg, kadar nikotin terendah adalah 8 μg, kadar nikotin yang paling banyak frekuensinya adalah 17 μg. Terbentuknya tabel ini juga akan mempermudah perhitungan-perhitungan tendensi sentral serta perhitungan lainnya.

3. Tabel Distribusi Frekuensi Bergolong

Tabel distribusi frekuensi bergolong digunakan apabila data-data yang akan disajikan cukup banyak, sehingga apabila penyajian menggunakan tabel distribusi tunggal berfrekuensi akan tidak efisien. Bentuk tabel ini akan lebih menguntungkan apabila data akan digunakan untuk pengujian normalitas data yang menggunakan kertas peluang normal.

Tabel 2.4 Distribusi Frekuensi Usia Penduduk RT. 01 Desa XYZ

Umur	**Frekuensi**
6 – 10	1
11 – 15	1
16 – 20	8
21 – 25	16
26 – 30	24
31 – 35	3
36 – 40	7
Jumlah	60

Sumber: Data pribadi

Dalam tabel distribusi frekuensi bergolong di atas terdapat istilah-istilah yang digunakan dalam distribusi frekuensi kelompok seperti:

1) Kelas Interval (*Class Interval*) merupakan sejumlah nilai yang terdapat dalam batas kelas tertentu, atau jarak antara nilai batas bawah dengan batas atas pada setiap kelas. Interval nilai bawah dengan atas sering disebut dengan panjang kelas.

   Tabel 2.4 berisikan 7 interval kelas, dimana 6-10 merupakan interval kelas pertama, dan 36 – 40 merupakan interval kelas ke tujuh. Batas bawah dari tabel 2.4 adalah 6, 11, 16, 21, 26, 31, dan 36 sedangkan batas atas ditunjukkan dengan nilai 10, 15, 20, 25, 30, 35, dan 40
2) Setiap interval kelas memiliki frekuensi
3) Batas kelas atau tepi kelas adalah bilangan akhir dari suatu interval kelas. Batas kelas ini berguna pada saat penyajian data dalam bentuk histogram. Pada interval kelas yang pertama yaitu 6 – 10, nilai ujung atas adalah 10 dan nilai ujung bawah adalah 6. Batas kelas akan diperoleh dengan menambahkan ujung kelas atas interval yang pertama dan ujung kelas bawah interval yang kedua kemudian dikalikan dengan ½. Secara aljabar dapat ditulis:

Batas kelas = $(ujung\ bawah\ kelas\ interval\ 1\ + ujung\ bawah\ kelas\ interval\ 2)\ x\ \frac{1}{2}$ (1)

4) Titik tengah (*class-mark/class-midpoint)* adalah titik tengah yang berada di tengah interval kelas. Pengertian lain dari titik tengah adalah hasil penjumlahan nilai ujung bawah kelas dan nilai ujung atas kelas dan dikalikan ½ (setengah). Secara aljabar dapat ditulis:

   Titik tengah = $(batas\ bawah\ + batas\ atas)\ x\ \frac{1}{2}$ (2)

   Berdasarkan tabel 2.4 maka:

   titik tengah kelas pertama = $(6 + 10)\ x\ \frac{1}{2} = 8$

   titik tengah kelas kedua = $(11 + 15)\ x\ \frac{1}{2} = 13$, dst

Teknik Pembuatan Distribusi Frekuensi

Langkah-Langkah dalam Pembuatan Tabel Distribusi Frekuensi Bergolong

1. Menghitung rentangan data (R)

   Rentangan data diperoleh dari selisih antara data tertinggi/terbesar dengan data yang terendah/terkecil.

   Rumus: R = data tertinggi – data terendah (3)
2. Menghitung jumlah kelas (K)

   Sebagai patokan dalam mencari jumlah kelas digunakan rumus **Sturges**, yaitu:

   K = 1 + 3,3 log n (4)

   Dimana:

   K = Jumlah Kelas

   n = banyak data,
3. Menghitung panjang interval kelas (P)

   Rumus $P = \frac{Rentangan\ data\ (R)}{Jumlah\ Kelas\ (K)}$ (5)
4. Menentukan batas kelas interval panjang kelas
5. Setelah jumlah kelas dan panjang interval kelas didapat, kemudian membuat tabel dan menghitung data dengan dibuat turus (penghitungan) untuk memasukan masing-masing data pengamatan dalam kelas interval yang telah dibuat.

Contoh

Diketahui hasil pengukuran kadar logam dalam sampel air sumur dusun XYZ tahun 2008 diperoleh data

23	20	16	18	30	22	26	15	13	18
20	23	24	17	18	16	27	16	28	26
23	24	15	24	28	19	24	22	17	19
19	11	23	17	23	13	17	26	26	14
15	**8**	15	14	17	29	18	16	27	17
19	11	**37**	28	18	18	21	23	20	21
16	10	12	25	33	20	19	24	20	22
21	20	21	27	22	22	25	19	20	21
18	22	21	15	19	19	18	16	17	20
19	22	21	14	17	21	24	19	16	20

Jumlah data pengukuran (n) = 100

1) Menghitung rentangan data (R)

   R = data tertinggi – data terendah

   R = 37 – 8 = 29

2) Menghitung jumlah kelas (K)

   K = 1 + 3,3 log n

   K = 1 + 3,3 log 100

   K = 1 + 3,3 . 2

   K = 1 + 6,6 = 7,6 ≈ 8

3) Menghitung panjang interval kelas (P)

   $$P = \frac{Rentangan\ data\ (R)}{Jumlah\ Kelas\ (K)}$$

   $$P = \frac{29}{8} = 3{,}6\ \approx 4$$

4) Menentukan batas kelas interval panjang kelas (P)

   (**8** + 4) = 12 – 1 = **11**

   (**12** + 4) = 16 – 1 = **15**

   (**16** + 4) = 20 – 1 = **19**

   (**20** + 4) = 24 – 1 = **23**

   (**24** + 4) = 28 – 1 = **27**

   (**28** + 4) = 32 – 1 = **31**

   (**32** + 4) = 36 – 1 = **35**

(**36** + 4) = 40 – 1 = **39**

5) Membuat tabel distribusi frekuensi sementara:

Tabel 2.5 Distribusi Frekuensi

Kadar Logam Dalam Sampel Air Sumur Dusun XYZ tahun 2008

Kelas Interval	Turus	Frekuensi
8 – 11	||||	4
12 – 15	~~||||~~ ~~||||~~ |	11
16 – 19	~~||||~~ ~~||||~~ ~~||||~~ ~~||||~~ ~~||||~~ ~~||||~~ |||	33
20 – 23	~~||||~~ ~~||||~~ ~~||||~~ ~~||||~~ ~~||||~~ ~~||||~~	30
24 – 27	~~||||~~ ~~||||~~ ~~||||~~	15
28 – 31	~~||||~~	5
32 – 35	|	1
36 - 39	|	1
	Jumlah	100

6) Membuat tabel distribusi frekuensi

Tabel 2.6 Distribusi Frekuensi

Kadar Logam Dalam Sampel Air Sumur Dusun XYZ tahun 2008

Kelas Interval	Frekuensi
8 – 11	4
12 – 15	11
16 – 19	33
20 – 23	30
24 – 27	15
28 – 31	5
32 – 35	1
36 - 39	1
Jumlah	100

C. DISTRIBUSI FREKUENSI LAINNYA

1. Distribusi Frekuensi Relatif

Distribusi frekuensi relatif ialah distribusi frekuensi yang nilai frekuensinya dinyatakan dalam bentuk presentase (%). Data akan akan lebih mudah dipahami bila dinyatakan dalam bentuk persen. Cara perhitungannya distribusi frekuensi *relative* dapat menggunakan rumus:

$$f_{relatif\ kelas\ ke-i} = \frac{f_{(mutlak)kelas-i}}{n}\ x\ 100\% \qquad (6)$$

Dengan menggunakan data tabel 2.6 maka dapat dihitung frekuensi relatif masing-masing kelas

$$f_{relatif\ kelas\ 1} = \frac{4}{100}\ x\ 100\% = 4\%$$
$$f_{relatif\ kelas\ 2} = \frac{11}{100}\ x\ 100\% = 11\%$$
$$f_{relatif\ kelas\ 3} = \frac{33}{100}\ x\ 100\% = 33\%$$
$$f_{relatif\ kelas\ 4} = \frac{30}{100}\ x\ 100\% = 30\%$$
$$f_{relatif\ kelas\ 5} = \frac{15}{100}\ x\ 100\% = 15\%$$
$$f_{relatif\ kelas\ 6} = \frac{5}{100}\ x\ 100\% = 5\%$$
$$f_{relatif\ kelas\ 7} = \frac{1}{100}\ x\ 100\% = 1\%$$
$$f_{relatif\ kelas\ 8} = \frac{1}{100}\ x\ 100\% = 1\%$$

Hasil perhitungan ini kemudian dimasukkan ke dalam tabel distribusi frekuensi relatif (Tabel 2.7) atau dapat digabungkan dengan tabel distribusi frekuensi (Tabel 2.8).

Tabel 2.7 Distribusi Frekuensi Relatif
Kadar Logam Dalam Sampel Air Sumur Dusun XYZ tahun 2008

Kelas Interval	Frekuensi (f)
8 – 11	4%
12 – 15	11%
16 – 19	33%
20 – 23	30%
24 – 27	15%

28 – 31	5%
32 – 35	1%
36 - 39	1%
Jumlah	100%

Tabel 2.8 Distribusi Frekuensi Relatif

Kadar Logam Dalam Sampel Air Sumur Dusun XYZ tahun 2008

Kelas Interval	Frekuensi	Frekuensi Felatif
8 – 11	4	4 %
12 – 15	11	11 %
16 – 19	33	33 %
20 – 23	30	30 %
24 – 27	15	15 %
28 – 31	5	5 %
32 – 35	1	1 %
36 - 39	1	1 %
Jumlah	100	100%

2. Distribusi Frekuensi Kumulatif

Distribusi frekuensi kumulatif disingkat dengan (F_{kum}) merupakan distribusi frekuensi yang nilai frekuensinya (f) didapat dari hasil penjumlahan frekuensi dari setiap kelas interval.

Tabel distribusi frekuensi kumulatif didasarkan pada tabel distribusi frekuensi. Tabel distribusi frekuensi kumulatif dibentuk dengan tujuan untuk menunjukkan frekuensi kumulatif "kurang dari" dan frekuensi kumulatif "lebih dari". Untuk memulai distribusi frekuensi kumulatif kurang dari dan lebih dari dimulai dari nilai ujung bawah dari setiap kelas interval. Data dalam tabel 2.4 distribusi frekuensi bergolong usia penduduk RT. 01 Desa XYZ dapat kita gunakan sebagai contoh pembuatan tabel distribusi frekuensi kumulatif "kurang dari" dan "lebih dari".

Tabel 2.9 Distribusi Frekuensi "kurang dari"

Usia Penduduk RT. 01 Desa XYZ

Usia Penduduk	Frek. Kumulatif mutlak
Kurang dari 6	0
Kurang dari 10	1
Kurang dari 15	2
Kurang dari 20	10
Kurang dari 25	26

Kurang dari 30	50
Kurang dari 35	53
Kurang dari 40	60

Tabel 2.10 Distribusi Frekuensi "lebih dari"
Usia Penduduk RT. 01 Desa XYZ

Usia Penduduk	Frek. Kumulatif mutlak
Lebih dari 6	60
Lebih dari 10	59
Lebih dari 15	58
Lebih dari 20	50
Lebih dari 25	34
Lebih dari 30	10
Lebih dari 35	7
Lebih dari 40	0

3. Distribusi Frekuensi Relatif Kumulatif

Distribusi frekuensi relatif kumulatif ($f_{kum\,(\%)}$) adalah tabel yang menggambarkan suatu keadaan dalam frekuensi kumulatif relatif yaitu frekuensi kumulatif dari persentase frekuensi terhadap total banyak pengamatan.

$$f_{kum(\%)\,ke-i} = \frac{f_{(kum)kelas-i}}{n}\ x\ 100\% \qquad (7)$$

Berdasarkan tabel 2.9 maka dapat dihitung distribusi frekuensi kumulatif relatif kurang dari:

$f_{kum(\%)} = \frac{0}{60}\ x\ 100\% = 0{,}00\%$

$f_{kum(\%)} = \frac{1}{60}\ x\ 100\% = 1{,}67\%$

$f_{kum(\%)} = \frac{2}{60}\ x\ 100\% = 3{,}33\%$

$f_{kum(\%)} = \frac{10}{60}\ x\ 100\% = 16{,}67\%$

$f_{kum(\%)} = \frac{26}{60}\ x\ 100\% = 43{,}33\%$

$f_{kum(\%)} = \frac{50}{60}\ x\ 100\% = 83{,}33\%$

$f_{kum(\%)} = \frac{53}{60}\ x\ 100\% = 88{,}33\%$

$f_{kum(\%)} = \frac{60}{60}\ x\ 100\% = 100\%$

Hasil perhitungan ini kemudian dimasukkan ke dalam tabel distribusi frekuensi kumulatif kurang dari seperti yang terlihat dalam tabel 2.11

Tabel 2.11 Distribusi Frekuensi Kumulatif Relatif "kurang dari"
Usia Penduduk RT. 01 Desa XYZ

Usia Penduduk	Frek. Kumulatif mutlak
Kurang dari 6	0,00%
Kurang dari 10	1,67%
Kurang dari 15	3,33%
Kurang dari 20	16,67%
Kurang dari 25	43,33%
Kurang dari 30	83,33%
Kurang dari 35	88,33%
Kurang dari 40	100%

Berdasarkan tabel 2.10 maka dapat dihitung distribusi frekuensi kumulatif relatif lebih dari:

$f_{kum(\%)} = \frac{60}{60} x$ 100% = 100%

$f_{kum(\%)} = \frac{59}{60} x$ 100% = 98,33%

$f_{kum(\%)} = \frac{58}{60} x$ 100% = 96,67%

$f_{kum(\%)} = \frac{50}{60} x$ 100% = 83,33%

$f_{kum(\%)} = \frac{34}{60} x$ 100% = 56,67%

$f_{kum(\%)} = \frac{10}{60} x$ 100% = 16,67%

$f_{kum(\%)} = \frac{7}{60} x$ 100% = 11,67%

$f_{kum(\%)} = \frac{0}{60} x$ 100% = 0,00%

Hasil perhitungan di ini kemudian dimasukkan ke dalam tabel distribusi frekuensi kumulatif lebih dari seperti pada tabel 2.12

Tabel 2.12 Distribusi Frekuensi Kumulatif Relatif "lebih dari"
Usia Penduduk RT. 01 Desa XYZ

Usia Penduduk	Frek. Kumulatif mutlak
Lebih dari 6	100%
Lebih dari 10	98,33%
Lebih dari 15	96,67%
Lebih dari 20	83,33%

Lebih dari 25	56,67%
Lebih dari 30	16,67%
Lebih dari 35	11,67%
Lebih dari 40	0,00%

BAB

3

UKURAN TENDENSI SENTRAL

Pengamatan sehari-hari menunjukkan bahwa setiap orang tidak menunjukkan kesamaan dalam sesuatu hal. Kecerdasan, tinggi badan, berat badan, penghasilan dan sebagainya, bagi setiap orang kebanyakan tidaklah sama. Salah satu tugas statistik adalah mencari suatu angka di sekitar mana nilai-nilai dalam suatu distribusi memusat. Angka yang menjadi pusat sesuatu distribusi disebut tendensi sentral.

Beberapa ukuran tendensi sentral yang sangat penting untuk dibicarakan adalah *Mean*, Median, Kuartil, Desil, Persentil dan Modus. Kesemuanya mempunyai cara-cara menghitung yang berbeda-beda, dan mempunyai arti yang berbeda pula sebagai alat untuk mengadakan deskripsi sesuatu distribusi.

A. *MEAN* (RATA-RATA)

a. Rata-rata Hitung

Rata-rata hitung adalah ukuran pemusatan lokasi yang banyak digunakan dalam statistika. Ukuran ini mudah dihitung dengan memanfaatkan semua data yang dimiliki. Jika ada sekelompok data, maka untuk menyebut ukuran numerik sebagai wakil dari data sering dipakai nilai rata-rata (hitung) baik terhadap populasi maupun terhadap sampel. Rata-rata hitung (sering disebut rata-rata saja) dapat ditentukan dengan cara membagi jumlah nilai data oleh banyaknya data. Bila sekelompok data X_1, X_2,…,X_n (tidak harus semuanya berbeda) menyusun sebuah populasi terhingga berukuran N, maka rata-rata populasi adalah

$$\mu = \frac{x_1 + x_2 + \cdots + Xn}{N}$$

Apabila dari populasi berukuran N dengan cara tertentu hanya diambil sebagian saja sehingga yang terukur nilainya hanya n buah amatan (n < N) yaitu $X_1, X_2, \ldots, X_n$ kumpulan data merupakan sebuah sampel terhingga berukuran n, sehingga rata-rata sampelnya adalah:

$$\bar{x} = \frac{X_1 + X_2 + \cdots + Xn}{n}$$

Bila datanya berkelompok atau disusun dalam kelas–kelas interval dengan titik tengah (xi) dan dalam bentuk tabel frekuensi (fi), maka rata–ratanya menjadi

$$\bar{x} = \frac{\sum f_i x_i}{n}$$

Contoh 1.

Hitunglah rata–rata nilai ujian dari populasi mahasiswa:

63, 67, 70, 73, 74, 77, 78, 81, 82, 84, dan 88.

$$\mu = \frac{63 + 67 + 70 + 73 + 74 + 77 + 78 + 81 + 82 + 84 + 88}{11}$$

$$= \frac{837}{11} = 76{,}1$$

Contoh 2.

Hitunglah rata-rata denyut nadi 30 pasien sebuah rumah sakit.

Detakan permenit (Xi)	Banyaknya pasien (fi)	Hasil kali fi xi
70	3	3x70 = 210
75	6	6x75 = 450
80	7	7x 80 =560
83	11	11x 83 =913
85	3	3x 85 = 255
-	30	2.388

Rata–rata denyut nadi $\bar{X} = \frac{2.388}{30}$ =79,66 detakan / menit.

Contoh 3.

Sebanyak 50 mahasiswa mengikuti ujian tengah semester mata kuliah Statistik. Pengawas ujian mencatat waktu yang digunakan dalam pengerjaan soal masing-masing mahasiswa. Hitunglah rata-rata waktu pengerjaan soal tersebut.

Selang (menit)	Titik tengah xi	Frekuensi fi	Hasil kali fi.xi
31-35	33	4	132
36-40	38	8	304
41-45	43	10	430
46-50	48	15	720
51-55	53	8	424
56-60	58	5	290
-	-	50	2300

$$\bar{x} = \frac{\sum f_i x_i}{n} = \frac{2300}{50} = 46$$

Rata–rata waktu pengerjaan 46 menit.

b. Rata-Rata Sementara

Menghitung rata-rata dapat dilakukan dengan menggunakan rata–rata sementara. Rata-rata hitung $\bar{x}$ yang diperoleh dari jumlah rata-rata sementara dan simpangan rata-rata dirumuskan:

$$\bar{X} = \bar{X}s + \frac{fi\,.di}{\sum fi}$$

Variabel $\bar{X}$s merupakan rata–rata sementara. nilai simpangan $\boldsymbol{d}$ diperoleh dari setiap titik tengah dikurangi rata–rata sementaranya $\overline{X}i - \overline{X}s$

Contoh 4.

Gunakan data ujian 50 mahasiswa untuk mata kuliah Statistik pada **Contoh 3** di atas. Hitunglah rata-rata dengan menggunakan rata-rata sementara.

Penyelesaian

Pertama-tama pilih salah satu titik tengah sembarang untuk rata–rata sementara, misalkan $\overline{X}$s = 43. maka simpangan pada nilai rata–rata sementara berharga d=0. Untuk titik tengah kurang dari $\overline{X}$s berturut-turut diberi harga d=-5, d=-10 sedangkan yang lebih dari $\overline{X}$s berturut- turut diberi harga d =+5, d =+10 dan d=+15.

Selang (menit)	Titik tengah Xi	Frekwensi fi	Simpangan Di =(Xi-$\overline{X}$s)	Hasil kali Fixdi
31 – 35	33	4	-10	4x-10 = -40
36 – 40	38	8	-5	8x-5 = -40
41 – 45	43	10	0	10x0 = 0
46 – 50	48	15	+5	15x 5 = 75
51 – 55	53	8	+10	8x 10 = 80
56 – 60	58	5	+15	5x15 = 75
		50		150

Rata – rata : $\overline{X} = 43 + \frac{150}{50} = 46$ menit.

Contoh 5.

Ulangi **contoh 4**. gunakan sandi dimana simpangan pada nilai rata– rata sementara $\overline{X}$s =38 berharga d_0=0 dan untuk titik tengah lebih kecil dari $\overline{X}$s berturut – turut diberi harga d=-1, d=-2 sedangkan yang lebih besar dari $\overline{X}$s berturut–turut diberi harga d=+1, d=+2 dan d=+3. Panjang kelas p = 5

Selang (menit)	Titik tengah	Frekwensi fi	Simpangan Di=(Xi–Xs)	Hasil kali fi x di
31-35	33	4	-1	4x-1 = -4
36-40	38	8	0	8x0 = 0
41-45	43	10	+1	10x1 = 10
46-50	48	15	+2	15x2 = 30
51-55	53	8	+3	8x3 = 24
56-60	58	5	+4	5x4 = 20
-	-	50		80

Rata–rata $\overline{X} = 38+5(\frac{80}{50})=38+8 =46$ menit (hasilnya sama)

c. Rata-Rata Tertimbang

Dalam hal tertentu sebuah nilai variabel statistik mempunyai timbangan, bila kita merata-ratakan n buah data X_1, X_2, X_n tetapi dengan asumsi bahwa sebagian lebih penting dari yang lainnya dapat dilakukan dengan memberi timbangan W_1, W_2, W_3,…W_n pada nilai–nilai tersebut. Oleh karena itu rata–rata tertimbangnya adalah:

$$\bar{X}_w = \frac{W_1\ .X\ _1 +\ \ W_2,\ \ X_2 + \cdots + Wn.Xn}{W_1 + W_2 + \cdots + Wn}$$

Contoh 6.

Hitunglah rata-rata nilai mata kuliah statistik seorang mahasiswa yang memperoleh nilai 70 dan 80 untuk dua kali ujian sisipan dan nilai 65 pada ujian akhir bila nilai ujian akhir dianggap tiga kali lebih penting dari masing-masing ujian sisipan.

Penyelesaian

$X_1 = 70$; $W_1 = 1$; $X_2 = 80$; $W_2 = 1$ dan $X_3=65$; $W_3 = 3$ sehingga

$\bar{X}\text{w} = \frac{1.(70)+1.(80)+3.(65)}{(1+1+3)} = 69$

Jadi rata–rata nilai mata kuliah statistik adalah 69

d. Rata–Rata Gabungan

Misalkan data k buah populasi terhingga dengan ukuran populasi N_1, N_2,….N_k dan mempunyai rata–rata populasi μ_1, μ_2, …. μ_k maka rata–rata gabungan μg dari semua populasi adalah:

$$\mu_g = \frac{N_1.\mu_1 + N_2.\mu_2 + \cdots. + N_k.\mu_k}{N_1 + N_2 + N_k}$$

Contoh 7.

Empat fakultas eksakta memiliki 40, 50, 35, dan 30 mahasiswa. pada ujian akhir mata kuliah statistik terapan rata-rata hasil ujian mencapai 60, 70, 80, dan 90. Berapakah nilai rata-rata populasi gabungan μ

$$\mu_g = \frac{40.(60) + 50.(70) + 35.(80) + 30.(90)}{(40 + 50 + 35 + 30)} = 72{,}3$$

Jadi nilai rata-rata gabungan empat fakultas eksakta adalah 72,3

e. Rata-Rata Ukur

Jika perbandingan tiap dua data berurutan tetap atau hampir tetap, rata-rata geometrik lebih baik daripada rata-rata hitung, apabila dikehendaki rata-ratanya. Untuk data bernilai x_1, x_2, …, x_n, maka rata-rata ukur U didefinisikan:

$$U = \sqrt[n]{x_1.x_2.x_3 \ldots x_n}$$

Yaitu akar pangkat n dari x_1, x_2, ..., x_n.

Contoh 8.

Hitung rata-rata ukur data 2, 4, dan 8

$$U = \sqrt[n]{x_1.x_2.x_3 \ldots x_n} = \sqrt[3]{2 \times 4 \times 8} = 4$$

Untuk bilangan-bilangan bernilai besar, lebih baik digunakan logaritma:

$$\log U = \frac{\sum \log x_i}{n}$$

Yakni logaritma rata-rata ukur U sama dengan jumlah logaritma tiap data dibagi oleh banyak data. Rata-rata ukur U akan didapat dengan jalan mencari kembali logaritmanya.

Contoh 9.

Sekedar menunjukkan penggunaan rumus di atas, Rata-rata ukur data 2, 4, dan 8:

$$\log U = \frac{\sum \log x_i}{n} = \frac{\log 2 + \log 4 + \log 8}{3} = 0{,}6021$$

Sehingga U = 4

Untuk fenomena yang bersifat tumbuh dengan syarat-syarat tertentu, seperti pertumbuhan penduduk, bakteri dan lain-lain, sering digunakan rumus yang mirip rata-rata ukur adalah:

$$P_t = P_0 \left(1 + \frac{\bar{x}}{100}\right)^t$$

Dengan:

Po = keadaan awal atau permulaan

Pt = keadaan akhir

x = rata-rata pertumbuhan setiap satuan waktu

t = satuan waktu yang digunakan

Contoh 10.

Penduduk Indonesia pada akhir tahun 1946 ada 60 juta sedangkan akhir tahun 1956 mencapai 78 juta. Hitunglah laju rata-rata pertumbuhan penduduk tiap tahun

Diketahui t = 10, Po=60 dan Pt = 78

$$P_t = P_0 \left(1 + \frac{\bar{x}}{100}\right)^t$$

$$78 = 60\left(1 + \frac{\bar{x}}{100}\right)^{10}$$
$$\log 78 = \log 60 + 10 \log\left(1 + \frac{\bar{x}}{100}\right)$$
$$\bar{x} = 2{,}67$$

f. Rata-Rata Harmonik

Dalam praktek nilai rata-rata harmonik H, paling sering digunakan untuk merata-ratakan kecepatan jarak tempuh, menentukan harga rata-rata komoditi tertentu, menghitung investasi sejumlah uang tertentu setiap periode dan lain-lain.

Rata-rata harmonik bagi nilai n bilangan positif X_1, X_2,..., X_n adalah n dibagi dengan jumlah kebalikan bilangan – bilangan itu .jadi

$$H = \frac{n}{\frac{1}{X_1} + \frac{1}{X_2} + \cdots + \frac{1}{X_n}}$$

Contoh 11.

Hitunglah harga rata-rata beras rojolele /kg. minggu pertama terjual dengan harga Rp.5000/kg, minggu kedua terjual dengan harga Rp.4000/kg dan minggu keempat terjual dengan harga Rp.4800/kg.

Gunakan rata-rata harmonik untuk untuk menghitung harga beras rojolele:

$$H = \frac{4}{\frac{1}{5000} + \frac{1}{4500} + \frac{1}{4000} + \frac{1}{4800}} = 4.543$$

Jadi harga rata–rata beras rojolele adalah Rp.4.543/kg.

Contoh 12.

Si A bepergian pulang pergi. Waktu pergi ia melakukan kecepatan 10 km/jam sedangkan waktu kembalinya 20 km/jam berapakah rata-rata kecepatan pulang pergi?

$$H = \frac{2}{\frac{1}{10} + \frac{1}{20}} = 13{,}3$$

Jadi rata–rata kecepatan pulang pergi adalah 13,3

g. Rata-rata Kuadratis (NRK)

Biasanya NRK disebut juga "Akar Nilai Rata-rata" atau dikatakan sebagai "Nilai Rata-rata Kuadratis" dari kumpulan bilangan yang merupakan urutan $X_1, X_2, ..., X_n$ dan diberi simbol dengan:

$$NRK = \sqrt{\frac{\sum X_i^2}{N}}$$

Contoh 13.

Ada suatu deret bilangan 2, 4, 6, 8 Tentukan rata-rata kuadratisnya

$$NRK = \sqrt{\frac{\sum X_i^2}{N}} = \sqrt{\frac{2^2 + 4^2 + 6^2 + 8^2}{4}} = 5{,}48$$

Jadi rata-rata kuadratis data tersebut adalah 5,48

B. *MEDIAN*

Bila sekumpulan data statistik sebanyak N telah diurutkan dari yang terkecil sampai yang terbesar atau sebaliknya, maka data statistik yang berada ditengah-tengahnya disebut median (Me). Bila banyak pengamatan data ganjil, data yang di tengah–tengahnya adalah medianya atau bila banyak pengamatan genap, rata–rata kedua pengamatan yang ditengahnya adalah medianya. Median ditentukan dengan membagi kumpulan data menjadi dua bagian yang sama.

Bila datanya berkelompok (kelas interval), maka rumus mediannya adalah:

$$Me = Bb + p\left(\frac{\frac{n}{2} - F}{f_m}\right)$$

Keterangan:

Bb = Batas bawah kelas interval yang mengandung median

p = panjang kelas

F = Frekuensi kumulatif sebelum kelas interval median

fm = Frekuensi kelas interval yang mengandung median

n = banyaknya data

Contoh 14.

Diberikan data hasil ujian statistika sebagai berikut: 63, 65, 67, 70, 73, 74, 77, 78, 81, 81, 81, 82, dan 88. Tentukan median populasi hasil ujian tersebut

$$\text{Letak } Me = \frac{n+1}{2} = \frac{13+1}{2} = 7$$

Jadi median populasi terletak pada data ke 7 yaitu 77

Contoh 15.

Berat gabah panenan yang berasal dari sebuah sampel acak 10 karung adalah 82, 90, 65, 74, 98, 76, 97, 75, 91, dan 74 kilogram. Tentukan medianya:

Kumpulan data diurutkan dari yang besar sampai yang terkecil, kita peroleh: 98, 97, 91, 90, 82, 76, 75, 74, 74, dan 65 kilogram.

Median merupakan rata–rata berat 82 dan 76 kilogram yaitu:

$$Me = \frac{82+76}{2} = 79$$

Jadi median berat gabah panenan adalah 79

Contoh 16.

Menurut hasil pencacahan tahun 2005, jumlah penduduk didaerah terpencil sebanyak 80 jiwa dengan komposisi usia (dalam tahun). Hitunglah mediannya.

Usia (tahun)	Titik tengah	Batas kelas	Frekuensi
01 – 10	5,5	0,5 - 10,5	6
11 – 20	15,5	10,5 – 20,5	9
21 – 30	25,5	20,5 – 30,5	14
31 – 40	35,5	30,5 – 40,5	20
41 – 50	45,5	40,5 – 50,5	15
51 – 60	55,5	50,5 - 60,5	10
61 – 70	65,5	60,5 – 70,5	6
-	-	-	80

Diketahui: Bb = 30,5; p=10; F = 6+9+14 = 29;

fm = 20 dan n = 80

$$Me = Bb + p\left(\frac{\frac{n}{2} - F}{f_m}\right) = 30{,}5 + 10\left(\frac{\frac{80}{2} - 29}{20}\right) = 36$$

Jadi median data di atas adalah 36.

C. *MODUS*

Sekumpulan pengamatan data yang nilai terjadinya sering muncul atau yang mempunyai frekuensi paling tinggi disebut modus (M_0) atau nilai yang paling banyak di dalam satu kelompok nilai. Suatu distribusi mungkin tidak mempunyai modus, dengan kata lain modus tidak selalu ada. Hal ini bila semua pengamatan hanya mempunyai satu frekuensi saja. untuk data-data tertentu kemungkinan muncul beberapa nilai dengan frekuensi yang sama dan dalam hal ini dimungkinkan kumpulan data mempunyai lebih dari satu modus.

Apabila data dikelompokkan dalam kelas-kelas interval dan disajikan dalam tabel–tabel frekuensi tertentu, maka dalam mencari modus dapat dipergunakan rumus sebagai berikut:

$$Mo = Bb + p\left(\frac{b_1}{b_1 + b_2}\right)$$

Keterangan:

Bb = Batas bawah kelas interval yang mengandung modus

p = Panjang kelas interval

b1 = Selisih frekuensi yang mengandung modus dengan frekuensi sebelumnya

b2 = Selisih frekuensi yang mengandung modus dengan frekuensi setelahnya

Contoh 23.

Kadar tar dan nikotin pada 10 rokok kretek berbagai merk tercatat sebagai berikut: kandungan tar 30, 32, 33, 30, 30, 29, 30, 33, 28, 31 mg per bungkus dan kandungan nikotin: 1,7; 1,5; 1,9; 1,6; 1,7; 1,8; 1,7; 1,9; 1,6; 1,6 mg per bungkus.

Tentukan modus tar dan nikotin rokok kretek.

a. Kadar tar rokok kretek:

28,29,30,30,30,31,32,33, dan 33 mg.

modusnya adalah data yang sering muncul, M_0 = 30 mg.

b. Kandungan nikotin rokok kretek:

1,5; 1,6; 1,6; 1,6; 1,7; 1,7; 1,7; 1,8; 1,9; dan 1,9 mg. Dalam kasus ini terdapat dua modus kandungan nikotin 1,6 mg dan 1,7 mg karena mempunyai frekuensi tertinggi yaitu sebanyak 3 kali. Sebaran demikian di sebut bimodus.

Contoh 24.

Tentukan modus dari tabel frekuensi di bawah ini.

Kelas	Frekuensi
30 – 39	3
40 – 49	7
50 – 59	8
60 – 69	12
70 – 79	9
80 – 89	8
90 – 99	3
	50

Penyelesaian

Dari tabel di atas terlihat bahwa frekuensi yang mengandung modus berada pada kelas interval keempat, maka modus dapat dihitung dengan Bb = 59,5; p=10; n=50; b1=12-8 = 4 dan b2 = 12-9 =3

$$Mo = Bb + p\left(\frac{b_1}{b_1 + b_2}\right)$$

$$Mo = 59{,}5 + 10\left(\frac{4}{4 + 3}\right) = 65{,}21$$

BAB

4

UKURAN PENYEBARAN DATA

A. PENGANTAR

Dalam statistik, penting untuk menggunakan ukuran penyebaran data. Ukuran penyebaran data berfungsi sebagai dasar untuk membantu memahami sejauh mana data tunggal dan data kelompok tersebar dan seberapa besar variasinya. Lebih penting lagi, dengan pemahaman yang lebih baik tentang ukuran penyebaran, kita dapat membuat keputusan berdasarkan pada fakta yang lebih akurat.

Pada bab sebelumnya telah dijelaskan mengenai data tunggal dan data kelompok. Dalam bab ini, penyebaran data juga disajikan dalam bentuk data tunggal dan kelompok. Dalam statistik, cara pengolahan data tunggal dan data kelompok diperlakukan berbeda. Secara umum, ukuran sebaran data merupakan suatu ukuran yang menggambarkan seberapa sebaran data terhadap nilai tengah atau rata-rata. Dengan kata lain, pengukuran yang menunjukkan seberapa besar nilai data bervariasi terhadap nilai ukuran pusat. Penyebaran data digunakan untuk mengetahui seberapa besar (ukuran) sebaran data yang tersedia.

B. UKURAN PENYEBARAN DATA TUNGGAL

Ukuran penyebaran data tunggal adalah cara untuk mengukur seberapa jauh atau bervariasi nilai-nilai data tunggal dari nilai pusat atau nilai tengah. Ukuran penyebaran data tunggal yang digunakan meliputi:

1. Rentang (*Range*)

Rentang (R) digunakan sebagai ukuran awal untuk melihat seberapa jauh data dapat tersebar dalam sebuah kumpulan data. Selain itu, rentang dapat digunakan sebagai ukuran untuk melihat seberapa besar perbedaan antara dua nilai dalam sebuah himpunan data. Rentang bisa memberikan informasi kasar tentang seberapa besar perbedaan antara nilai terkecil dan terbesar.

Rentang merupakan selisih antara data tertinggi dengan data terendah dalam suatu kumpulan data. Rentang memberikan gambaran kasar tentang seberapa jauh nilai-nilai data tunggal dari nilai pusat. Semakin besar jangkauan artinya data semakin menyebar. Rentang dapat dihitung dengan menggunakan rumus:

$$R = X_{max} - X_{min} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Keterangan:

R = Rentang

X_{max} = Data tertinggi

X_{min} = Data terendah

Contoh 1:

Data luas lahan (hektar) petani kelapa, yaitu: 3, 6, 1, 2, 2, 4, 3. Tentukan rentang dari data tersebut.

Jawab:

Nilai terurut dari data tersebut adalah 1, 2, 2, 3, 3, 4, 6.

$X_{max} = 6$

$X_{min} = 1$

$R = X_{max} - X_{min}$

$R = 6 - 1 = 5$

Jadi, rentang luas lahan kelapa milik petani adalah 5 hektar.

2. Varians (*Variance*)

Varians menggambarkan suatu ukuran seberapa jauh nilai-nilai data tunggal tersebar dari nilai rata-rata. Nilai rata-rata digunakan sebagai acuan untuk menghitung varians. Varians menghasilkan nilai positif, Semakin tinggi nilai varians, semakin besar variasi nilai dalam data tunggal.

Simbol varians untuk populasi disimbolkan dengan σ^2, sedangkan untuk sampel, varians disimbolkan dengan S^2. Varians bisa dihitung dengan menggunakan dua metode, yaitu (1) metode biasa, dan (2) metode angka kasar. Perbedaan utama antara metode biasa dan metode angka kasar adalah pada bagaimana rata-rata dihitung. Metode biasa menghitung nilai rata-rata langsung dari nilai data, sementara metode angka kasar lebih fokus pada kuadrat setiap nilai dan selisih kuadrat setiap nilai dari kuadrat rata-rata. Meskipun begitu, nilai akhir perhitungan varians dengan kedua metode harus sama, karena data yang digunakan sama hanya berbeda dalam penggunaan rumus saja.

1) Metode Biasa

Metode biasa juga dikenal sebagai metode pengujian kuadrat rata-rata. Metode biasa merupakan salah satu metode yang paling sering digunakan dalam menghitung varians. Metode biasa sangat sesuai digunakan pada sampel yang cukup besar. Dengan demikian, perhitungan dibedakan berdasarkan pada jumlah sampel.

a. Untuk Sampel Besar (N > 30)

$$S^2 = \frac{\sum (X_i - \bar{X})^2}{n} \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (2)$$

b. Untuk Sampel Kecil (N ≤ 30)

$$S^2 = \frac{\sum (X_i - \bar{X})^2}{n-1} \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (3)$$

2) Metode Angka Kasar

Metode angka kasar juga dikenal sebagai metode deviasi terstandarisasi. Metode angka kasar merupakan metode penghitungan varians dengan cara menghitung kuadrat setiap nilai dan selisih kuadrat setiap nilai dari kuadrat rata-rata, kemudian menjumlahkan semua nilai tersebut hingga diperoleh varians.

a. Untuk Sampel Besar (N > 30)

$$S^2 = \frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2 \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (4)$$

b. Untuk Sampel Kecil (N ≤ 30)

$$S^2 = \frac{\sum X^2}{n-1} - \frac{(\sum X)^2}{n(n-1)} \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (5)$$

Contoh 2:

Tentukan nilai varians dari 1, 2, 2, 3, 3, 4, 6. Hitunglah dengan menggunakan kedua metode perhitungan varians.

Jawab:

Menghitung varians, dilakukan dengan menggunakan langkah-langkah sebagai berikut:

Langkah 1. Menghitung rata-rata ($\bar{X}$)

$$\overline{\mathrm{X}} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{1+2+2+3+3+4+6}{7} = 3$$

Langkah 2. Menghitung rumus ke dalam bentuk tabel

Tabel 4.1 Perhitungan Varians

X	$(\mathbf{X}-\overline{\mathbf{X}})$	$(\mathbf{X}-\overline{\mathbf{X}})^2$	$\mathbf{X}^2$
1	1 – 3 = - 2	-2 x -2 = 4	$1^2 = 1$
2	2 – 3 = - 1	-1 x -1 = 1	$2^2 = 4$
2	2 – 3 = - 1	-1 x -1 = 1	$2^2 = 4$
3	3 – 3 = 0	0 x 0 = 0	$3^2 = 9$
3	3 – 3 = 0	0 x 0 = 0	$3^2 = 9$
4	4 – 3 = 1	1 x 1 = 1	$4^2 = 16$
6	6 – 3 = 3	3 x 3 = 9	$6^2 = 36$
21		**16**	**79**

Langkah 3. Menghitung varians dengan kedua metode

Berdasarkan tabel pada Langkah 2, jumlah sampel (nilai X) berjumlah 21, maka menggunakan rumus dengan sampel kecil (N≤ 30).

Metode Biasa: $\mathrm{S}^2 = \frac{\sum(X_i-\bar{X})^2}{n-1} = \frac{16}{7-1} = 2{,}67$

Metode Angka Kasar: $\mathrm{S}^2 = \frac{\sum \mathrm{X}^2}{\mathrm{n}-1} - \frac{(\sum \mathrm{X})^2}{\mathrm{n}(\mathrm{n}-1)} = \frac{79}{7-1} - \frac{21^2}{7(7-1)} = \frac{79}{6} - \frac{441}{42}$

$= 13{,}17 - 10{,}5 = 2{,}67$

Jadi, nilai varians dari data tersebut adalah 2,67

3. Standar Deviasi (Simpangan Baku)

Deviasi standar atau standar deviasi adalah cara lain untuk mengukur penyebaran data tunggal dari nilai rata-rata. Standar deviasi digunakan untuk mengukur seberapa jauh nilai-nilai data tersebar dari nilai rata-rata. Standar deviasi dikenal pula sebagai akar kuadrat dari varians. Standar deviasi disimbolkan dengan huruf S.

Sama halnya pada varians, menghitung standar deviasi menggunakan dua metode rumus, yaitu metode biasa dan metode angka kasar. Perbedaan antara standar deviasi metode biasa dan metode angka kasar adalah pada metode penghitungan nilai rata-rata dan variansnya. Standar deviasi metode biasa dan metode angka kasar hanya berbeda dalam penghitungan varians.

1) Metode Biasa

Standar deviasi metode biasa dihitung dengan mengambil akar kuadrat dari varians yang telah dihitung pada metode biasa. Dalam penghitungan ini, selisih kuadrat setiap nilai data dan nilai rata-rata dijumlahkan kemudian dibagi dengan jumlah seluruh data pada sampel.

a. Untuk Sampel Besar (N > 30)

$$S = \sqrt{\frac{\sum(X-\bar{X})^2}{n}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (6)$$

b. Untuk Sampel Kecil (N ≤ 30)

$$S = \sqrt{\frac{\sum(X-\bar{X})^2}{n-1}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (7)$$

2) Metode Angka Kasar

Standar deviasi metode angka kasar juga dihitung dengan cara mengambil akar kuadrat dari varians yang dihitung menggunakan rumus metode angka kasar.

a. Untuk Sampel Besar (N > 30)

$$S = \sqrt{\frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (8)$$

b. Untuk Sampel Kecil (N ≤ 30)

$$S = \sqrt{\frac{\sum X^2}{n-1} - \frac{(\sum X)^2}{n(n-1)}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (9)$$

Contoh 3:

Tentukan standar deviasi dari data pada contoh Soal 2.

Jawab:

Diketahui data sebagai berikut: 1, 2, 2, 3, 3, 4, 6

Dari contoh soal 2, telah menghitung bahwa nilai varians dari data tersebut adalah $S^2 = 2{,}67$. Dengan demikian, standar deviasi dari data tersebut yakni:

Metode biasa: $S = \sqrt{\frac{\sum(X-\bar{X})^2}{n-1}} = \sqrt{\frac{16}{7-1}} = \sqrt{\frac{16}{6}} = 1{,}63$

Metode angka kasar:

$$S = \sqrt{\frac{\sum X^{2}}{n-1} - \frac{(\sum X)^{2}}{n(n-1)}} = \sqrt{\frac{79}{7-1} - \frac{21^{2}}{7(7-1)}} = \sqrt{13{,}17 - 10{,}5} = \sqrt{2{,}67} = 1{,}63$$

Jadi simpangan baku dari contoh di atas adalah 1,63.

4. Kuartil (*Quartiles*)

Kuartil adalah nilai yang membagi data tunggal menjadi empat bagian sama besar. Atau dengan kata lain, kuartil merupakan nilai yang membagi data dari sampel atau populasi menjadi empat bagian yang sama. Kuartil pertama (Q_1) membagi data menjadi 25% terbawah, kuartil kedua (Median) membagi data menjadi 50% dan kuartil ketiga (Q_3) membagi data menjadi 75% teratas.

Untuk menghitung kuartil, anda harus terlebih dahulu menentukan kuartil ketiga (Q_3) dan kuartil pertama (Q_1). Rentang antar kuartil merupakan selisih antara kuartil ketiga dengan kuartil pertama. Dengan menggunakan rumus:

$$\text{Rentang antar kuartil} = \frac{1}{2}\, Q_3 - Q_1 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (10)$$

$$\text{Variasi Kuartil} = \frac{\text{Rentang antar kuartil}}{\text{Median}} \text{ x } 100\% \ldots\ldots\ldots\ldots (11)$$

Contoh 4:

Tentukan rentang antar kuartil dari data pada contoh Soal 2.

Jawab:

Diketahui data sebagai berikut: 1, 2, 2, 3, 3, 4, 6

Langkah 1. Urutkan data dari data terendah ke data tertinggi 1, 2, 2, 3, 3, 4, 6

Langkah 2. Tentukan nilai Q1 dan Q3

Menentukan kuartil pertama (Q1) = 2

Menentukan kuartil pertama (Q3) = 4

Menentukan kuartil kedua (median) = 3

Langkah 3. Menentukan rentang antar kuartil

$$\text{Rentang antar kuartil} = \frac{1}{2}\,Q_3 - Q_1 = \frac{1}{2}(4-2) = 1$$

$$\text{Variasi Kuartil} = \frac{1}{3}\text{ x }100\% = 33{,}33\%$$

C. UKURAN PENYEBARAN DATA KELOMPOK

Ukuran penyebaran data kelompok adalah cara untuk mengukur penyebaran data antar kelompok data yang telah dikelompokkan atau diklasifikasikan. Berikut adalah beberapa ukuran penyebaran data kelompok yang paling umum digunakan:

1. Deviasi Rata-Rata

Deviasi adalah ukuran jarak antara setiap nilai dalam kumpulan data dari nilai rata-rata atau nilai tengah. Deviasi adalah indikator yang dapat digunakan untuk menentukan variabilitas data dalam populasi atau sampel. Semakin tinggi nilai deviasi, semakin besar variabilitas atau keragaman data dalam sampel atau populasi. Sebaliknya, semakin kecil nilai deviasi maka semakin homogen data dalam sampel atau populasi. Penghitungan deviasi pada data kelompok mengikuti langkah-langkah berikut:

1) Menentukan kelas-kelas dalam data dan menghitung titik tengah setiap kelas.
2) Tentukan frekuensi (f) dari setiap kelas
3) Menghitung nilai rata-rata dari seluruh titik tengah kelas ($\bar{x}$), dengan menggunakan rumus: $\bar{x} = (\sum \text{fx})/\text{n}$
4) Menghitung deviasi setiap data terhadap rata-rata ($\bar{x}$) dengan melakukan pengurangan antara rata-rata dengan titik tengah kelas
5) Menghitung deviasi rata-rata (DR)

Catatan: Dalam mendefinisikan dan cara menentukan kelas dan frekuensi telah dibahas pada bab sebelumnya.

$$\text{DR} = \frac{1}{\text{n}}\sum \text{f}(\text{x}_\text{i} - \bar{\text{x}}) = \frac{|\sum \text{f}\,(\text{x}_\text{i} - \bar{\text{x}})|}{\text{n}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots . (12)$$

Pemberian tanda mutlak “| |” dimaksudkan agar penyebaran data selalu bernilai positif.

Contoh 5:

Untuk meningkatkan produksi kelapa rakyat, telah diteliti sebanyak 50 sampel petani kelapa di suatu desa. Adapun salah satu faktor yang diteliti yaitu penggunaan bibit unggul. Hasil penelitian tersebut adalah:

Penggunaan Bibit Unggul (Batang/ Ha)	Petani Kelapa (F)
35 – 39	3
40 – 44	6
45 – 49	7
50 – 54	13
55 – 59	9
60 – 64	8
65 – 69	4
Jumlah	50

Tentukan deviasi rata-rata dari data didalam tabel.

Jawab:

Langkah 1. Menentukan $\bar{x}$. Untuk memudahkan perhitungan digunakan tabel. Penggunaan bibit unggul dinotasikan dengan "Interval" dan Petani kelapa disimbolkan dengan "f".

Interval	Titik Tengah (x)	f	f.x
35 – 39	37	3	111
40 – 44	42	6	252
45 – 49	47	7	329
50 – 54	52	13	676
55 – 59	57	9	513
60 – 64	62	8	496
65 – 69	67	4	268
Jumlah		50	2645

Keterangan: titik tengah = nilai tertinggi – nilai terendah di setiap kelas

$\bar{x} = (\sum fx)/n = 2645/50 = 52{,}9$

Langkah 2. Menghitung deviasi pada setiap data didalam tabel.

Interval	x	f	(x - x̄)	f (x - x̄)
35 – 39	37	3	-15,9	-47,7
40 – 44	42	6	-10,9	-65,4
45 – 49	47	7	-5,9	-41,3
50 – 54	52	13	-0,9	-11,7
55 – 59	57	9	4,1	36,9
60 – 64	62	8	9,1	72,8
65 – 69	67	4	14,1	56,4
Jumlah		50		332,2

Langkah 3. Menghitung deviasi rata-rata

$$\mathrm{DR} = \frac{332{,}2}{50} = 6{,}64$$

Jadi deviasi rata-rata adalah 6,64

2. Variansi (*Variance*)

Varians untuk data kelompok digunakan untuk menggambarkan sebaran data kelompok. Varians adalah ukuran statistik yang menunjukkan seberapa jauh setiap titik data dalam kumpulan data dari rata-rata. Varians dalam data kelompok dihitung dengan menggunakan frekuensi relatif, dengan menghitung selisih dari setiap nilai dengan *mean*, kemudian mengkuadratkan selisih ini, lalu membagi dengan jumlah nilai dalam data. Varians dapat dihitung dengan menggunakan 3 metode, yaitu; (1) metode biasa; (2) metode angka kasar, dan (3) metode *coding*.

1) Metode Biasa

Metode biasa pada varians atau disebut juga dengan metode jangka pendek atau metode simpangan kuadrat yang tak biasa (*unbiased squared deviation*) digunakan untuk menghitung varians data sampel. Metode ini digunakan karena sampel yang diambil merupakan bagian dari populasi, sehingga varians yang dihasilkan harus disesuaikan dengan ukuran sampel yang lebih kecil. Metode biasa masih umum digunakan karena memiliki keuntungan yaitu mudah dihitung dan diinterpretasikan.

a. Untuk Sampel Besar (N > 30)

$$S^2 = \frac{\sum f(x-\bar{x})^2}{n} \text{ .... (13)}$$

b. Untuk Sampel Kecil (N ≤ 30)

$$S^2 = \frac{\sum f(x-\bar{x})^2}{n-1} \text{ .... (14)}$$

Contoh 6:

Tentukan varians dari data pada Contoh 5.

Jawab:

Frekuensi dengan jumlah 50, sehingga rumus yang digunakan yaitu sampel besar (sampel >30)

Interval	x	f	(x - x̄)	(x - x̄)²	f (x - x̄)²
35 – 39	37	3	-15,9	252,81	758,43
40 – 44	42	6	-10,9	118,81	712,86
45 – 49	47	7	-5,9	34,81	243,67
50 – 54	52	13	-0,9	0,81	10,53
55 – 59	57	9	4,1	16,81	151,29
60 – 64	62	8	9,1	82,81	662,48
65 – 69	67	4	14,1	198,81	795,24
Jumlah		50		705,67	3.334,5

$$S^2 = \frac{3.334,5}{50} = 66,69$$

2) Metode Angka Kasar

Metode angka kasar juga dikenal sebagai metode jangka panjang atau metode simpangan kuadrat yang biasa (*biased squared deviation*) yang digunakan untuk menghitung varians data populasi. Metode ini digunakan pada populasi data karena asumsi bahwa semua populasi data yang tersedia diketahui. Selain itu, metode ini mengabaikan sebagian besar variasi antara sampel besar atau dekat dengan populasi.

a. Untuk Sampel Besar (N > 30)

$$S^2 = \frac{\sum fX^2}{n} - \left(\frac{\sum fX}{n}\right)^2 \text{ .... (15)}$$

b. Untuk Sampel Kecil (N ≤ 30)

$$S^2 = \frac{\sum fX^2}{n-1} - \left(\frac{\sum fX}{n-1}\right)^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (16)$$

Contoh 7:

Tentukan varians dari data pada Contoh 5.

Jawab:

Frekuensi dengan jumlah 50, sehingga rumus yang digunakan yaitu sampel besar (sampel >30)

Interval	x	f	X²	fx	fx²
35 – 39	37	3	1369	111	4107
40 – 44	42	6	1764	252	10584
45 – 49	47	7	2209	329	15463
50 – 54	52	13	2704	676	35152
55 – 59	57	9	3249	513	29241
60 – 64	62	8	3844	496	30752
65 – 69	67	4	4489	268	17956
Jumlah		50		2645	143255

$$S^2 = \frac{\sum fX^2}{n} - \left(\frac{\sum fX}{n}\right)^2 = \frac{143255}{50} - \left(\frac{2645}{50}\right)^2 = 2865{,}1 - 2798{,}41 = 66{,}69$$

3) Metode *Coding*

Metode *coding* adalah pendekatan untuk menghitung varians dari sekelompok data dengan menggunakan sistem penomoran atau penandaan tertentu untuk data tersebut. Pendekatan ini berguna ketika membandingkan variabilitas antara beberapa kelompok data atau ingin menghindari perhitungan yang rumit.

a. Untuk Sampel Besar (N > 30)

$$S^2 = C^2 \frac{\sum fu^2}{n} - \left(\frac{\sum fu}{n}\right)^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (17)$$

b. Untuk Sampel Kecil (N ≤ 30)

$$S^2 = C^2 \frac{\sum fu^2}{n-1} - \left(\frac{\sum fu}{n-1}\right)^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (18)$$

Contoh 8:

Tentukan varians dari data pada Contoh 5.

Jawab:

Frekuensi dengan jumlah 50, sehingga rumus yang digunakan yaitu sampel besar (sampel >30). Pada metode *coding*, simbol C merupakan interval kelas. Jadi, interval pada soal adalah 5.

Interval	x	f	u	u^2	fu	fu^2
35 – 39	37	3	-3	9	-9	27
40 – 44	42	6	-2	4	-12	24
45 – 49	47	7	-1	1	-7	7
50 – 54	52	13	0	0	0	0
55 – 59	57	9	1	1	9	9
60 – 64	62	8	2	4	16	32
65 – 69	67	4	3	9	12	36
Jumlah		50			9	135

$$S^2 = C^2 \frac{\sum fu^2}{n} - \left(\frac{\sum fu}{n}\right)^2 = 5^2 \frac{135}{50} - \left(\frac{9}{50}\right)^2 = 25\,\{2{,}7 - (0{,}18)^2\}$$

$$= 25\,(2{,}6676) = 66{,}69$$

3. Simpangan Baku (*Standard Deviation*)

Simpangan baku adalah ukuran statistik yang mengukur sejauh mana data dalam suatu kelompok tersebar dari rata-ratanya. Simpangan baku memberikan gambaran tentang variasi atau penyebaran data disekitar nilai rata-rata. Semakin besar nilai simpangan baku, semakin banyak variasi data yang dimiliki. Sama halnya dengan varians, menghitung simpangan baku juga menggunakan 3 metode, yaitu; (1) metode biasa; (2) metode angka kasar, dan (3) metode *coding*.

1) Metode Biasa

a. Untuk Sampel Besar (N > 30)

$$S = \sqrt{\frac{\sum f(x-\bar{x})^2}{n}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(19)}$$

b. Untuk Sampel Kecil (N ≤ 30)

$$S = \sqrt{\frac{\sum f(x-\bar{x})^2}{n-1}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(20)}$$

2) Metode Angka Kasar

a. Untuk Sampel Besar (N > 30)

$$S = \sqrt{\frac{\sum fX^2}{n} - \left(\frac{\sum fX}{n}\right)^2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (21)$$

b. Untuk Sampel Kecil (N ≤ 30)

$$S = \sqrt{\frac{\sum fX^2}{n-1} - \frac{(\sum fX)^2}{n(n-1)}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (22)$$

3) Metode *Coding*

a. Untuk Sampel Besar (N > 30)

$$S = C\sqrt{\frac{\sum fu^2}{n} - \left(\frac{\sum fu}{n}\right)^2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (23)$$

b. Untuk Sampel Kecil (N ≤ 30)

$$S = C\sqrt{\frac{\sum fu^2}{n-1} - \frac{(\sum fu)^2}{n(n-1)}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (24)$$

Contoh 9:

Tentukan simpangan baku dari data pada Contoh 5. Hitunglah dengan menggunakan ketiga metode.

Jawab:

Frekuensi dengan jumlah 50, sehingga rumus yang digunakan yaitu sampel besar (sampel >30). Pada metode *coding*, simbol C merupakan interval kelas. Jadi, interval pada soal adalah 5.

<u>Metode Biasa</u>

Interval	x	f	(x - x̄)	fx	(x - x̄)²	f (x - x̄)²
35 – 39	37	3	-15,9	111	252,81	758,43
40 – 44	42	6	-10,9	252	118,81	712,86
45 – 49	47	7	-5,9	329	34,81	243,67
50 – 54	52	13	-0,9	676	0,81	10,53
55 – 59	57	9	4,1	513	16,81	151,29
60 – 64	62	8	9,1	496	82,81	662,48

65 – 69	67	4	14,1	268	198,81	795,24
Jumlah		50		2645		3.334,5

$$S = \sqrt{\frac{\sum f(x-\bar{x})^2}{n}} = \sqrt{\frac{3334,5}{50}} = \sqrt{66,69} = 8,17$$

Metode Angka Kasar

Interval	x	f	X²	fx	fx²
35 – 39	37	3	1369	111	4107
40 – 44	42	6	1764	252	10584
45 – 49	47	7	2209	329	15463
50 – 54	52	13	2704	676	35152
55 – 59	57	9	3249	513	29241
60 – 64	62	8	3844	496	30752
65 – 69	67	4	4489	268	17956
Jumlah		50		2645	143255

$$S = \sqrt{\frac{\sum fX^2}{n} - \left(\frac{\sum fX}{n}\right)^2} = \sqrt{\frac{143255}{50} - \left(\frac{2645}{50}\right)^2} = \sqrt{66,69} = 8,17$$

Metode *Coding*

Interval	x	f	u	u²	fu	fu²
35 – 39	37	3	-3	9	-9	27
40 – 44	42	6	-2	4	-12	24
45 – 49	47	7	-1	1	-7	7
50 – 54	52	13	0	0	0	0
55 – 59	57	9	1	1	9	9
60 – 64	62	8	2	4	16	32
65 – 69	67	4	3	9	12	36
Jumlah		50			9	135

$$S = C\sqrt{\frac{\sum fu^2}{n} - \left(\frac{\sum fu}{n}\right)^2} = 5\sqrt{\frac{135}{50} - \left(\frac{9}{50}\right)^2} = 5\sqrt{2,6676} = 8,17$$

4. Koefisien Variasi (*Variance Coefficient*)

Koefisien Variasi (KV) adalah ukuran statistik yang digunakan untuk mengukur tingkat relatif dari variasi atau dispersi dalam suatu kelompok data. Koefisien variasi mengukur seberapa besar simpangan baku dari rata-rata dalam proporsi terhadap nilai rata-rata itu sendiri.

Analisis ini memungkinkan untuk membandingkan tingkat variasi antara dua kelompok data yang memiliki skala yang berbeda. Dengan kata lain, dapat membandingkan tingkat variasi antara dua kumpulan data yang berbeda. Koefisien variasi memiliki keterbatasan yaitu hanya relevan bila rata-rata data tidak mendekati nol. Jika rata-rata mendekati nol atau nol, maka koefisien variasi akan menghasilkan nilai yang tidak bermakna. Rumus koefisien variasi sebagai berikut:

$$\mathrm{KV} = \frac{\mathrm{s}}{\bar{\mathrm{x}}} \times 100\% \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. (25)$$

Keterangan:
KV = Koefisien Variasi
s = Simpangan Baku
$\bar{X}$ = Rata-rata

Dengan kriteria penilaian:

KV < 15%: kategori KV rendah, ini menunjukkan bahwa data cenderung homogen dan memiliki sedikit variasi relatif terhadap rata-ratanya. Data dengan KV rendah biasanya memiliki konsistensi dan relatif stabil.

KV 15% - 30%: kategori KV sedang, menunjukkan bahwa tingkat variasi relatif dalam data masih cukup terkendali.

KV > 30%: kategori KV tinggi, mengindikasikan bahwa data memiliki variasi yang signifikan relatif terhadap rata-ratanya.

Contoh 10:

Dua perusahaan PT. A dan PT. B memiliki karyawan masing-masing sebanyak 50 orang. Untuk keperluan penelitian variasi gaji karyawan, diambil sampel sebanyak 7 orang setiap perusahaan. Dengan gaji setiap karyawan perminggu (dalam ribuan rupiah): 400, 350, 450, 500, 700, 600, 650, *dan* 300, 550, 350, 400, 450, 850, 600. Tentukan dispersi relatif dari kedua Perusahaan tersebut.

Jawab:

Langkah 1. Menentukan nilai x dan $\overline{\mathrm{X}}$ untuk kedua Perusahaan

$\sum$A = 400+350+450+500+700+600+650 = 3.650

$\sum$B = 300+550+350+400+450+850+600 = 3.500

$$\overline{\mathrm{X}}_{\mathrm{A}} = \frac{\sum \mathrm{X}_{\mathrm{A}}}{\mathrm{n}} = \frac{400 + 350 + 450 + 500 + 700 + 600 + 650}{7} = \frac{3650}{7} = 521{,}43$$

$$\overline{\mathrm{X}}_{\mathrm{B}} = \frac{\sum \mathrm{X}_{\mathrm{B}}}{\mathrm{n}} = \frac{300 + 550 + 350 + 400 + 450 + 850 + 600}{7} = \frac{3500}{7} = 500$$

Langkah 2. Menentukan nilai $\sum X^2$

$$\sum \mathrm{X}^2{}_{\mathrm{A}} = 400^2 + 350^2 + 450^2 + 500^2 + 700^2 + 600^2 + 650^2 = 2.007.500$$

$$\sum \mathrm{X}^2{}_{\mathrm{B}} = 300^2 + 550^2 + 350^2 + 400^2 + 450^2 + 850^2 + 600^2 = 1.960.000$$

Langkah 3. Menentukan simpangan baku (s)

$$\mathrm{S}_A = \sqrt{\frac{\sum x^2}{7} - \left(\frac{\sum x}{7}\right)^2} = \sqrt{\frac{2.007.500}{7} - \left(\frac{3.650}{7}\right)^2} = \sqrt{286.785{,}71 - 271.888} = 122{,}06$$

$$\mathrm{S}_{\mathrm{B}} = \sqrt{\frac{\sum x^2}{7} - \left(\frac{\sum x}{7}\right)^2} = \sqrt{\frac{1.960.000}{7} - \left(\frac{3.500}{7}\right)^2} = \sqrt{280.000 - 250.000} = 173{,}20$$

Langkah 4. Menentukan koefisien variasi (KV)

$$\mathrm{KV}_A = \frac{\mathrm{s}}{\overline{\mathrm{x}}} \text{ x } 100\% = \frac{122{,}06}{521{,}43} x 100\% = 23{,}41\ \%$$

$$\mathrm{KV}_B = \frac{\mathrm{s}}{\overline{\mathrm{x}}} \text{ x } 100\% = \frac{173{,}20}{500} x 100\% = 34{,}64\ \%$$

Kesimpulan: Berdasarkan dua Perusahaan yang telah dianalisis, maka Perusahaan A dengan nilai KV 23,41% (kategori sedang) sedangkan untuk Perusahaan B dengan nilai KV 34,64% (kategori tinggi).

BAB
5

BENTUK DISTRIBUSI DATA

A. PENGANTAR

Pada bab terdahulu sudah dibahas mengenai statistik deskriptif tentang distribusi frekuensi, ukuran tendensi sentral, dan ukuran dispersi (penyebaran).

Adapun distribusi frekuensi, ukuran tendensi sentral, dan ukuran dispersi tersebut dapat menggambarkan distribusi datanya, tetapi itu tidak cukup untuk menggambarkan sifat-sifatnya.

Bentuk distribusi dapat dilihat dari:

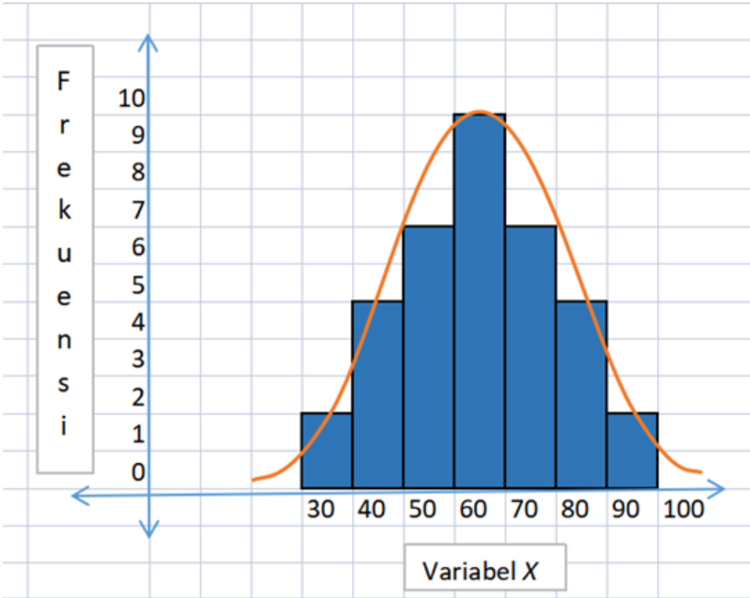
a. Kesimetriannya
b. *Skewness*
c. Jumlah puncaknya
d. Kurtosis

Selanjutnya penjelasan simetri, *skewness*, puncak, dan kurtosis, dibahas dalam sub bab berikut.

B. SIMETRI

Suatu distribusi dikatakan simetri jika sisi kanan distribusinya sama dengan sisi kirinya.

Gambar 5.1 adalah contoh distribusi simetri:



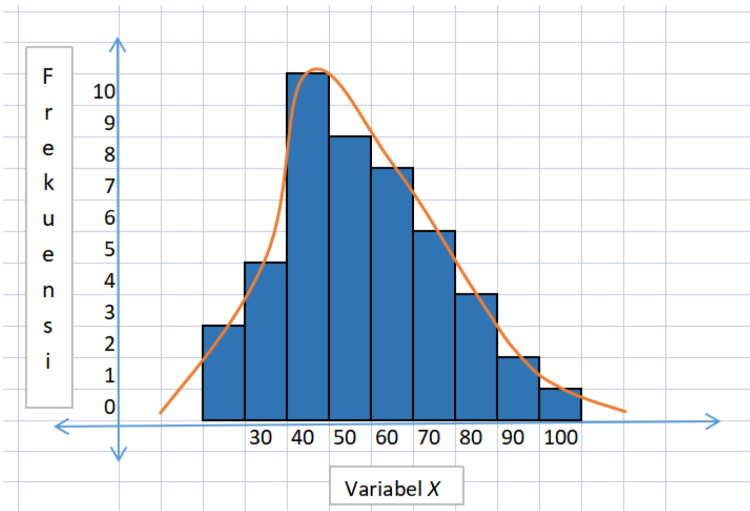
Gambar 5.1 Contoh Distribusi Simetri

Dari distribusi simetri pada gambar 5.1 dapat dihitung ukuran tendensi sentralnya bahwa nilai *mean*, median, dan modus adalah sama.

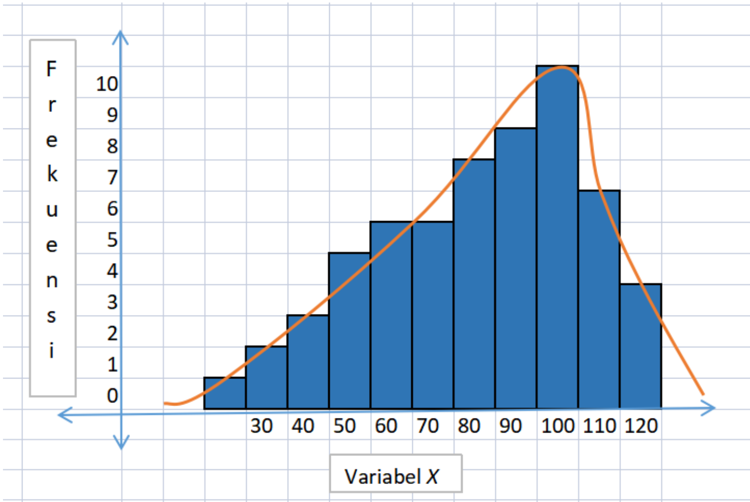
Apabila hubungan *mean*, median, dan modus tidak sama maka distribusinya asimetri. Bentuk distribusi asimetri dapat dibagi dua, yaitu:

a. Jika modus < median < *mean*, maka distribusi asimetri bentuknya miring ke kanan, contohnya lihat gambar 5.2.

b. Jika *mean* < median < modus, maka distribusi asimetri bentuknya miring ke kiri, , contohnya lihat gambar 5.3.



Gambar 5.2 Contoh Distribusi Asimetri Miring Kanan



Gambar 5.3 Contoh Distribusi Asimetri Miring Kiri

C. *SKEWNESS*

Istilah *skewness* sering juga menggunakan sebutan kemiringan, kemencengan, kecondongan, kemenjuluran data, dan ada juga yang menyebut juling (Hadi, Sutrisno, 2015).

Menurut DR. Boediono (2014) dalam bukunya yang berjudul Teori dan Aplikasi Statistika da Probabilitas, bahwa *kemiringan adalah derajat atau ukuran dari ketidaksimetrian (asimetri) suatu distribusi data.*

Berarti *skewness* adalah nilai statistik yang memberitahu apakah suatu distribusi simetri atau asimetri.

Terdapat 3 (tiga) formula untuk menghitung *skewness*, yaitu:

a. Formula Pearson

Pearson memformulasikan dengan menggunakan ukuran tendensi sentral dan dispersi, yaitu:

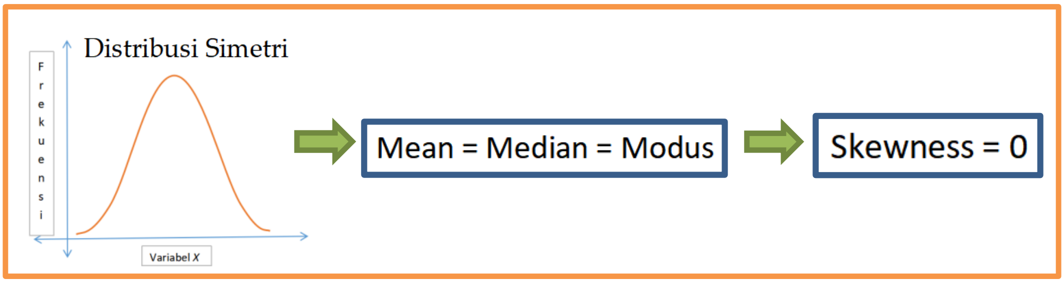
$$\text{Skewness} = \frac{\text{Mean} - \text{Modus}}{\text{Deviasi Baku}}$$

atau

$$\text{Skewness} = \frac{3 \text{ x (Mean} - \text{Median)}}{\text{Deviasi Baku}}$$

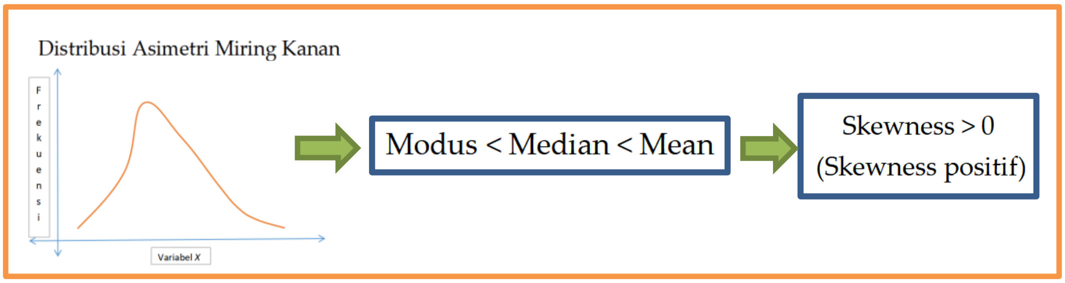
Dari penjelasan tentang distribusi simetri di sub bab B, dan dikaitkan dengan formula Pearson, dapat diambil pengertian bahwa:

1) Jika distribusinya simetri, maka nilai *mean*, median, dan modus adalah sama, dan mengakibatkan *skewness* bernilai nol (*skewness* = 0).
   Ilustrasinya dapat dilihat pada gambar 5.4.



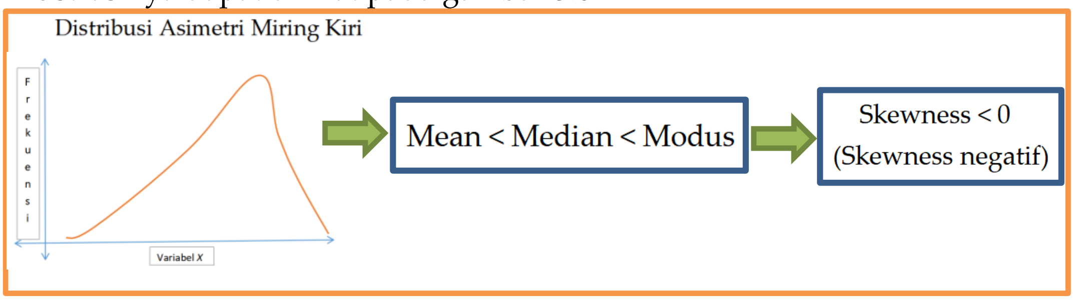
Gambar 5.4 Ilustrasi Distribusi Simetri

2) Jika distribusi asimetri miring kanan, maka modus < median < *mean*, dan mengakibatkan *skewness* > 0 (*skewness* positif).
   Ilustrasinya dapat dilihat pada gambar 5.5.



Gambar 5.5 Ilustrasi Distribusi Asimetri Miring Kanan

3) Jika distribusi asimetri miring kiri, maka *mean* < median < modus, dan mengakibatkan *skewness* < 0 (*skewness* negatif).
   Ilustrasinya dapat dilihat pada gambar 5.6.



Gambar 5.6 Ilustrasi Distribusi Asimetri Miring Kiri

b. Formula Bowley

Bowley memformulasikan dengan menggunakan ukuran letak data yang telah diurutkan terlebih dahulu dari nilai terendah sampai dengan nilai tertinggi. Bowley memakai ukuran letak data kuartil, yaitu sekumpulan datanya tersebut dibagi menjadi 4 (empat) bagian, sehingga terdapat 3 (tiga) buah kuartil, masing-masing urutannya adalah kuartil satu (K_1), kuartil dua (K_2), dan kuartil tiga (K_3).

Adapun Bowley mendefinisikan formulanya adalah sebagai berikut:

$$\text{Skewness} = \frac{(\mathrm{K}_3 - \mathrm{K}_2) - (\mathrm{K}_2 - \mathrm{K}_1)}{(\mathrm{K}_3 - \mathrm{K}_2) + (\mathrm{K}_2 - \mathrm{K}_1)}$$

yang dapat disederhanakan menjadi

$$\text{Skewness} = \frac{\mathrm{K}_3 - 2\mathrm{K}_2 + \mathrm{K}_1}{\mathrm{K}_3 - \mathrm{K}_1}$$

Telah diketahui bahwa distribusi simetri menghasilkan *skewness* = 0, maka dari formula Bowley didapat perhitungan:

$$0 = \frac{(\mathrm{K}_3 - \mathrm{K}_2) - (\mathrm{K}_2 - \mathrm{K}_1)}{(\mathrm{K}_3 - \mathrm{K}_2) + (\mathrm{K}_2 - \mathrm{K}_1)}$$

$$\Leftrightarrow \quad (\mathrm{K}_3 - \mathrm{K}_2) - (\mathrm{K}_2 - \mathrm{K}_1) = 0$$

$$\Leftrightarrow \quad \mathrm{K}_3 - \mathrm{K}_2 = \mathrm{K}_2 - \mathrm{K}_1$$

Jadi distribusi simetri menurut Bowley jika selisih K_3 dengan K_2 sama dengan selisih K_2 dengan K_1.

c. Formula Momen

Mengingatkan kembali, disini momen yang digunakan merupakan ukuran dispersi data dengan derajat 3 (tiga).

Formula momen untuk data yang tidak digrupkan:

$$\text{Skewness} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^3}{nS^3}$$

Keterangan:

n = banyaknya data

X_i = data ke-i

$\bar{X}$ = mean

S = deviasi baku

Sedangkan untuk data yang digrupkan:

$$\text{Skewness} = \frac{\sum_{i=1}^{n} f_i (X_i - \bar{X})^3}{(\sum_{i=1}^{n} f_i\ ) S^3}$$

Keterangan:

n = banyaknya data

f_i = banyaknya data pada kelas ke-i

X_i = nilai tengah ke-i

$\bar{X}$ = mean

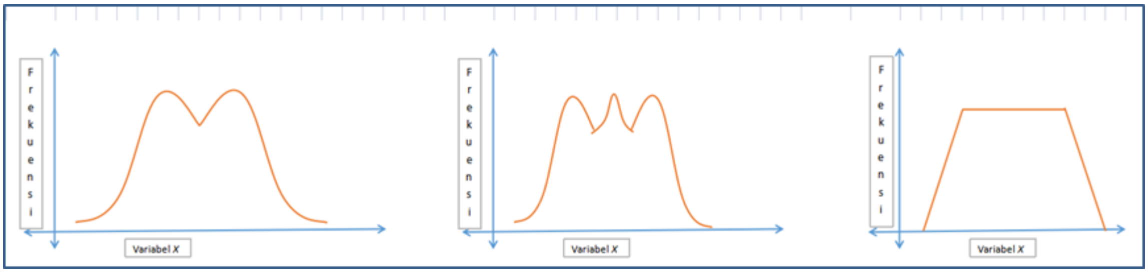
S = deviasi baku

D. PUNCAK

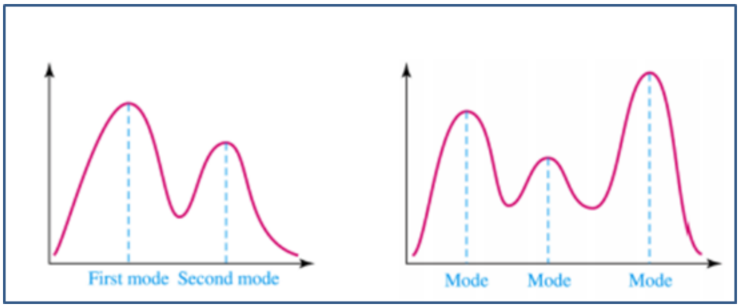
Dalam bentuk distribusi data, ada yang disebut puncak. Jelas puncak merupakan nilai dari modus.

Jika distribusi datanya mempunyai satu puncak, berarti nilai modus hanya ada satu, biasa dinamakan *unimode.* Untuk dua puncak, berarti dua nilai modus yaitu *dwimode* (Hadi, Sutrisno, 2015). Selanjutnya untuk bentuk lebih dari dua puncak atau nilai modusnya lebih dari pada dua, disebut *multiple mode*. Contohnya dapat dilihat di gambar 5.7.

Beberapa puncak dapat menjadi puncak 'mayor' dan beberapa bisa menjadi puncak 'minor'. Contohnya dapat dilihat di gambar 5.8.



Gambar 5.7 Contoh *Dwimode, Multiple Mode*



Gambar 5.8 Contoh Puncak Mayor, Minor

E. KURTOSIS

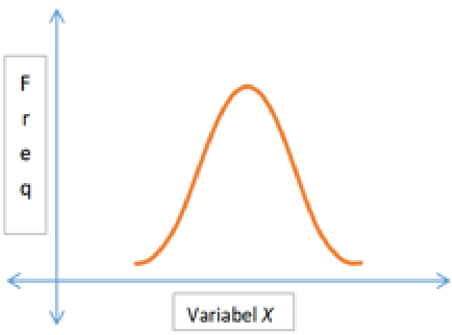
Kurtosis mengacu pada tingkat ketajaman dari puncaknya, atau kerataan di bagian atas pada distribusi simetri.

Distribusi simetri dengan model normal dikatakan sebagai distribusi normal.

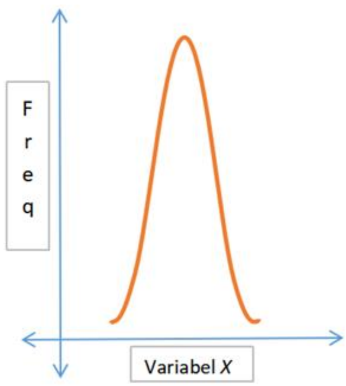
Adapun dalam buku Metoda Statistika oleh Sudjana (2005) dituliskan bahwa, bertitik tolak dari distribusi normal, tinggi rendahnya atau runcing datarnya bentuk kurva disebut kurtosis.

Ada dikenal 3 (tiga) kelompok tingkat ketajaman puncak pada distribusi simetri, yaitu:

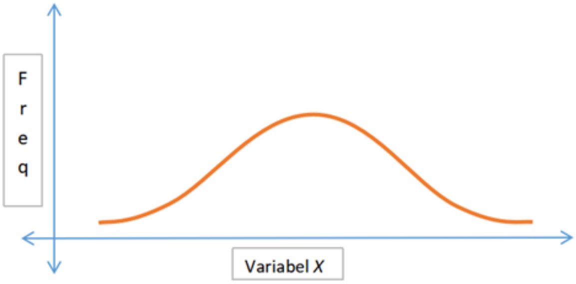
a. Tingkat ketajaman puncak yang normal, selain dikatakan sebagai distribusi normal, juga dapat disebut distribusi mesokurtik (Gambar 5.9).
b. Tingkat ketajaman puncak yang tinggi (*high*) disebut distribusi leptokurtik (Gambar 5.10).
c. Tingkat ketajaman puncak yang lebih rendah (*low*) dari yang normal atau puncaknya lebih datar disebut distribusi platikurtik (Gambar 5.11).



Gambar 5.9 Distribusi Mesokurtik



Gambar 5.10 Distribusi Leptokurtik



Gambar 5.11 Distribusi Platikurtik

Untuk perhitungan nilai kurtosis terdapat 2 (dua) cara:

a. Formula Momen

Istilah yang digunakan Supranto (2008) yaitu *Moment coefficient of kurtosis*, dengan menggunakan simbol α_4 .

Untuk data yang tidak digrupkan, formulanya:

$$\alpha_4 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^4}{nS^4}$$

Keterangan:

n = banyaknya data

X_i = data ke-i

$\bar{X}$ = mean

S = deviasi baku

Sedangkan untuk data yang digrupkan, formulanya:

$$\alpha_4 = \frac{\sum_{i=1}^{n} f_i(X_i - \bar{X})^4}{(\sum_{i=1}^{n} f_i\,)S^4}$$

Keterangan:

n = banyaknya data

f_i = banyaknya data pada kelas ke-i

X_i = nilai tengah ke-i

$\bar{X}$ = mean

S = deviasi baku

Selanjutnya kriteria nilai α_4 untuk masing-masing distribusinya adalah:

- $\alpha_4 = 3$, menyatakan distribusi mesokurtik (distribusi normal)
- $\alpha_4 > 3$, menyatakan distribusi leptokurtik
- $\alpha_4 < 3$, menyatakan distribusi platikurtik

b. Formula yang diistilahkan oleh Supranto (2008) yaitu *Quartile Coeffisient of Kurtosis (QCK),* adapun oleh Hooda (2013) disebut *Percentile Coeffisient of Kurtosis:*

$$QCK = \frac{\frac{1}{2}(K_3 - K_1)}{P_{90} - P_{10}}$$

Keterangan:

$K_1 =$ Kuartil satu

$K_3 =$ Kuartil tiga

$P_{10} =$ Persentil sepuluh

$P_{90} =$ Persentil sembilan puluh

Selanjutnya kriteria nilai QCK untuk masing-masing distribusinya adalah:

- $QCK = 0{,}263$, menyatakan distribusi mesokurtik (distribusi normal)
- $QCK > 0{,}263$, menyatakan distribusi leptokurtik
- $QCK < 0{,}263$, menyatakan distribusi platikurtik

F. CONTOH SOAL DAN PENYELESAIANNYA

Berdasarkan tabel data nilai ujian bergrup berikut, hitunglah *skewness* dan kurtosisnya dengan menggunakan formula momen.

NILAI UJIAN	Frekuensi
11 - 20	2
21 - 30	6
31 - 40	8
41 - 50	12
51 - 60	20
61 - 70	15
71 - 80	9
81 - 90	7
91 - 100	4

Penyelesaian:

NILAI UJIAN	Frekuensi	Nilai Tengah X_i	$f_i X_i$	$f_i (X_i - \bar{X})^2$	$f_i (X_i - \bar{X})^3$	$f_i (X_i - \bar{X})^4$
11 - 20	2	15,5	31	3515,87	-147412,21	6180656,42
21 - 30	6	25,5	153	6116,27	-195278,57	6234797,85
31 - 40	8	35,5	284	3846,60	-84347,05	1849537,62
41 - 50	12	45,5	546	1707,24	-20363,51	242890,01
51 - 60	20	55,5	1110	74,32	-143,27	276,18
61 - 70	15	65,5	982,5	977,43	7890,08	63691,00
71 - 80	9	75,5	679,5	2939,47	53122,93	960052,93
81 - 90	7	85,5	598,5	5516,37	154857,24	4347197,33
91 - 100	4	95,5	382	5798,00	220743,01	8404191,74
JUMLAH	**83**		**4766,5**	**30491,57**	**-10931,34**	**28283291,09**
Mean	**57,43**		**S =**	**18,30**	**6133,49**	**112273,35**

Formula dari *skewness* yaitu:

$$\text{Skewness} = \frac{\sum_{i=1}^{n} f_i (X_i - \bar{X})^3}{(\sum_{i=1}^{n} f_i)S^3}$$

Dari tabel penyelesaian didapat informasi bahwa

$$\sum_{i=1}^{n} f_i (X_i - \bar{X})^3 = -10931{,}34$$

$$\sum_{i=1}^{n} f_i = 83$$

$$S^3 = 6133{,}49$$

Jadi Skewness $= \frac{\sum_{i=1}^{n} f_i (X_i - \bar{X})^3}{(\sum_{i=1}^{n} f_i)S^3} = \frac{-10931{,}34}{83 * 6133{,}49} = -0{,}021472756$

Karena nilai *skewness* mendekati 0, maka distribusi datanya simetri.

Selanjutnya menghitung nilai kurtosisnya dengan formula momen, yaitu:

$$\alpha_4 = \frac{\sum_{i=1}^{n} f_i (X_i - \bar{X})^4}{(\sum_{i=1}^{n} f_i)S^4}$$

Dari tabel penyelesaian didapat informasi bahwa

$$\sum_{i=1}^{n} f_i (X_i - \bar{X})^4 = 28283291{,}09$$

$$\sum_{i=1}^{n} f_i = 83$$
$S^4 = 112273,35$

Jadi $\alpha_4 = \frac{\sum_{i=1}^{n} f_i (X_i - \bar{X})^4}{(\sum_{i=1}^{n} f_i) S^4} = \frac{28283291,09}{83 * 112273,35} = 3,035115123$

Karena nilai α_4 dekat 3, maka data membentuk distribusi mesokurtis (distribusi normal).

BAB

6

UKURAN KETERKAITAN (KORELASI DAN REGRESI) PARAMETRIK DAN *NON* PARAMETRIK

A. PENGANTAR

Statistika inferensi atau *statistics inference* merupakan suatu metode statistika yang dipergunakan untuk menganalisis data sampel dan hasil perhitungannya diberlakukan secara umum. Adapun teknik pengambilan sampel dari populasi tersebut dilakukan secara random. (Sugiono, 2012 dalam (Anggraeni, Dewi, & Rio Satriyantara, 2021)

Statistika inferensi terdiri dari dua jenis yaitu statistika parametrik dan *non*-parametrik.

B. KORELASI DAN REGRESI PARAMETRIK

1. Korelasi Parametrik

Korelasi parametrik umumnya digunakan pada data yang memenuhi asumsi normalitas

Suatu bentuk hubungan antara satu variabel terikat (*dependent variable*) dengan dua atau lebih variabel bebas (*independent variable*) disebut korelasi. (Usman & Akbar, 2020)

Variabel terikat sering dinotasikan dengan Y meupakan variabel yang dipengaruhi, sedangkan variabel bebas dinotasikan dengan X diartikan sebagai variabel yang mempengaruhi.

Korelasi sebagai suatu analisis mempunyai berbagai jenis menurut tingkatannya, yaitu (UCEO, 2016)

a. Korelasi *Pearson*

Korelasi *Pearson* yang paling cocok digunakan untuk data kontinyu maupun diskrit yaitu korelasi *Pearson Product Moment* atau *PPM*

Secara umum, rumus korelasi *Pearson* adalah:

$$r_{xy} = \frac{n \sum Y_i X_i - (\sum Y_i)\,(\sum X_i)}{\sqrt{\{n \sum Y_i^2 - (\sum X_i)^2\}}\sqrt{\{n \sum Y_i^2 - (\sum X_i)^2\}}}$$

Interpretasi terhadap nilai r adalah sebagai berikut:

a) 0,0 : tidak ada hubungan
b) 0,0 - 0,20 : hubungan sangat rendah
c) 0,21 - 0,40 : rendah
d) 0,41 – 0,60 : cukup
e) 0,61 – 0,80 : kuat
f) 0,81 - 1,00 : sangat kuat

b. Korelasi Parsial

Korelasi parsial digunakan untuk melihat besarnya hubungan antara dua variabel bebas dengan variabel terikatnya. Korelasi parsial dilambangkan dengan $r_{yx1.x2}$ yang menyatakan hubungan antara Y dengan X_1 dimana X_2 dianggap tetap, $r_{yx2.x1}$, yang menyatakan hubungan antara Y dengan X_2 dimana X_1 dianggap tetap, serta , $r_{x1x2.y}$ yang menyatakan hubungan antara X_1 dengan X_2 dimana Y dianggap tetap. (Purwanto, 2016)

Rumus korelasi parsial untuk tiga variabel adalah sebagai berikut:

$$r_{YX1.X2} = \frac{r_{YX1} - r_{YX2} r_{X1X2}}{\sqrt{1 - r_{YX2}^2}\sqrt{1 - r_{X1X2}^2}}$$

$$r_{YX2.X1} = \frac{r_{YX2} - r_{YX1} r_{X1X2}}{\sqrt{1 - r_{YX1}^2}\sqrt{1 - r_{X1X2}^2}}$$

$$r_{X1X2.Y} = \frac{r_{X1X2} - r_{YX1} r_{YX2}}{\sqrt{1 - r_{YX1}^2}\sqrt{1 - r_{YX2}^2}}$$

c. Korelasi Berganda

Korelasi berganda merupakan alat ukur mengenai hubungan yang terjadi antara variabel terikat (variabel Y) dengan dua atau lebih variabel bebas (X_2 ,X_3.....X_k). (Hasan, 2017)

Rumus korelasi berganda untuk dua variabel bebas (X_1 dan X_2) adalah sebagai berikut : (Usman & Akbar, 2020)

$$R_y X_1 X_2 = \sqrt{\frac{r_{yx1}^2 + r_{yx2}^2 - 2r_{yx1}\, r_{yx2} r_{x1x2}}{1 - r_{x1x2}^2}}$$

dimana: $R_y X_1 X_2$ adalah koefisien korelasi berganda antara variabel X_1 dan X_2 secara bersama sama dengan variabel Y

r_{yx1} adalah koefisien korelasi variabel X_1 dan Y

r_{yx2} adalah koefisien korelasi variabel X_2 dan Y

r_{x1x2} adalah koefisien korelasi variabel X_1 dan X_2

2. Regresi Parametrik

Regresi parametrik dapat digunakan apabila bentuk kurve$\int(x_i$ diketahui. Ada beberapa jenis regresi parametrik yaitu : (Yusi & Idris, 2019)

a. Regresi linier sederhana

Hubungan antara dua variabel yang jika digambarkan secara grafis (*scatter diagram*), maka semua nilai X dan Y akan berada pada suatu garis lurus, yang dalam ilmu ekonomi disebut *garis regresi*. Oleh karena hubungan ini hanya melibatkan dua variabel dan berpangkat satu, maka disebut regresi linier sederhana. (Hasan, 2017)

Persamaan regresi linier sederhana yaitu : (Mulyono, 2017)

$$\hat{Y}t = a + bXi + \mu$$

dimana:

$\hat{Y}t$ adalah penduga terhadap ∈ (Y|X)

a adalah penduga terhadap A

b adalah penduga terhadap B

μ adalah error term

Nilai a dan b diperoleh dengan rumus:

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

$$a = \frac{\sum Y - b\sum X}{n}$$

Suatu regresi linier dapat disebut model regresi linier klasik apabila memenuhi asumsi-asumsi sebagai berikut:

a) Error term μ memiliki disribusi normal
b) Rata-rata atau *expected value* dari error term untuk setiap nilai X sama dengan nol.
c) Varians error term adalah konstan pada setiap periode dan untuk semua nilai x_i
d) Error term pada suatu observasi tidak berhubungan dengan error term pada observasi yang lain.
e) Variabel bebas merupakan variabel *non*-stokastik sehingga variabel bebas tidak berhubungan dengan error term.

b. Regresi *Non*-linier Sederhana

Regresi *non*-linier sederhana adalah suatu pola hubungan yang berbentuk garis tidak lurus antara satu variabel terikat dengan satu variabel bebas.

Ada beberapa bentuk pola hubungan garis tidak lurus antara lain:

a) Model parabola kuadratik

Persamaan umumnya adalah:

$$\hat{Y} = a + bX + cX^2$$

dimana $\hat{Y}$ adalah variable yang diprediksi, X adalah variable bebas, sedangkan a, b dan c adalah parameter atau koefisien regresi.

Untuk menentukan nilai koefisien regresi dapat digunakan *least square* yang meminimalisasikan kesalahan prediksi.

Rumus umum *least square* untuk regresi kuadratik adalah:

$$\sum Y = na + b\sum X + c\sum X^2$$

$$\sum XY = a\sum X + b\sum X^2 + c\sum X^3$$

$$\sum X^2 Y = a\sum X^2 + b\sum X^3 + c\sum X^4$$

b) Model Eksponensial

Persamaan model ini adalah

$$\hat{Y} = ab^X$$

Model eksponensial juga disebut dengan model pertumbuhan karena sering digunakan dalam menganalisis data hasil pengamatan mengenai fenomena yang bersifat tumbuh.

$$\hat{Y} = \ ae^{bX}$$

dimana e = merupakan bilang pokok logaritma asli, dengan nilai e = 2,7183

c) Model parabola kubik

Model ini mempunyai persamaan umum sebagai berikut:

$$\hat{Y} = a + bX + cX^2 + dX^3$$

Sementara itu, persamaan untuk menghitung parameter atau koefisien regresi adalah sebagai berikut:

$$\sum Y = na + b\sum X + c\sum X^2 + d\sum X^3$$

$$\sum XY = a\sum X + b\sum X^2 + c\sum X^3 + d\sum X^4$$

$$\sum X^2 Y = a\sum X^2 + b\sum X^3 + c\sum X^4 + d\sum X^5$$

$$\sum X^3 Y = a\sum X^3 + b\sum X^4 + c\sum X^5 + d\sum X^6$$

Makin tinggi pangkat X dalam persamaan tersebut , makin banyak model persamaan yang harus diselesaikan.

d) Model Geometrik

Model ini mempunyai persamaan umum:

$$Y = aXb$$

Jika dipergunakan persamaan logaritma, maka diperoleh persamaan:

$$Log\ Y = log\ a + b\ log\ X$$

Nilai koefisien regresi dapat dihitung dengan rumus:

$$loga = \frac{\sum logY}{n} - b\frac{\sum logX}{n}$$

$$logb = \frac{n(\sum logX.logY) - (\sum logX)(\sum logY)}{n(\sum log^2 X) - (\sum logX)^2}$$

c. Regresi Linier Berganda

Menurut Simon dalam (Simamora, 2023) regresi linier berganda digunakan apabila variabel bebas (*independent*) terdiri dari dua atau lebih. Persamaan regresi untuk k variabel adalah:

$$\hat{Y}i = \beta_0 + \beta_1(X_1) + \beta_2(X_2) + \beta_3(X_3) + \cdots + \beta_k(X_k)$$

Keberhasilan regresi tergantung pada seberapa dekat nilai prediksi *Y* ($\hat{Y}1, \hat{Y}2, \hat{Y}3 \ldots \hat{Y}n$) dengan nilai observasi *Y* ($Y_1, Y_2, Y_3 \ldots Y_n$) atau seberapa kecil *error* kuadrat.

Adapun syarat-syarat regresi linier berganda adalah: Rawling *et al* (1991) dalam (Simamora, 2023)

a) Error atau residual berdistribusi normal

Apabila tidak normal, maka langkah yang dilakukan adalah:

- Menambah jumlah sampel, karena semakin besar jumlah sampel, semakin cenderung kedistribusi normal
- Mendeteksi keberadaan data *outlier* yang menyebabkan distribusi tidak normal
- Melakukan transformasi regresi linier menjadi regresi *polynomial*.
- Melakukan transformasi regresi linier menjadi regresi logaritmik. Ada tiga pilihan model yaitu *log-linier*, *linier-log*,dan *log-log*.

b) Tidak terdapat multikolinearitas

Dampak yang ditimbulkan adanya multikolinearitas adalah:

- Terjadi peningkatan varians dan *standard error* koefisien regresi estimator
- Terdapat satu atau lebih *variable* independen berpengaruh negatif
- Penambahan atau pengurangan variabel *independent* akan menyebabkan perubahan pada koefisien estimasi dan tandanya.
- Pengurangan data akan menyebabkan perubahan besar pada koefisien estimasi
- Dapat terjadi nilai *F* signifikan, tetapi nilai *t* tidak signifikan.

c) Tidak terjadi heteroskedastisitas

Heteroskedastisitas dapat menyebabkan gangguan pada temuan dan kemungkinan terjadi kesalahan Tipe 1(*Type 1 Error*).

d) Tidak terjadi otokorelasi

otokorelasi biasanya terjadi pada data *time series* atau data runtut waktu. Akan tetapi oto-korelasi dapat juga terjadi pada data tidak runtut waktu, yaitu pada data satu titik waktu (*single cross sectional*)

C. KORELASI DAN REGRESI *NON*-PARAMETRIK

1. Korelasi *Non*-Parametrik

Korelasi *non*-parametrik terdiri beberapa macam antara lain:

a. Koefisien Korelasi Rank *Spearman*

Menurut (Yusi & Idris, 2019) koefisien korelasi *Spearman* adalah yang pertama kali dikembangkan dan diperkirakan yang paling banyak diketahui saat ini. Nilai statistiknya kadang-kadang disebut dengan *rho* disimbolkan *rs*. Karena data berskala ordinal maka objek-objek dalam individu-individu yang diteliti harus dirank (diurutkan) dari kedua gugusnya.

Rumus korelasi *spearman* untuk jumlah sampel < = 30 adalah:

$$rs = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

Keterangan:

Rs = koefisien korelasi *Spearman*

d = selisih antara X dan Y

n = jumlah sampel penelitian

6 = angka konstan

b. Koefisien Korelasi Rank *Tau Kendall*

Koefisien korelasi rank *Kendall* $\tau(ta)$ mempunyai kegunaan untuk mengukur derajat hubungan dari dua variabel seperti halnya dengan r_s , dengan syarat minimal pengukurannya dalam skala ordinal bagi kedua variabel tersebut ((Yusi & Idris, 2019)

Rumus korelasi rank *Kendall* adalah :

$$r = \frac{S}{½(N-1)}$$

dimana: S = total skor

N = jumlah objek atau individu yang di rank

c. Koefisien Korelasi *Phi*

Koefisien *Phi* (kadang-kadang disebut koefisien kontingensi kuadrat rata-rata) adalah ukuran hubungan antara dua variabel biner.((Zach, 2020)

Rumus korelasi *Phi* adalah:

$$\phi = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

Koefisien *Phi* mengambil nilai antara -1 dan 1 di mana: (Zach, 2020)

a) -1 menunjukkan hubungan negatif sempurna antara kedua variabel.
b) 0 menunjukkan tidak ada hubungan antara kedua variabel.
c) 1 menunjukkan hubungan positif sempurna antara kedua variabel.

d. Koefisien Kontingensi

Salah satu uji statistik *non* parametrik yang bertujuan untuk mengetahui keeratan hubungan (korelasi) antara dua variabel atau lebih dan biasanya digunakan untuk jenis data berskala nominal adalah Koefisien Kontigensi atau *Contigency Coefficient* C. (jagostatitsik, 2022)

Rumus koefisien kontingensi adalah sebagai berikut: (Hidayat, 2023)

$$C = \sqrt{\frac{X^2}{N + X^2}}$$

dimana:

$$X^2 = \sum_{i=1}^{h} \sum_{j=1}^{k} \frac{(A_{ij} - H_{ij)})^2}{H_{ij}}$$

2. Regresi *Non*-Parametrik

Metode regresi *non* parametrik mulai di kenal sejak abad XIX, yaitu pada saat Engel, seorang ekonom telah mengkonstruksi suatu kurve yang dikenal sebagai regressogram. (Abdy, 2019). Perkembangan regresi *non*-parametrik tidak terlalu pesat seperti regresi parametrik. Akan tetapi sejak beberapa dekade terakhir, regresi *non*-parametrik berkembang pesat seiring dengan perkembangan teknologi komputer.

Apabila bentuk dari kurve regresi tidak diketahui maka digunakan regresi *non*-parametrik. Dalam regresi *non*-parametrik kurva regresi diasumsikan mulus (*smooth*) dalam arti bahwa kurve tersebut termuat dalam suatu ruang fungsi tertentu sehingga mempunyai sifat fleksibilitas yang tinggi (Syam, Sanusi, & Adwiyah, 2019). Dengan demikian, teknik regresi *non*-parametrik dapat menjadi alternatif karena penggunaannya tidak terikat pada asumsi-asumsi seperti dalam regresi parametrik.

Pendekatan regresi *non*-parametrik yang telah dikembangkan antara lain *spline*, *kernel*, *polinomial lokal*, *wavelet*, dan *fourier*. Salah satu model regresi *non*-parametrik yang sering digunakan untuk melakukan estimasi terhadap kurva regresi adalah Regresi *Spline* dan *Kernel*.

a. *Spline*

Spline merupakan suatu model polinom (*piecewise polynomial*) yang tersegmen atau terpotong-potong yang mulus dan dapat menghasilkan fungsi regresi yang sesuai dengan data (Alwi, Muh.Irwan, & Musfirah, 2021). Menghitung estimasi *Spline* tergantung pada titik *knot*, dimana titik *knot* merupakan suatu titik perpaduan yang terjadi karena perubahan pola perilaku dari suatu fungsi pada selang yang berbeda.

Bentuk umum dari fungsi *Spline* linier dengan *r* orde dan *m* titik knot adalah sebagai berikut : (Kurniasari, Kusnandar, & Sulistianingsih, 2019)

$$f x_i = \sum_{q=0}^{r} \beta_q \ x_i^q + \sum_{j=1}^{m} \beta_{(r+j)} \ (x_i - K_j)_+^r$$

Dimana $f(x_i)$ adalah fungsi dari variabel ke i , β_q dan $\beta_{(r+j)}$ menunjukkan koefisien regresi *Spline* dengan q = 1,2,…,r dan j=1,2,..,m serta $(x_i - K_j)_+^r$ yang menunjukkan fungsi linier pada variabel ke i dengan titik knot j pada orde ke r.

Metode yang digunakan untuk mencari titik knot optimal yaitu *Generelized Cross Validation* (*GCV*) dan *Mean Square Error* (*MSE*).

b. *Kernel*

Regresi *Kernel* merupakan suatu teknik statistik *non*-parametrik untuk menaksir nilai ekspektasi bersyarat dari suatu variabel random dengan tujuan untuk memperoleh hubungan *non*-linier antara *variable* X dan Y. (Wulandary & Purnama, 2020)

Proses penghalusan (*smoothing*) yang dikenal dengan penghalusan *Kernel* (*Kernel Smoothing*) lebih ditekankan pada penerapan regresi *Kernel* ini. Proses penghalusan ini sangat tergantung pada fungsi *Kernel* dan *bandwidth* yang akan menentukan taksiran kepadatan.

Tingkat pengaruh titik-titik pada penentuan lokasi dapat dikontrol melalui parameter pemulus atau *bandwidth Kernel* yang dinyatakan dengan notasi *h*.

Pada tahun 1964 Nadaraya dan Watson mempublikasikan metode penaksir m(.) yang disebut metode *Nadaraya-Watson Estimator (NEW)* dengan rumus sebagai berikut: Härdle (1994) dalam (Wulandary & Purnama, 2020)

$$\widehat{m}\,(X) = \frac{\sum_{i=1}^{n} K_h\ (X_i - x)Y_i}{\sum_{i=1}^{n} K_h\,(X_i - x)}$$

Penaksir $\widehat{\theta}$ yang meminimumkan pada *X fixed* adalah*:*

$$\sum_{i=1}^{n} (Y_i - \theta)^2\ K_h\,(X_i - x)$$

Memiliki bentuk $\sum_{i=1}^{n} a_i\ Y_i$, *NEW* merupakan *minimizer* persamaan diatas dengan:

$$a_i\ = \frac{K_h - (X_i\ - x)}{\sum_{i=1}^{n} K_h\ (X_i\ - x)}$$

D. PENUTUP

Statistik parametrik digunakan apabila informasi tentang populasi tersedia pada saat melakukan penelitian dan penggunaan jenis minimal yaitu data interval dan data rasio.

Sedangkan statistik *non*-parametrik dapat digunakan pada (1) data yang distribusi populasinya tidak diketahui, (2) data terdistribusi tidak normal, (3) data tidak diambil secara random, (4) data berskala ordinal atau nominal, dan (5) jumlah data <30.

BAB

7

PENGANTAR TEORI KEMUNGKINAN

A. PENGANTAR

Kita selalu membuat keputusan berdasarkan ketidakpastian setiap hari. Ketika Anda membeli iPhone baru, apakah Anda juga membeli perpanjangan masa garansinya? Itu tergantung pada kemungkinan terjadi kerusakan selama masa garansi. Selanjutnya apakah Anda membutuhkan waktu 45 menit untuk sampai ke kelas jam 8 pagi, atau apakah cukup 35 menit saja? Dari pengalaman, Anda mungkin tahu bahwa hampir setiap pagi Anda dapat berkendara ke kampus dan parkir dalam 25 menit atau kurang. Waktu yang diperlukan untuk berjalan kaki dari tempat parkir Anda ke kelas adalah 5 menit atau kurang. Tapi seberapa sering perjalanan ke kampus atau jalan kaki ke kelas memakan waktu lebih lama dari yang Anda mengharapkan? Ketika dibutuhkan waktu lebih lama dari biasanya untuk berkendara ke kampus, apakah kemungkinan besar hal itu akan terjadi juga membutuhkan waktu lebih lama untuk berjalan ke kelas? lebih kecil kemungkinannya? Atau apakah waktu mengemudi dan berjalan tidak berhubungan?

Beberapa pertanyaan melibatkan ketidakpastian yang lebih serius: Jika jantung buatan memiliki empat bagian utama, seberapa besar kemungkinan masing-masing gagal? Seberapa besar kemungkinan bahwa setidaknya satu akan gagal? Kita dapat menjawab pertanyaan seperti ini dengan menggunakan metode probabilitas yaitu studi sistematis tentang ketidakpastian.

B. PENGERTIAN

Probabilitas adalah cabang matematika yang mempelajari peluang dan kejadian acak. Ini adalah alat yang digunakan untuk mengukur sejauh mana suatu kejadian mungkin terjadi atau tidak terjadi. Dalam kehidupan sehari-hari, kita sering menghadapi situasi di mana kita tidak dapat memprediksi hasilnya dengan pasti, dan itulah saat probabilitas menjadi relevan. Pemahaman tentang probabilitas dapat membantu kita membuat keputusan yang lebih baik, merencanakan strategi, menganalisis risiko, dan menginterpretasikan data yang tidak pasti. Dalam dunia ilmu pengetahuan dan ilmu terapan, probabilitas diterapkan dalam berbagai bidang, seperti statistik, ilmu komputer, ekonomi, fisika, dan banyak lagi.

Beberapa konsep penting dalam probabilitas meliputi:

- **Eksperimen Acak (*Random Experiment*):** Ini adalah situasi di mana hasilnya tidak dapat diprediksi dengan pasti. Contohnya, melempar dadu adalah eksperimen acak karena kita tidak tahu hasilnya sebelumnya.
- **Kejadian (*Event*):** Ini adalah hasil atau hasil tertentu dari eksperimen acak. Kejadian dapat berupa satu kejadian tunggal atau kombinasi beberapa kejadian.
- **Ruang Sampel (*Sample Space*):** Ini adalah himpunan dari semua hasil yang mungkin dari eksperimen acak. Misalnya, jika kita melempar sebuah koin, ruang sampelnya akan terdiri dari "H" (kepala) dan "T" (ekor).
- **Peluang (*Probability*):** Ini adalah ukuran seberapa besar kemungkinan suatu kejadian terjadi. Peluang dinyatakan sebagai bilangan antara 0 (tidak mungkin terjadi) dan 1 (pasti terjadi). Peluang juga dapat dinyatakan dalam bentuk persentase.
- **Hukum Peluang (*Laws of Probability*):** Ada beberapa hukum dan prinsip dalam probabilitas yang membantu dalam menghitung dan menganalisis peluang berbagai kejadian. Beberapa hukum ini termasuk hukum penjumlahan, hukum perkalian, dan hukum komplementer.
- **Distribusi Peluang (*Probability Distribution*):** Ini adalah cara untuk menggambarkan peluang berbagai hasil dalam eksperimen acak. Distribusi peluang dapat dinyatakan dalam bentuk tabel, grafik, atau persamaan matematika.

Dalam probabilitas, kita dapat menerapkan konsep-konsep ini untuk menghitung peluang suatu kejadian, menganalisis eksperimen acak, membuat keputusan berdasarkan risiko, dan banyak lagi. Probabilitas memiliki dampak signifikan dalam berbagai aspek kehidupan kita dan membantu kita menghadapi ketidakpastian dengan cara yang lebih terstruktur dan ilmiah.

C. JENIS PROBABILITAS

Terdapat beberapa jenis probabilitas yang digunakan dalam berbagai konteks. Berikut adalah beberapa jenis probabilitas yang umum:

1. Probabilitas Klasik (*Classical Probability*)

Probabilitas klasik, juga dikenal sebagai probabilitas *a priori*, adalah jenis probabilitas yang mendasarkan perhitungan peluang pada penghitungan yang sama atau homogen. Ini adalah pendekatan yang paling sederhana dalam mengukur peluang dan biasanya digunakan dalam situasi di mana semua hasil yang mungkin memiliki peluang yang sama. Konsep probabilitas klasik didasarkan pada prinsip bahwa semua hasil dalam himpunan sampel memiliki peluang yang sama untuk terjadi. Ini sering digunakan dalam kasus eksperimen acak dengan hasil yang dapat dihitung secara teoritis. Contohnya termasuk melempar dadu adil atau mengambil kartu dari setumpuk kartu yang dikocok dengan baik.

Contoh penerapan probabilitas klasik:

- **Melempar Dadu:** Dalam melempar dadu adil, ada 6 hasil yang mungkin (angka 1 hingga 6), dan setiap hasil memiliki peluang 1/6.
- **Mengambil Kartu dari Dek Kartu:** Dalam setumpuk kartu yang terdiri dari 52 kartu, setiap kartu memiliki peluang 1/52 untuk ditarik.
- **Mengambil Bola dari Wadah:** Dalam wadah yang berisi bola-bola berwarna yang identik secara visual, jika semua bola memiliki peluang yang sama untuk ditarik, maka probabilitas mengambil bola tertentu adalah 1/jumlah total bola.

Namun, penting untuk diingat bahwa probabilitas klasik hanya berlaku dalam situasi di mana semua hasil memiliki peluang yang sama. Dalam situasi yang lebih kompleks, seperti dengan peluang variabel atau kejadian yang tidak homogen, metode probabilitas lainnya mungkin lebih sesuai, seperti probabilitas frekuensi relatif atau probabilitas subyektif.

2. Probabilitas Frekuensi Relatif (*Relative Frequency Probability*)

Probabilitas frekuensi relatif adalah jenis probabilitas yang dihitung berdasarkan pengamatan atau pengukuran empiris dari kejadian dalam sejumlah percobaan. Ini adalah pendekatan yang mengandalkan data nyata untuk mengestimasi peluang suatu kejadian terjadi. Probabilitas frekuensi relatif didefinisikan sebagai rasio antara jumlah kejadian yang dianggap sukses (muncul dalam pengamatan) dengan jumlah total percobaan.

Formula untuk menghitung probabilitas frekuensi relatif dari suatu kejadian adalah:

$$\mathrm{P(A)} = \frac{Jumlah\ kejadian\ sukses}{Jumlah\ total\ percobaan}$$

Di mana:

- P(A) adalah probabilitas frekuensi relatif dari kejadian A.
- "Jumlah kejadian sukses" adalah jumlah kali kejadian A muncul dalam percobaan.
- "Jumlah total percobaan" adalah jumlah total percobaan atau pengamatan yang dilakukan.

Probabilitas frekuensi relatif menjadi semakin akurat ketika jumlah percobaan meningkat. Dalam situasi di mana probabilitas teoritis tidak diketahui atau tidak dapat dihitung, probabilitas frekuensi relatif adalah cara untuk memperkirakan peluang berdasarkan data empiris.

Contoh penerapan probabilitas frekuensi relatif: Dalam pengamatan 100 lemparan koin, kepala (H) muncul sebanyak 45 kali. Maka probabilitas frekuensi relatif munculnya kepala dalam lemparan koin adalah:

$$\mathrm{P(H)} = \frac{45}{100}$$

Ini berarti dalam pengamatan ini, peluang munculnya kepala dalam lemparan koin adalah sekitar 0.45 atau 45%. Namun, perlu diingat bahwa probabilitas frekuensi relatif mungkin berbeda dalam setiap pengamatan berbeda. Hal ini dapat menghasilkan perkiraan yang lebih baik jika banyak pengamatan dilakukan.

3. Probabilitas Subyektif (*Subjective Probability*)

Probabilitas subyektif adalah jenis probabilitas yang didasarkan pada keyakinan atau penilaian subjektif seseorang. Ini adalah pendekatan yang digunakan ketika tidak ada data empiris yang tersedia untuk mengukur peluang, atau ketika situasinya sangat kompleks sehingga sulit untuk menghitung peluang secara objektif. Dalam probabilitas subyektif, seseorang mengestimasi peluang suatu kejadian berdasarkan pengetahuannya, pengalaman, intuisi, atau informasi yang ada. Ini bisa sangat bervariasi antara individu dan sering kali tidak memiliki dasar matematis yang kuat. Namun, probabilitas subyektif dapat berguna dalam situasi di mana pendekatan matematis yang lebih formal tidak memungkinkan atau relevan.

Contoh penerapan probabilitas subyektif:

- **Perdagangan Saham:** Seorang pedagang saham mungkin mengestimasi peluang naik atau turunnya harga saham berdasarkan pengetahuannya tentang industri, tren pasar, dan berita terkait.
- **Prediksi Cuaca:** Seseorang dapat memberikan peluang subjektif tentang apakah hari berikutnya akan cerah atau berawan berdasarkan pengalaman cuaca sebelumnya dan informasi cuaca saat ini.
- **Keberhasilan Proyek:** Manajer proyek dapat memberikan estimasi subjektif tentang peluang keberhasilan atau kegagalan proyek berdasarkan pengetahuan mereka tentang tim, sumber daya, dan risiko yang terlibat.

Probabilitas subyektif dapat menjadi alat yang berguna dalam mengambil keputusan di saat ketidakpastian, tetapi juga dapat terpengaruh oleh bias, persepsi pribadi, atau informasi yang terbatas. Karena itu, dalam situasi di mana data empiris dapat diakses, pendekatan lain seperti probabilitas frekuensi relatif atau probabilitas teoritis lebih dapat diandalkan untuk mengukur peluang secara objektif.

4. Probabilitas Gabungan (*Joint Probability*)

Probabilitas gabungan, atau *joint probability*, mengacu pada peluang terjadinya dua atau lebih kejadian bersamaan. Ini adalah probabilitas bahwa dua atau lebih kejadian yang berbeda terjadi dalam satu percobaan atau situasi. Probabilitas gabungan dinyatakan sebagai hasil dari perkalian probabilitas masing-masing kejadian yang terlibat. Jika A dan B adalah dua kejadian yang berbeda, probabilitas gabungan $P(A \cap B)$ dihitung sebagai:

$$P(A \cap B) = P(A) \times P(B|A)$$

Di mana:

- $P(A)$ adalah probabilitas terjadinya kejadian A.
- $P(B|A)$ adalah probabilitas terjadinya kejadian B jika kejadian A telah terjadi.

Probabilitas gabungan dapat diterapkan dalam berbagai konteks, seperti statistik, teori probabilitas, ilmu sosial, dan lain-lain. Ini membantu kita memahami peluang terjadinya kombinasi kejadian yang berbeda.

Contoh penerapan probabilitas gabungan:

- **Melempar Dadu**: Misalkan kita ingin mengetahui peluang mendapatkan angka genap (kejadian A) dan peluang mendapatkan angka lebih dari 3 (kejadian B) dalam satu lemparan dadu. Jika dadu adalah adil, maka $P(A) = \frac{3}{6}$ dan $P(B) = \frac{3}{6}$. Jika kejadian A terjadi (mendapatkan angka genap), peluang mendapatkan angka lebih dari 3 (kejadian B) adalah $P(B|A) = 32$. Probabilitas gabungan $P(A \cap B)$ adalah $\frac{3}{6} x \frac{2}{3} = \frac{1}{3}$
- **Peluang Penyakit**: Dalam uji medis, A mewakili hasil positif tes untuk suatu penyakit dan B mewakili penyakit yang sebenarnya. Probabilitas bahwa tes positif (A) dan individu memiliki penyakit (B) dalam satu waktu dapat dihitung berdasarkan probabilitas kesalahan positif dan probabilitas penyakit sebenarnya.

Probabilitas gabungan adalah alat penting dalam analisis peluang dan membantu kita memahami bagaimana kejadian-kejadian yang berbeda saling berhubungan dalam situasi yang kompleks.

5. Probabilitas Tersyarat (*Conditional Probability*)

Probabilitas tersyarat, atau *conditional probability*, mengukur peluang terjadinya suatu kejadian tertentu berdasarkan informasi atau kejadian lain yang telah terjadi. Dalam probabilitas tersyarat, kita ingin menghitung peluang terjadinya kejadian A jika kita tahu bahwa kejadian B telah terjadi. Ini dinyatakan sebagai $P(A|B)$, yang dibaca sebagai "peluang A diberikan B".

Rumus umum untuk menghitung probabilitas tersyarat adalah:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Di mana:

- $P(A \cap B)$ adalah probabilitas gabungan atau peluang terjadinya kedua kejadian A dan B bersamaan.
- $P(B)$ adalah probabilitas terjadinya kejadian B.

Probabilitas tersyarat sangat penting dalam situasi di mana kejadian-kejadian tidak terjadi secara independen, dan informasi tentang kejadian yang telah terjadi dapat mempengaruhi peluang kejadian lain.

Contoh penerapan probabilitas tersyarat:

- **Uji Medis**: Misalkan kita ingin menghitung peluang bahwa seseorang benar-benar memiliki penyakit (A) jika hasil tesnya positif (B). Probabilitas ini akan sangat bergantung pada akurasi tes dan prevalensi penyakit di populasi.
- **Pandemi**: Dalam kasus pandemi, A mewakili seseorang terinfeksi virus, dan B mewakili hasil positif tes. Peluang bahwa seseorang benar-benar terinfeksi (A) jika hasil tesnya positif (B) akan tergantung pada tingkat akurasi tes dan prevalensi infeksi di populasi.

Probabilitas tersyarat membantu kita membuat keputusan berdasarkan informasi yang ada dan memahami bagaimana kejadian yang satu dapat mempengaruhi peluang kejadian yang lain.

6. Probabilitas Acak (*Random Probability*)

Probabilitas acak atau *random probability* merujuk pada konsep probabilitas dalam konteks acak atau kejadian tak terduga. Dalam statistika dan teori probabilitas, probabilitas menggambarkan seberapa mungkin suatu kejadian akan terjadi. Dalam konteks "*random probability*," istilah ini mungkin merujuk pada probabilitas yang terkait dengan hasil yang dihasilkan secara acak atau kebetulan.

Sebagai contoh, jika Anda melempar dadu yang adil (setiap sisi memiliki peluang yang sama), probabilitas munculnya masing-masing angka (1 sampai 6) adalah 1/6, atau sekitar 16.67%. Ini adalah contoh probabilitas acak karena hasilnya akan bervariasi secara acak setiap kali Anda melempar dadu.

Penting untuk memahami konsep probabilitas ini dalam berbagai konteks, termasuk dalam analisis data, peramalan, pengambilan keputusan, dan banyak bidang lainnya di mana tidak mungkin untuk memprediksi hasil dengan pasti, tetapi kita dapat memahami seberapa mungkin suatu kejadian akan terjadi berdasarkan probabilitasnya.

7. Probabilitas Diskrit (*Discrete Probability*)

Probabilitas diskrit adalah jenis probabilitas yang terkait dengan himpunan hasil yang terhitung atau dapat dihitung. Ini mengacu pada situasi di mana ruang sampel hanya terdiri dari sejumlah terbatas hasil yang bisa dihitung satu per satu. Dalam probabilitas diskrit, kita mengukur peluang terjadinya setiap hasil individual. Contoh probabilitas diskrit termasuk melempar dadu, mengambil kartu dari dek kartu, atau memilih anggota dalam suatu kelompok. Dalam kasus ini, kita tahu dengan pasti berapa banyak hasil yang mungkin dan kita dapat menghitung peluang setiap hasil.

Probabilitas diskrit dinyatakan dengan menggunakan fungsi probabilitas, yang memberikan peluang terjadinya setiap hasil dalam himpunan hasil. Fungsi probabilitas ini memiliki sifat-sifat khusus, seperti:

- Semua nilai probabilitas harus berada dalam rentang 0 hingga 1.
- Total probabilitas dari semua hasil dalam ruang sampel harus sama dengan 1.

Probabilitas diskrit sering diilustrasikan dalam bentuk tabel atau diagram batang. Dalam kasus probabilitas diskrit, kita dapat dengan tepat menghitung peluang untuk setiap hasil individual.

Contoh penerapan probabilitas diskrit: Misalnya, jika kita melempar sebuah dadu adil, ruang sampelnya adalah {1, 2, 3, 4, 5, 6}. Setiap hasil memiliki peluang yang sama, yaitu $\frac{1}{6}$, karena dadu adil. Sebagai contoh, peluang mendapatkan angka 3 adalah $\frac{1}{6}$. Probabilitas diskrit adalah jenis probabilitas yang mendasarkan perhitungan pada hasil yang terhitung dan banyak digunakan dalam berbagai masalah matematika, statistik, dan ilmu terapan lainnya.

8. Probabilitas Kontinu (*Continuous Probability*)

Probabilitas kontinu adalah jenis probabilitas yang terkait dengan himpunan hasil yang kontinu dan tak terhitung. Ini mengacu pada situasi di mana ruang sampel terdiri dari interval bilangan riil yang tidak dapat dihitung satu per satu. Dalam probabilitas kontinu, kita mengukur peluang terjadinya rentang nilai tertentu daripada hasil individual.

Contoh probabilitas kontinu termasuk mengukur tinggi badan, berat badan, waktu tempuh, atau suhu. Dalam kasus ini, ada tak terhitung banyak nilai dalam interval tertentu, dan probabilitas dinyatakan dalam bentuk distribusi peluang. Probabilitas kontinu dinyatakan dengan menggunakan fungsi probabilitas kontinu, yang menggambarkan cara probabilitas didistribusikan dalam interval-nilai. Fungsi probabilitas kontinu memiliki sifat-sifat khusus, seperti:

- Nilai probabilitas selalu lebih besar dari atau sama dengan 0.
- Integral dari fungsi probabilitas kontinu dalam seluruh rentang interval-nilai adalah sama dengan 1.

Probabilitas kontinu sering diilustrasikan dengan menggunakan kurva distribusi, seperti kurva normal (distribusi Gauss) atau distribusi eksponensial. Dalam kasus probabilitas kontinu, kita mengukur peluang terjadinya interval-nilai tertentu.

Contoh penerapan probabilitas kontinu: Misalnya, dalam mengukur waktu tempuh perjalanan, ruang sampelnya adalah interval bilangan riil positif. Peluang bahwa waktu tempuh berada dalam suatu interval tertentu dapat dihitung menggunakan fungsi distribusi peluang yang sesuai. Probabilitas kontinu adalah jenis probabilitas yang penting dalam analisis statistik dan ilmu terapan lainnya, terutama ketika berurusan dengan variabel yang dapat memiliki banyak nilai di dalam interval-nilai tertentu.

9. Probabilitas Margin (*Marginal Probability*)

Probabilitas margin, atau *marginal probability*, mengukur peluang terjadinya satu kejadian tanpa memperhatikan kejadian lain yang mungkin terjadi bersamaan. Ini adalah probabilitas tunggal dari kejadian itu sendiri, tidak mempertimbangkan hubungannya dengan kejadian lain. Dalam konteks yang lebih luas, istilah "probabilitas margin" sering digunakan dalam analisis data multivariat, terutama dalam tabel kontingensi atau analisis silang. Dalam hal ini, probabilitas margin mengacu pada peluang terjadinya satu kejadian dalam satu variabel tertentu, tanpa mempertimbangkan variabel lain.

Contoh penerapan probabilitas margin dalam tabel kontingensi: Misalnya, kita memiliki data tentang preferensi makanan berdasarkan jenis kelamin dalam populasi tertentu. Kami memiliki tabel kontingensi yang menunjukkan berapa banyak pria dan wanita yang memilih setiap jenis makanan:

	Pizza	Burger	Salad
Pria	20	25	15
Wanita	15	10	30

Di sini, probabilitas margin untuk memilih pizza oleh semua responden adalah $\frac{3}{6} x \frac{2}{3} = \frac{20+15}{20+25+15+15+10+30}$. Ini adalah probabilitas bahwa seseorang memilih pizza, tanpa memperhatikan jenis kelamin. Probabilitas margin membantu kita memahami distribusi peluang dalam satu variabel tanpa memperhatikan hubungannya dengan variabel lain.

Setiap jenis probabilitas memiliki penggunaan dan aplikasi yang berbeda tergantung pada situasi dan konteksnya. Pemahaman tentang berbagai jenis probabilitas penting untuk menginterpretasikan dan menganalisis peluang dalam berbagai aspek kehidupan dan ilmu pengetahuan.

D. PERMUTASI

Teori permutasi adalah bagian dari teori kombinatorik yang berkaitan dengan pengaturan ulang atau anagram elemen-elemen dalam suatu himpunan. Permutasi berfokus pada berapa banyak cara kita dapat mengatur elemen-elemen tersebut dengan mempertimbangkan urutan mereka. Dalam permutasi, urutan elemen memiliki arti yang berbeda.

Ada beberapa konsep dan istilah penting dalam teori permutasi:

1. Permutasi dengan Unsur Berbeda

Permutasi dengan unsur berbeda adalah pengaturan ulang atau anagram elemen-elemen dalam suatu himpunan di mana semua elemen dianggap unik atau berbeda. Dalam permutasi ini, urutan elemen memiliki arti yang berbeda dan permutasi mewakili semua cara yang mungkin untuk mengatur elemen-elemen tersebut. Jumlah permutasi yang mungkin dari n elemen yang berbeda adalah $n!$, di mana n adalah jumlah elemen dalam himpunan. Ini berarti ada $n!$ cara yang berbeda untuk mengatur elemen-elemen tersebut dalam berbagai urutan.

Contoh permutasi dengan unsur berbeda: Misalkan Anda memiliki tiga buah bola berwarna: merah, hijau, dan biru. Anda ingin mengatur ulang bola-bola ini untuk membuat urutan warna yang berbeda. Jumlah permutasi yang mungkin adalah 3!, atau $3\times2\times1=6$. Berikut adalah enam permutasi yang berbeda:

1) Merah, Hijau, Biru
2) Merah, Biru, Hijau
3) Hijau, Merah, Biru
4) Hijau, Biru, Merah
5) Biru, Merah, Hijau
6) Biru, Hijau, Merah

Dalam contoh di atas, karena ada tiga bola berbeda, kita memiliki 3! atau 6 permutasi yang berbeda. Permutasi dengan unsur berbeda adalah dasar dari konsep permutasi dalam matematika dan memiliki aplikasi dalam berbagai masalah yang melibatkan pengaturan ulang objek-objek atau elemen-elemen yang berbeda.

2. Permutasi dengan Unsur Sama

Permutasi dengan unsur sama terjadi ketika Anda memiliki beberapa elemen yang berulang dalam himpunan dan Anda ingin menghitung berapa banyak pengaturan ulang yang mungkin dengan mempertimbangkan elemen-elemen yang sama. Dalam permutasi ini, meskipun beberapa elemen memiliki nilai yang sama, urutan elemen masih memiliki arti yang berbeda.

Jumlah permutasi yang mungkin dalam permutasi dengan unsur yang sama dihitung dengan mempertimbangkan jumlah kemunculan setiap elemen yang sama dan menggunakan rumus faktorial untuk menghitung jumlah pengaturan ulang. Jika ada n elemen total dan ada n_1 elemen dari jenis pertama, n_2 elemen dari jenis kedua, dan seterusnya, maka jumlah permutasi yang mungkin adalah:

$$\frac{n!}{n1! \times n2! \times \ldots \times nk!}$$

Di mana k adalah jumlah jenis elemen yang berulang.

Contoh permutasi dengan unsur sama: Misalkan Anda memiliki kata "ABB". Anda ingin menghitung berapa banyak anagram yang mungkin dengan kata ini. Ada tiga huruf, di mana "A" muncul dua kali dan "B" muncul satu kali. Jumlah permutasi yang mungkin adalah:

$$\frac{3!}{2! \times 1!} = 3$$

Berikut adalah tiga anagram yang mungkin: "ABB", "BAB", dan "BBA".

Dalam contoh di atas, kita menggunakan rumus permutasi dengan unsur yang sama untuk menghitung jumlah permutasi yang mungkin dengan elemen-elemen yang berulang. Permutasi dengan unsur yang sama sering muncul dalam masalah yang melibatkan kata atau objek yang mengandung elemen yang muncul beberapa kali.

3. Permutasi Sederet

Permutasi sederet adalah bentuk sederhana dari permutasi di mana seluruh elemen dalam himpunan diatur dalam urutan tertentu. Ini berarti semua elemen dalam himpunan harus diatur dalam urutan yang ditentukan. Permutasi sederet melibatkan pengaturan ulang elemen-elemen tanpa ada elemen yang sama atau berulang. Jumlah permutasi sederet dari n elemen adalah $n!$, di mana n adalah jumlah elemen dalam himpunan. Ini berarti ada $n!$ cara yang berbeda untuk mengatur elemen-elemen tersebut dalam urutan yang berbeda.

Contoh permutasi sederet: Misalnya, Anda memiliki kata "ABC". Jumlah permutasi sederet yang mungkin adalah 3!, atau $3\times2\times1=6$. Berikut adalah enam permutasi sederet yang berbeda:

1) ABC
2) ACB
3) BAC
4) BCA
5) CAB
6) CBA

Dalam contoh di atas, karena tidak ada elemen yang sama, kita memiliki 3! atau 6 permutasi sederet yang berbeda. Permutasi sederet merupakan pengaturan ulang yang dasar dan sering digunakan dalam konteks di mana urutan elemen sangat penting, seperti kata-kata atau kombinasi angka dalam kode.

4. Permutasi Siklik

Permutasi siklik adalah bentuk khusus dari permutasi di mana elemen-elemen diatur dalam bentuk siklus atau lingkaran tertentu. Dalam permutasi siklik, urutan elemen dalam suatu siklus memiliki arti yang penting, sementara urutan siklus secara keseluruhan juga berdampak pada hasil permutasi. Dalam sebuah permutasi siklik, setiap elemen dalam himpunan akan menjadi bagian dari satu dan hanya satu siklus. Elemen-elemen dalam siklus akan diatur dalam urutan tertentu, dan siklus-siklus ini dapat diatur dalam urutan yang berbeda untuk menghasilkan berbagai permutasi siklik.

Jumlah permutasi siklik dari n elemen adalah $(n-1)!$, di mana n adalah jumlah elemen dalam himpunan. Hal ini disebabkan oleh fakta bahwa hanya ada satu elemen yang menentukan siklus pertama, dan setelah itu elemen-elemen lain akan mengikuti urutan yang telah ditentukan.

Contoh permutasi siklik: Misalnya, Anda memiliki tiga buah angka: 1, 2, dan 3. Jumlah permutasi siklik yang mungkin adalah (3–1)! = 2!, atau 2×1 = 2. Berikut adalah dua permutasi siklik yang berbeda:

- (1 2 3) - artinya elemen 1 berada di posisi pertama, elemen 2 di posisi kedua, dan elemen 3 di posisi ketiga.
- (1 3 2) - artinya elemen 1 berada di posisi pertama, elemen 3 di posisi kedua, dan elemen 2 di posisi ketiga.

Dalam contoh di atas, kita memiliki (3−1)! atau 2 permutasi siklik yang berbeda. Permutasi siklik sering ditemukan dalam konsep grup dalam matematika dan memiliki aplikasi dalam berbagai masalah yang melibatkan pengaturan ulang elemen dalam bentuk siklus.

5. Permutasi Terbatas

Permutasi terbatas adalah permutasi dari sejumlah k elemen yang diambil dari himpunan n elemen yang berbeda. Dalam permutasi terbatas, kita memperhatikan sejumlah k elemen tertentu dari himpunan n elemen dan menghitung berapa banyak pengaturan ulang yang mungkin dari elemen-elemen tersebut. Jumlah permutasi terbatas dari n elemen yang diambil k elemen sekaligus adalah $P(n,k)$ atau nPk, dan dihitung dengan rumus:

$$P(n,k) = \frac{n!}{(n-k)!}$$

Di sini, n adalah jumlah elemen dalam himpunan, k adalah jumlah elemen yang diambil untuk diatur ulang, dan $(n-k)!$ adalah hasil faktorial dari perbedaan antara n dan k. Permutasi terbatas terutama digunakan ketika Anda ingin menghitung berapa banyak pengaturan ulang yang mungkin dari elemen-elemen tertentu dalam himpunan, tanpa memperhatikan elemen-elemen yang tidak diambil.

Contoh permutasi terbatas: Misalnya, Anda memiliki tiga buah huruf: A, B, dan C. Anda ingin mengatur ulang dua huruf untuk membuat kata dua huruf yang berbeda. Jumlah permutasi terbatas yang mungkin adalah

$$P(3,2) = \frac{3!}{(3-2)!} = 6$$

Berikut adalah enam permutasi terbatas yang berbeda:

1) AB
2) AC
3) BA
4) BC
5) CA
6) CB

Dalam contoh di atas, kita menggunakan rumus permutasi terbatas untuk menghitung jumlah permutasi yang mungkin dari dua elemen yang diambil dari himpunan tiga elemen yang berbeda. Permutasi terbatas adalah alat yang bermanfaat dalam menghitung pengaturan ulang elemen-elemen dalam situasi di mana kita hanya mempertimbangkan sejumlah tertentu dari elemen dalam himpunan yang lebih besar.

6. **Notasi Faktorial**

Permutasi notasi faktorial adalah cara yang ringkas untuk mengekspresikan permutasi terbatas dalam bentuk faktorial. Ini membantu menyederhanakan perhitungan permutasi terbatas dengan menggabungkan faktorial dalam rumus.

Notasi permutasi terbatas dalam bentuk faktorial adalah

$$P(n,k) = \frac{n!}{(n-k)!}$$

Di mana n adalah jumlah total elemen dalam himpunan dan k adalah jumlah elemen yang diambil untuk diatur ulang.

Notasi ini dapat disederhanakan menjadi

$$P(n,k) = \frac{n!}{n-k!}$$

Di mana $n-k!$ adalah faktorial dari perbedaan antara n dan k.

Contoh permutasi notasi faktorial: Misalnya, Anda memiliki enam buah buku yang berbeda, dan Anda ingin mengatur ulang tiga buku untuk menempatkan mereka di atas rak. Permutasi terbatas yang mungkin adalah

$$P(6,3) = \frac{6!}{(6-3)!} = \frac{6!}{3!}$$

yang juga dapat ditulis sebagai ${}_6P_3$.

Dalam hal ini, Anda dapat menghitung nilai faktorial dari 6 dan 3 secara terpisah, lalu menghitung permutasi dengan membagi 6! dengan 3! untuk mendapatkan hasil yang sama. Permutasi notasi faktorial membantu dalam menghitung permutasi terbatas secara lebih efisien dan menyederhanakan perhitungan dengan mengurangi pengulangan perhitungan faktorial.

E. KOMBINASI

Statistika kombinasi adalah cabang statistika yang melibatkan analisis dan penerapan konsep kombinasi dalam konteks pengambilan sampel dan analisis data. Ini melibatkan penghitungan berapa banyak cara yang mungkin untuk memilih sejumlah elemen dari suatu populasi atau sampel, serta menganalisis data yang berkaitan dengan pengelompokan atau kombinasi elemen-elemen tersebut.

Beberapa konsep dalam statistika kombinasi meliputi:

- **Sampel acak**: Dalam statistika, kita sering bekerja dengan sampel acak yang diambil dari populasi. Konsep kombinasi digunakan untuk menghitung berapa banyak cara yang mungkin untuk memilih sejumlah sampel dari populasi.
- **Distribusi sampel**: Distribusi sampel mengacu pada bagaimana elemen-elemen dalam sampel dikelompokkan atau diatur. Konsep kombinasi dapat membantu dalam memahami berapa banyak cara yang mungkin elemen-elemen dalam sampel dapat dikelompokkan.

- **Pengambilan *non*-urutan**: Kombinasi memainkan peran penting dalam statistika karena data dalam statistika seringkali tidak berurutan. Dalam pengambilan sampel atau analisis data, urutan elemen tidak memiliki arti yang khusus, sehingga konsep kombinasi lebih relevan daripada permutasi.
- **Estimasi parameter populasi**: Dalam statistika inferensial, kita sering ingin mengestimasi parameter populasi berdasarkan sampel yang diambil. Konsep kombinasi dapat membantu dalam memahami variasi dan ketidakpastian dalam estimasi ini.
- **Koefisien kombinasi**: Koefisien kombinasi digunakan untuk menghitung jumlah cara yang mungkin untuk memilih sejumlah elemen dari himpunan tanpa memperhatikan urutan. Ini penting dalam menghitung probabilitas, peluang, dan analisis lainnya.

Statistika kombinasi sangat penting dalam analisis data dan penelitian ilmiah. Ini membantu dalam menghitung peluang, membuat estimasi, menghitung interval kepercayaan, dan memahami variasi dalam data. Dalam banyak masalah statistika, konsep kombinasi memainkan peran penting dalam mengatasi tantangan pengelompokan dan pemilihan data yang tepat. Ada beberapa konsep dan istilah penting dalam teori kombinasi:

1. Kombinasi dengan Unsur Berbeda

Kombinasi dengan unsur berbeda terjadi ketika Anda ingin memilih sejumlah k elemen dari himpunan n elemen yang semuanya berbeda. Dalam kombinasi ini, urutan elemen tidak penting, yang berarti hasil yang sama dianggap setara terlepas dari urutan elemen yang dipilih. Jumlah kombinasi dengan unsur berbeda dari n elemen yang diambil k elemen sekaligus adalah $C(n,k)$ atau nCk, dan dihitung dengan rumus:

$$C(n,k) = \frac{n!}{k! \times (n-k)!}$$

Di sini, n adalah jumlah total elemen dalam himpunan, k adalah jumlah elemen yang diambil untuk diatur ulang, $k!$ adalah faktorial dari k, dan $(n-k)!$ adalah faktorial dari perbedaan antara n dan k.

Contoh kombinasi dengan unsur berbeda: Misalnya, Anda memiliki lima buah buah berbeda: apel, pisang, ceri, jeruk, dan anggur. Anda ingin memilih tiga buah untuk dimasukkan ke dalam keranjang tanpa memperhatikan urutan. Jumlah kombinasi dengan unsur berbeda yang mungkin adalah

$$C(5,3) = \frac{5!}{3! \times (5-3)!} = 10$$

Berikut adalah sepuluh kombinasi yang berbeda:

1) Apel, Pisang, Ceri
2) Apel, Pisang, Jeruk
3) Apel, Pisang, Anggur
4) Apel, Ceri, Jeruk
5) Apel, Ceri, Anggur
6) Apel, Jeruk, Anggur
7) Pisang, Ceri, Jeruk
8) Pisang, Ceri, Anggur
9) Pisang, Jeruk, Anggur
10) Ceri, Jeruk, Anggur

Dalam contoh di atas, kita menggunakan rumus kombinasi dengan unsur berbeda untuk menghitung jumlah cara yang mungkin memilih tiga buah dari himpunan lima buah yang berbeda.

2. Kombinasi dengan Unsur Sama

Kombinasi dengan unsur sama terjadi ketika Anda ingin memilih sejumlah k elemen dari himpunan n elemen, di mana beberapa elemen mungkin memiliki nilai yang sama. Dalam kombinasi ini, Anda masih tidak memperhatikan urutan elemen yang dipilih, tetapi Anda juga harus mempertimbangkan kemunculan setiap elemen yang sama dalam pemilihan. Jumlah kombinasi dengan unsur sama dari n elemen yang diambil k elemen sekaligus dapat dihitung dengan mempertimbangkan kemunculan setiap elemen yang sama dalam kombinasi. Ini dihitung dengan rumus:

$$C(n,k) = \frac{n!}{k! \times (n-k)!}$$

Di mana n adalah jumlah total elemen dalam himpunan, k adalah jumlah elemen yang diambil untuk diatur ulang, $k!$ adalah faktorial dari k, dan $(n-k)!$ adalah faktorial dari perbedaan antara n dan k.

Namun, dalam kasus kombinasi dengan unsur sama, Anda juga perlu mempertimbangkan faktor pembagi untuk setiap kemunculan elemen yang sama dalam himpunan yang diambil. Jika ada n_1 elemen dari jenis pertama, n_2 elemen dari jenis kedua, dan seterusnya, maka faktor pembagi adalah $n_1! \times n_2! \times \ldots$.

Sehingga, jumlah kombinasi dengan unsur sama dapat dinyatakan sebagai:

$$C(n,k) = \frac{n!}{k! \times (n-k)! \times (n1! \times n2! \times \ldots)}$$

Contoh kombinasi dengan unsur sama: Misalnya, Anda memiliki kata "MISSISSIPPI" dan ingin memilih empat huruf dari kata tersebut tanpa memperhatikan urutan. Ada 11 huruf total, dengan 4 "I", 4 "S", 2 "P", dan 1 "M". Jumlah kombinasi dengan unsur sama yang mungkin adalah:

$$C(11,4) = \frac{11!}{4! \times (11-4)! \times (4! \times 4! \times 2! \times 1!)}$$

Dalam contoh di atas, kita menggunakan rumus kombinasi dengan unsur sama untuk menghitung jumlah cara yang mungkin memilih empat huruf dari kata "MISSISSIPPI".

3. Kombinasi dengan Pengulangan

Kombinasi dengan pengulangan terjadi ketika Anda ingin memilih sejumlah k elemen dari himpunan n elemen, di mana elemen-elemen dapat dipilih lebih dari sekali. Dalam kombinasi ini, urutan elemen yang dipilih tidak penting, dan Anda tidak memperhatikan kemunculan berulang dari elemen yang sama. Jumlah kombinasi dengan pengulangan dari n elemen yang diambil k elemen sekaligus dapat dihitung dengan rumus:

$$C(n+k-1,k) = \frac{(n+k-1)!}{k! \times (n-1)!}$$

Di sini, n adalah jumlah total elemen dalam himpunan, k adalah jumlah elemen yang diambil, $k!$ adalah faktorial dari k, dan $(n-1)!$ adalah faktorial dari $n-1$.

Contoh kombinasi dengan pengulangan: Misalnya, Anda memiliki kue dengan tiga jenis hiasan: ceri, cokelat chip, dan kacang. Anda ingin memilih lima hiasan untuk diletakkan di atas kue. Anda dapat memilih jenis yang sama lebih dari sekali. Jumlah kombinasi dengan pengulangan yang mungkin adalah:

$$C(3+5-1{,}5) = \frac{(3+5-1)!}{5! \times (3-1)!} = \frac{7!}{5! \times 2!} = 21$$

Dalam contoh di atas, kita menggunakan rumus kombinasi dengan pengulangan untuk menghitung berapa banyak cara yang mungkin memilih lima hiasan dari tiga jenis yang tersedia untuk diletakkan di atas kue.

4. Kombinasi Tanpa Pengulangan

Kombinasi tanpa pengulangan terjadi ketika Anda ingin memilih sejumlah k elemen dari himpunan n elemen, di mana setiap elemen hanya dapat dipilih sekali dan urutan elemen tidak penting. Dalam kombinasi ini, Anda tidak memperhatikan urutan pemilihan, dan setiap elemen hanya dapat dipilih satu kali. Jumlah kombinasi tanpa pengulangan dari n elemen yang diambil k elemen sekaligus dapat dihitung dengan rumus yang sama seperti kombinasi dengan unsur berbeda:

$$C(n,k) = \frac{n!}{k! \times (n-k)!}$$

Di sini, n adalah jumlah total elemen dalam himpunan, k adalah jumlah elemen yang diambil untuk diatur ulang, $k!$ adalah faktorial dari k, dan $(n-k)!$ adalah faktorial dari perbedaan antara n dan k.

Contoh kombinasi tanpa pengulangan: Misalnya, Anda memiliki enam kartu berbeda dan ingin memilih tiga kartu untuk ditampilkan. Jumlah kombinasi tanpa pengulangan yang mungkin adalah:

$$C(6{,}3) = \frac{6!}{3! \times (6-3)!} = \frac{6!}{3! \times 3!} = 20$$

Dalam contoh di atas, kita menggunakan rumus kombinasi tanpa pengulangan untuk menghitung berapa banyak cara yang mungkin memilih tiga kartu dari enam kartu yang berbeda.

5. Notasi Faktorial

Kombinasi dalam notasi faktorial adalah cara yang lebih sederhana untuk menghitung kombinasi tanpa harus menghitung faktorial untuk setiap angka. Notasi kombinasi dalam bentuk faktorial dinyatakan sebagai $C(n,k)$ atau nCk, dan dihitung dengan rumus:

$$C(n,k) = \frac{n!}{k! \times (n-k)!}$$

Di mana n adalah jumlah total elemen dalam himpunan, k adalah jumlah elemen yang diambil untuk diatur ulang, $k!$ adalah faktorial dari k, dan $(n-k)!$ adalah faktorial dari perbedaan antara n dan k. Namun, dalam notasi faktorial, Anda tidak perlu menghitung faktorial untuk setiap angka secara terpisah. Sebaliknya, Anda dapat memanfaatkan sifat faktorial untuk mempermudah perhitungan.

Contoh kombinasi notasi faktorial: Misalnya, untuk menghitung $C(6,3)$, Anda dapat menggunakan notasi faktorial:

$$C(6,3) = \frac{6!}{3! \times (6-3)!} = \frac{6!}{3! \times 3!}$$

Namun, daripada menghitung 6! dan 3! secara terpisah, Anda dapat menggunakan sifat faktorial $n!=n\times(n-1)!$ untuk menyederhanakan perhitungan:

$$C(6,3) = \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{3 \times 2 \times 1 \times (3 \times 2 \times 1)} = \frac{6 \times 5 \times 4}{3 \times 2 \times 1} = 20$$

Dalam contoh di atas, kita menggunakan notasi faktorial dan sifat faktorial untuk menghitung $C(6,3)$ dengan lebih efisien.

BAB 8

DISTRIBUSI PROBABILITAS

A. PENGANTAR

Distribusi probabilitas merupakan himpunan yang nilai-nilainya didistribusikan berdasarkan teori probabilitas. Nilai-nilai tersebut merupakan kemungkinan peristiwa dari sebuah percobaan yang berkaitan dengan ketidakpastian dari peristiwa tersebut, selanjutnya nilai ini disebut sebagai *variable random*.

Tujuan dari distribusi probabilitas adalah untuk mendeskripsikan semua kemungkinan nilai dan kemungkinan-kemungkinan lain yang dapat diambil dari berbagai *variable* random pada rentang tertentu. Faktor-faktor yang mempengaruhi distribusi probabilitas diantaranya adalah *mean*, deviasi standar, kurtosis dan *skwenes*. Karakteristik khas dari distribusi probabilitas adalah memiliki kurva berbentuk lonceng.

Pada sub bab ini akan dibahas macam-macam distribusi probabilitas, yaitu distribusi probabilitas diskrit yang terdiri dari distribusi binomial dan poisson, kemudian distribusi probabilitas kontinu yaitu distribusi normal dan yang terakhir distribusi probabilitas sampling.

B. DISTRIBUSI PROBABILITAS DISKRIT

Suatu ruang sampel mengandung sederetan anggota yang banyaknya sebanyak bilangan bulat atau berhingga banyaknya disebut ruang sampel diskrit. Dengan demikian distribusi probabilitas diskrit merupakan beberapa peristiwa atau kejadian yang terdistribusi dengan setiap kemungkinan kejadiannya terhitung.

1. Distribusi Binomial/Distribusi Bernoulli

Distribusi binomial merupakan distribusi probabilitas yang memuat suatu kejadian dengan hasil yang memuat dua kemungkinan. Suatu percobaan dikatakan binomial jika memenuhi syarat-syarat berikut:

a. Percobaan dilakukan lebih dari satu kali (berulang-ulang),
b. Termasuk dalam kejadian saling bebas (hasil suatu percobaan tidak bergantung pada hasil percobaan lainnya),
c. Setiap kejadian memiliki probabilitas tetap untuk setiap percobaan,
d. Setiap kejadian hanya memiliki dua kemungkinan (ya-tidak, sukses-gagal, baik-cacat, gambar-angka, dll).

Secara umum rumus probabilitas binomial dari suatu kejadian adalah:

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

Dimana:

x : banyaknya peristiwa sukses
n : banyaknya percobaan
p : probabilitas peristiwa sukses
q : banyaknya peristiwa gagal / $1 - p$

Contoh

Sebuah dadu dilempar sebanyak 5 kali. Berapakah peluang munculnya mata dadu berjumlah 7 sebanyak 3 kali pada pelemparan tersebut?

Pembahasan:

$n = 5 \qquad x = 3 \qquad P(X = 3) = \cdots ?$

Kemungkinan munculnya mata dadu berjumlah 7 adalah $\{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$. Dengan demikian, probabilitas munculnya mata dadu berjumlah 7 adalah:

$$p = P(7) = \frac{6}{36} = \frac{1}{6}$$

Probabilitas munculnya mata dadu berjumlah selain 7 adalah:

$$q = 1 - p = 1 - \frac{1}{6} = \frac{5}{6}$$

Probabilitas munculnya mata dadu berjumlah 7 sebanyak 3 kali adalah:

$$P(X = 3) = C_3^5 \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^2 = 0{,}0322$$

2. Distribusi Poisson

Suatu percobaan binomial yang memiliki n lebih dari 50 dan p kurang dari 0,1 akan sulit dihitung dengan menggunakan distribusi binomial. Masalah tersebut dapat diselesaikan dengan distribusi Poisson sebagai pendekatan dari distribusi binomial.

Distribusi Poisson merupakan distribusi probabilitas diskrit yang menyatakan peluang jumlah peristiwa yang terjadi pada interval waktu tertentu. Distribusi Poisson dirumuskan sebagai berikut:

$$\boxed{\boldsymbol{P(X = x) = \frac{\mu^x \cdot e^{-\mu}}{x!}}}$$

Dimana:

μ : rata-rata kejadian suatu peristiwa = $n \cdot p$

e : bilangan alam = 2,7182

Contoh

Dalam sebuah novel yang terdiri dari 200 halaman terdapat 80 kata yang salah cetak dan berdistribusi secara acak dalam halaman-halaman novel tersebut. Tentukan probabilitas bila sebuah halaman novel itu dibuka, maka terdapat 4 kata salah cetak.

Penyelesaian:

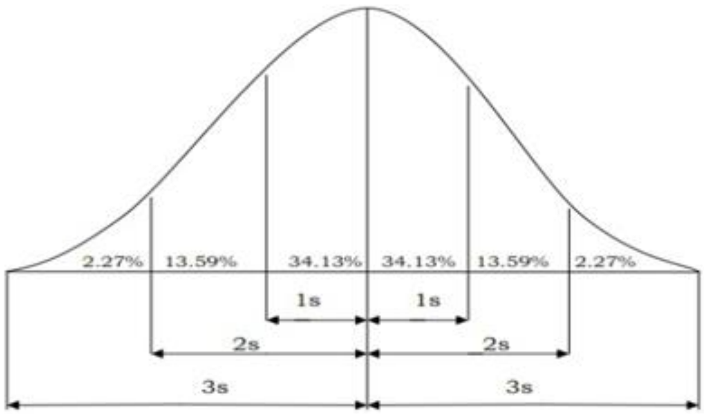
$n = 80 \qquad p = \frac{1}{200} \qquad \mu = n \cdot p = 80 \cdot \frac{1}{200} = 0{,}4$

$$P(X = 4) = \frac{(0{,}4)^4 \cdot (2{,}71828)^{-0{,}4}}{4!} = \frac{(0{,}0256)(0{,}67032)}{24} = 0{,}00071$$

C. DISTRIBUSI NORMAL

Salah satu distribusi probabilitas kontinu yang akan dibahas adalah distribusi Normal. Disebut kontinu karena variabel yang didapat merupakan hasil dari perhitungan bukan pencacahan, sehingga variabelnya merupakan bilangan real. Distribusi Normal merupakan distribusi dengan *variable* random kontinu, seperti tinggi badan, suhu tubuh dan sebagainya.

Distribusi Normal atau distribusi Gauss memiliki grafik berbentuk lonceng yang disebut kurva normal dengan puncaknya berada pada *mean*, serta merupakan distribusi yang simetris.



Gambar 8.1. Kurva distribusi normal
(https://www.konsultanstatistik.com/2021/07/apa-itu-tabel-z.html)

Kurva normal sangat dipengaruhi oleh rata-rata dan simpangan baku. Jika rata-rata dan simpangan baku besar maka kurva akan semakin rendah, sebaliknya kurva akan semakin runcing. Bentuk fungsi distribusi normal adalah:

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

Dimana:

x : nilai data

π : 3,14

σ : simpangan baku

μ : rata-rata dari x

e : 2,71828

1. Distribusi Normal Standar

Distribusi normal standar atau yang dikenal dengan distribusi Z digunakan untuk menghitung luas daerah di bawah kurva normal dengan menggunakan *table* distribusi Z (*table* lihat di akhir bab 8). Distribusi normal standar merupakan distribusi normal yang memiliki rata-rata 0 dan simpangan baku 1.

Luas keseluruhan dari kurva adalah 1 (nilai total peluang) dan kurva simetris terhadap $\mu = 0$, sehingga luas daerah kiri dari 0 dan luas daerah kanan setelah 0 mempunyai luas sebesar 0,5 atau $P(Z > 0) = P(Z < 0) = 0{,}5$. Dengan demikian nilai negatif di daerah kiri mempunyai luas yang sama dengan nilai positif di daerah kanan. Luas daerah di bawah kurva normal pada interval tertentu ditulis dengan $P(0 < Z < b)$.

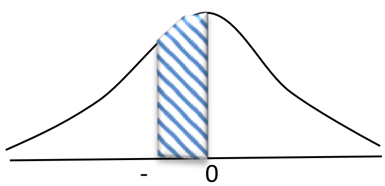
Contoh 1 penggunaan tabel distribusi Z:

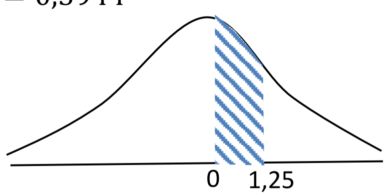
Hitunglah nilai peluang dengan menggunakan *table* Z

a. $P\,(-1{,}25 < Z < 0)$
b. $P(1{,}00 < Z < 2{,}01)$
c. $P(-0{,}05 < Z < 1.00)$
d. $P(-2{,}01 < Z < -1{,}25)$
e. $P(Z > 2{,}01)$

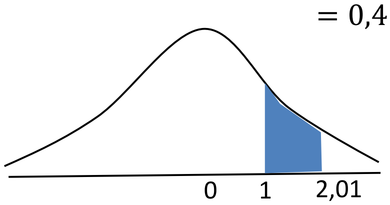
Penyelesaian:

a. Dengan menggunakan *table* didapat
$P(-1{,}25 < Z < 0) = P(0 < Z < 1{,}25) = 0{,}3944$





b. Dengan menggunakan *table* didapat
$$P(1{,}00 < Z < 2{,}01) = P(0 < Z < 2{,}01) - P(0 < Z < 1{,}00)$$
$$= 0{,}4778 - 0{,}3413 = 0{,}1365$$

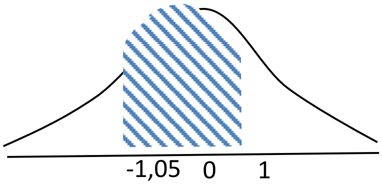


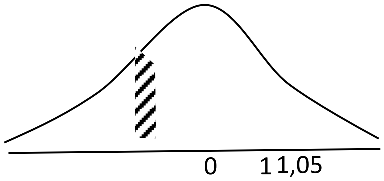
c. Dengan menggunakan *table* didapat

$P(-1{,}05 < Z < 1{,}00) = P(1{,}00 < Z < 1{,}05)$

$= P(0 < Z < 1{,}05) - P(0 < Z < 1{,}00)$

$= 0{,}3531 - 0{,}3413 = 0{,}0118$



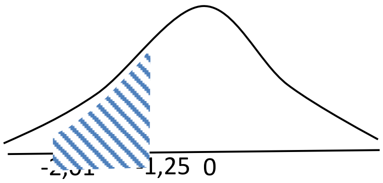


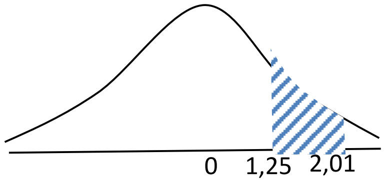
d. Dengan menggunakan *table* didapat

$P(-2{,}01 < Z < -1{,}25) = P(1{,}25 < Z < 2{,}01)$

$= P(0 < Z < 2{,}01) - P(0 < Z < 1{,}25)$

$= 0{,}4778 - 0{,}3944 = 0{,}0834$





e. Dengan menggunakan *table* didapat

$$P(Z > 2{,}01) = 0{,}5 - P(0 < Z < 2{,}01) = 0{,}5 - 0{,}4778 = 0{,}0222$$

Cat: 0,5 adalah total luas daerah sisi sebelah kanan 0

Perlu diingat bahwa *table* Z tidak bisa digunakan untuk distribusi normal umum. Sehingga untuk mencari luas di bawah kurva normal dengan $\mu \neq 0$ dan $\sigma \neq 1$ dapat dicari dengan mengubah distribusi normal umum menjadi distribusi normal standar dengan menggunakan rumus:

$$Z = \frac{X - \mu_X}{\sigma_X}$$

Dimana:

Z: variable normal standar

X: nilai variable random (distribusi normal umum)

μ_X: rata-rata variable random

σ_X: simpangan baku *variable* random

Contoh 2

Rata-rata nilai perusahaan listing di BEJ adalah Rp 5.750,- dengan simpangan baku Rp 2.250,- yang berdistribusi normal. Jika diketahui perusahaan-perusahaan yang memiliki nilai saham antara Rp 3.150,- dan Rp 3.500,- berjumlah 119 perusahaan, berapakah jumlah perusahaan yang listing di BEJ?

Penyelesaian;

$N = 119 \quad \mu_X = 5.750 \qquad\qquad \sigma_X = 2.250$

Tentukan $P(3.150 < X < 3.500) =?$

Untuk $X = 3.150 \to Z = \frac{3.150-5.750}{2.250} = -1{,}16$

Untuk $X = 3.500 \to Z = \frac{3.500-5.750}{2.250} = -1$

Sehingga $P(3.150 < X < 3.500) = P(-1{,}16 < Z < -1)$. Dengan menggunakan table Z nilai peluangnya adalah:

$P(-1{,}16 < Z < -1) = P(1 < Z < 1{,}16)$

$= P(0 < Z < 1{,}16) - P(0 < Z < 1)$

$= 0{,}3770 - 0{,}3413 = 0{,}0357$

D. DISTRIBUSI SAMPLING

1. Distribusi Sampling Rata-Rata

Distribusi sampling rata-rata merupakan distribusi yang terbentuk dari rata-rata semua sampel yang ada. Distribusi ini bergantung kepada populasi, ukuran sampel dan metode pemilihan sampel.

- Bila populasi terbatas, untuk pengambilan sampel tanpa pengembalian atau $\frac{n}{N} > 5\%$:

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

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

- Bila populasi tidak terbatas, untuk pengambilan sampel dengan pengembalian atau $\frac{n}{N} \leq 5\%$:

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

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

Distribusi normal standar untuk distribusi sampling rata-rata dapat ditentukan dengan:

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}}$$

Contoh:

Suatu bank akan menghitung tabungan seluruh nasabahnya . Diketahui bahwa bank tersebut mempunyai 500 nasabah dengan rata-rata tabungan setiap nasabah adalah Rp 300,- dan deviasi standar Rp 220,-. Apabila seorang peneliti mengambil sampel 200 nasabah, berapa probabilitas jika rata-rata sampel terletak antara Rp 270,- dan Rp 305,-

Penyelesaian:

$\mu_{\bar{X}} = \mu = 300 \qquad N = 500 \qquad n = 200 \qquad \sigma = 220$

$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{n-1}} = \frac{250}{\sqrt{200}}\sqrt{\frac{500-200}{200-1}} = 21{,}7050$

$P(270 < X < 305) =?$

Untuk mencari nilai Z, digunakan rumus distribusi normal standar sehingga:

Untuk $X = 270 \rightarrow Z = \frac{\bar{X}-\mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{270-300}{21{,}7050} = -1.38$

Untuk $X = 305 \rightarrow Z = \frac{\bar{X}-\mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{305-300}{21{,}7050} = 0{,}23$

Maka $P(270 < X < 305) = P(-1{,}38 < Z < 0{,}23) = P(0{,}23 < Z < 1{,}38)$

$= P(0 < Z < 1{,}38) - P(0 < Z < 0{,}23) = 0{,}4162 - 0{,}0910 = 0{,}3252$

2. Distribusi Sampling Beda Dua Rata-Rata

Distribusi sampling beda dua rata-rata merupakan distribusi yang terbentuk dari selisih rata-rata seluruh sampel pada dua populasi.

- Rata-rata

  $\mu_{\bar{X}_1-\bar{X}_2} = \mu_{\bar{X}_1} - \mu_{\bar{X}_2}$

- Simpangan baku

  $\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

Untuk ukuran sampel yang cukup besar ($n_1, n_2 > 30$) maka distribusi akan mendekati normal. Sehingga untuk penyelesaiannya dapat menggunakan bentuk distribusi normal standar.

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_{\bar{X}_1 - \bar{X}_2})}{\sigma_{\bar{X}_1 - \bar{X}_2}}$$

Contoh

100 orang mahasiswa kelas A memiliki rata-rata nilai ujian 70 dengan deviasi standar 10. Sedangkan rata-rata nilai ujian kelas B dengan jumlah 90 mahasiswa adalah 65 dengan deviasi standar 25. Apabila diambil 25 mahasiswa kelas A dan 20 mahasiswa kelas B sebagai sampel, berapa probabilitas perbedaan rata-rata kedua kelompok sampel tersebut 10 atau lebih?

Penyelesaian

$\mu_1 = 70 \quad \sigma_1 = 10 \qquad n_1 = 25$

$\mu_2 = 65 \quad \sigma_2 = 25 \qquad n_2 = 20$

$P(\bar{X}_1 - \bar{X}_2 \geq 10) = ?$

$\mu_{\bar{X}_1 - \bar{X}_2} = 70 - 65 = 5$

$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{10^2}{25} + \frac{25^2}{20}} = \sqrt{35,25} = 5,9372$

$Z = \frac{10-5}{5,9372} = 0,8421$

Jadi $P(\bar{X}_1 - \bar{X}_2 \geq 10) = P(Z \geq 0,84) = 0,5 - P(0 < Z < 0,84)$

$= 0,5 - 0,2995 = 0,2005 = 20,05\%$

3. Distribusi Sampling Proporsi

Distribusi sampling proporsi merupakan distribusi yang terbentuk dari proporsi (persentase) dari semua sampel yang sama besar. Proporsi dari suatu populasi dinyatakan dengan $P = \frac{X}{N}$ sedangkan untuk sampel dinyatakan dengan $p = \frac{X}{n}$.

- Bila populasi terbatas, untuk pengambilan sampel tanpa pengembalian atau $\frac{n}{N} >$ 5%:

  $\mu_P = P$

  $\sigma_P = \sqrt{\frac{PQ}{n}} \sqrt{\frac{N-n}{N-1}}$

- Bila populasi tidak terbatas, untuk pengambilan sampel dengan pengembalian atau $\frac{n}{N} \leq 5\%$:
  $\mu_P = P$
  $\sigma_P = \sqrt{\frac{PQ}{n}}$

Distribusi normal standar untuk distribusi sampling proporsi dapat dinyatakan dengan:

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

Dimana
P: proporsi populasi kejadian sukses
Q: proporsi populasi kejadian gagal $(1 - P)$
p: proporsi sampel kejadian sukses

Contoh
Sebuah toko roti Bernama "Halal Bakery" menemukan bahwa 20% pembelian dilakukan secara *online* dengan menggunakan aplikasi. Jika diambil sampel acak sebanyak 180 orang yang membeli roti di toko tersebut, maka berapa probabilitas pelanggan yang membeli dengan aplikasi kurang dari 15%?

Penyelesaian
$n = 180 \quad P = 20\% = 0{,}20 \qquad Q = 1 - 0{,}2 = 0{,}8$
$p = 15\% = 0{,}15$
$P(p < 15\%) = ?$

$\mu_P = 0{,}20$
$\sigma_P = \sqrt{\frac{0{,}20(0{,}8)}{180}} = 0{,}0298$
$Z = \frac{0{,}15 - 0{,}20}{0{,}0298} = -1{,}68$

Jadi $P(p < 15\%) = P(Z < -1{,}68) = P(Z > 1{,}68)$
$= 0{,}5 - P(0 < Z < 1{,}68) = 0{,}5 - 0{,}4535 = 0{,}04465$

4. Distribusi Sampling Beda Dua Proporsi

Distribusi sampling beda dua proporsi merupakan distribusi yang terbentuk dari selisih proporsi dari seluruh sampel pada dua populasi.

- Rata-rata

  $\mu_{P_1-P_2} = P_1 - P_2$
- Simpangan baku

  $\sigma_{P_1-P_2} = \sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}$

Untuk ukuran sampel yang cukup besar ($n_1, n_2 > 30$) maka distribusi akan mendekati normal. Sehingga untuk penyelesaiannya dapat menggunakan bentuk distribusi normal standar.

$$Z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sigma_{P_1-P_2}}$$

Contoh

Dari hasil survei, diketahui bahwa 300 dari 2000 pasien di RS "Selalu Sehat" adalah penderita penyakit jantung. Sementara itu, 10 dari 100 pasien di RS "Selalu Aman" adalah penderita penyakit jantung. Bila diambil sampel masing-masing 20 pasien dari kedua RS tersebut, berapa probabilitas selisih proporsi pasien penyakit jantung di kedua RS adalah 6% atau lebih?

Penyelesaian

$P_1 = \frac{300}{1000} = 0{,}15 \quad P_2 = \frac{10}{100} = 0{,}1 \qquad P_1 - P_2 = 0{,}15 - 0{,}1 = 0{,}05$

$n_1, n_2 = 20$

$P(p_1 - p_2 \geq 6\%) = ?$

$\sigma_{P_1-P_2} = \sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}} = \sqrt{\frac{0{,}15(0{,}85)}{20} + \frac{0{,}1(0{,}9)}{20}} = 0{,}11$

Maka $P(p_1 - p_2 \geq 6\%) = P(Z \geq 0{,}11) = 0{,}5 - P(0 < Z < 0{,}11)$

$= 0{,}5 - 0{,}0359 = 4641 = 46{,}41\%$

BAB

9

STATISTIKA INFERENSI

A. PENGANTAR

Dalam bab 9 buku ini akan dijelaskan mengenai pengenalan statistik inferensi, Uji hipotesis, Uji-t, Uji-X^2 dan Uji-F. Inferensi atau inferensial dibuat oleh peneliti pendidikan dan ilmu sosial umumnya untuk melihat perbedaan dan korelasi antara dua variabel *independent* maupun antara beberapa variabel sekaligus. Beberapa nama lain statistik Inferensial yaitu statistik induktif atau statistik lanjut atau statistik mendalam. Statistik inferensial yaitu statistik yang menyediakan aturan atau cara, dimana cara tersebut dapat dipergunakan sebagai alat dalam rangka mencoba menarik kesimpulan yang bersifat umum dari sekumpulan data yang telah dikumpul dan diolah. Kesimpulan lazimnya dilakukan dalam rangka menguji hipotesis penelitian yang telah dirumuskan dan melakukan generalisasi hasil penelitian. Dengan kata lain untuk mempelajari statistik inferensial seseorang diperlukan mempelajari statistik deskriptif terlebih dahulu.

Estimasi parameter merupakan teknik statistika dalam menduga nilai parameter pada populasi yang dapat berupa estimasi titik yaitu memperkirakan parameter berdasarkan satu nilai saja misalkan μ (rata-rata populasi) dengan $\bar{x}$ (rata-rata sampel) dimana $\bar{x} \rightarrow \mu = \bar{x}$ tentu saja hasil estimasi ini tidak memberikan tingkat keyakinan tertentu, ataupun estimasi interval yaitu memperkirakan suatu parameter berdasarkan nilai dalam interval tertentu sehingga hasilnya akan memberikan tingkat keyakinan tertentu. Misalnya untuk mengestimasi μ digunakan interval estimasi: $\bar{x} - d < \mu < \bar{x} + d$ atau $\mu = \bar{x} \pm d$ dimana d adalah perbedaan *true value* dan *estimate value* (*difference*) yang dikehendaki atau disebut juga *estimation* *error* atau kekeliruan estimasi atau galat estimasi. Penelitian merupakan cara mencari kebenaran melalui metode ilmiah mulai dari merumuskan masalah, melakukan studi literatur yaitu studi mengenai teori dan atau hasil penelitian di masa lampau berkenaan permasalahan yang akan dikaji, bila perlu merumuskan hipotesis, mengumpulkan data, mengolah data, sampai mengambil kesimpulan. Bidang permasalahan yang telah diteliti secara mendalam, biasanya sumber-sumber bacaannya telah banyak tersedia sehingga peneliti bisa terhindar dari penelaahan yang kurang penting. Akan tetapi, bagi permasalahan yang belum atau jarang diteliti perlu melakukan studi literatur bidang lain yang berhubungan dengan permasalahan yang kita bahas agar rasional dan rumusan hipotesisnya logis. Dalam penelitian kesimpulan dibuat berdasarkan penaksiran yang dilakukan ataupun melalui pengujian hipotesis. Penelitian eksperimen menggunakan statistika inferensial, yaitu statistika yang dipergunakan untuk membuat generalisasi hasil penelitian terhadap populasinya atau terhadap yang lain yang karakteristiknya mirip dengan populasi itu. Beberapa karakteristik percobaan, antara lain:

1. Adanya kesetaraan subjek dalam kelompok-kelompok yang berbeda.
2. Menggunakan statistika inferensial.
3. Adanya kontrol terhadap variabel-variabel luar.
4. Paling tidak, ada satu variabel bebas yang dimanipulasikan.

Dalam penelitian, peneliti tidak membuat ancang-ancang untuk membuktikan hipotesis tetapi lebih bertindak mengumpulkan data untuk melihat apakah hipotesis oleh data itu disokong atau tidak. Jadi, tujuan sebenarnya dari melaksanakan penelitian adalah untuk memperoleh data. Data yang kita perlukan untuk pengujian hipotesis diperoleh dengan instrumen. Data yang diperoleh, lalu diolah ke dalam ukuran statistik yang akan kita pergunakan. Jadi, bila ukuran sampel terhadap populasinya besar dan atau populasinya kecil, pengambilan sampel tanpa dan dengan pengembalian, ditinjau dari statistik inferensial, perbedaan maknanya besar. Setelah menyatakan hipotesis, peneliti merancang penelitian. Peneliti memilih uji statistik yang tepat, memilih tingkat signifikansi yang sesuai, dan merumuskan rencana untuk melakukan penelitian, seperti yang akan dibahas pada bab ini mengenai Uji Hipotesis, Uji-t, Uji-Chi Kuadrat dan Uji-F.

B. UJI HIPOTESIS

Dalam suatu penelitian, ukuran-ukuran dalam populasi tidak selalu diketahui, ketika hal ini yang terjadi, maka yang harus dilakukan adalah dengan pengujian hipotesis. Literatur atau pengalaman terdahulu membuat suatu hipotesis terhadap populasi, kemudian data yang didapat dari sampel dilakukan pengujian terhadap hipotesis yang dibuat. Kesimpulan dibuat dengan mempelajari ukuran pemusatan dan penyebaran data. Agar kemudian kesimpulan dapat dipercaya, maka kaidah kaidah pengujian hipotesis, seperti asumsi sebaran dan distribusi rataan sampel harus dipenuhi. Kaidah asumsi-asumsi dasar tentang distribusi data serta jenis data yang terkumpul dari populasi dan sampel yaitu:

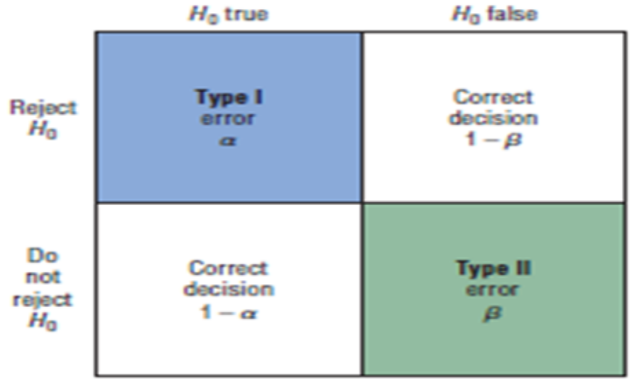
1. Sampel diambil secara acak (saat peneliti menggunakan penelitian sampel).
2. Data-data yang terkumpul dari sampel-sampel atau populasi, bersifat homogen (sama atau mendekati sama), terutama jika jumlah sampel atau populasinya kecil.
3. Jenis data berskala interval atau rasio.

Hipotesis adalah penjelasan atau jawaban tentatif (sementara) tentang tingkah laku, fenomena (gejala), atau kejadian yang akan terjadi ataupun kejadian yang sedang berjalan. Berdasarkan bagaimana diperolehnya hipotesis dibagi menjadi hipotesis induktif dan hipotesis deduktif. Hipotesis induktif hanya didasarkan pada observasi, generalisasi hasil observasi. Hipotesis deduktif adalah hipotesis yang dijabarkan dari teori dengan menyiapkan bukti sebagai pendukung untuk memperluas atau menentang teori yang sudah ada. Selain berdasarkan bagaimana diperolehnya hipotesis, terdapat juga pembagian secara lain yaitu: hipotesis statistik, hipotesis penelitian, dan hipotesis riset atau deklaratif. Hipotesis statistik dibedakan menjadi dua yaitu hipotesis nol (*Null Hypothesis*) yang dibuat dengan tujuan untuk ditolak dan dilambangkan H_0, perumusannya mengandung pengertian sama atau memiliki perbedaan antara dua parameter dengan nilai tertentu serta hipotesis alternatif (*Alternative Hypothesis*) dilambangkan dengan H_1 atau Ha mengandung pengertian tidak sama, lebih kecil atau lebih besar. Dalam bab ini, hipotesis nol selalu dinyatakan dengan menggunakan tanda sama dengan (=), sedangkan hipotesis alternatif menggunakan tanda $>$ atau $<$ atau $\neq$. Frase hipotesis nol dengan tanda = artinya; Sama dengan; Apakah sama dengan; Tidak berubah dari. Frase hipotesis alternatif dengan tanda > artinya; Lebih besar dari; Apakah di atas; lebih tinggi dari; Lebih panjang dari; Apakah bertambah. Frase hipotesis alternatif dengan tanda < artinya; Lebih kecil dari; Apakah di bawah; Lebih rendah dari; Lebih pendek dari; Apakah berkurang. Frase hipotesis alternatif dengan tanda ≠ artinya; Tidak sama dengan; Apakah berbeda dari; Telah berubah dari.

Berdasarkan pengertian hipotesis, dimana H_0 dibuat untuk ditolak maka jelas kecenderungan suatu penelitian diarahkan untuk menolak H_0. Meskipun definisi hipotesis nol atau hipotesis alternatif yang menggunakan kata parameter, definisi ini dapat diperluas untuk menyertakan istilah lain seperti distribusi dan keacakan. Hipotesis nol dan hipotesis alternatif dinyatakan bersama-sama. Hipotesis kerja atau hipotesis alternatif (Ha) senantiasa diformulasikan berupa kalimat positif (mengiyakan) oleh sebab itu, terkadang peneliti menyebut hipotesis alternatif dengan hipotesis penelitian, sebaliknya hipotesis nol (H_0) diformulasikan berupa kalimat negatif (menyangkal). Semestinya, hipotesis lebih mendekati kebenaran dari pada kekeliruan, tetapi tidak mustahil perkiraannya meleset sama sekali. Berikut langkah-langkah pengujian Hipotesis, yaitu :

1. Hipotesis dinyatakan dengan H, supaya nampak adanya dua pilihan perlu didampingi oleh pernyataan lain yang isinya berlawanan.
2. Tentukan kriteria pengujian, meliputi daerah penerimaan dan daerah penolakan hipotesis (daerah kritis).
3. Jika perumusannya mengandung pengertian sama atau tidak memiliki perbedaan, disebut Hipotesis nol (H_0) melawan hipotesis yang mengandung pengertian tidak sama, lebih besar atau lebih kecil (H_1).
4. Pilih bentuk statistik mana yang harus digunakan, apakah z, t, X^2, F dll.

Uji statistik menggunakan data yang diperoleh dari sampel untuk membuat keputusan tentang apakah hipotesis nol harus ditolak. Tingkat signifikansi adalah probabilitas maksimum melakukan kesalahan tipe I. Probabilitas ini dilambangkan dengan α (alpha). Sedangkan, lambang probabilitas kesalahan tipe II yaitu β (beta). Tingkat keyakinan/tingkat kepercayaan dalam penelitian dilambangkan (1- α). Misal: Jika α = 5% maka tingkat kepercayaan = 95% , α (alpha) dalam penelitian adalah 1% sampai dengan 10%, namun biasanya dipakai alpha 5%. Misalnya α = 5 % , maksudnya bahwa sekitar 5 dari tiap 100 kesimpulan akan menolak hipotesis yang seharusnya diterima dan 95 yakin bahwa kita telah membuat kesimpulan yang benar. Semakin kecil alpha semakin bagus, ada kalanya peneliti memiliki pilihan dua atau lebih uji statistik untuk menguji hipotesis. Namun, penting untuk diingat bahwa uji statistik memiliki asumsi yang perlu dipertimbangkan. Jika asumsi tidak terpenuhi, maka uji lain dengan kekuatan yang lebih rendah harus digunakan. Kekuatan uji dapat ditingkatkan dengan meningkatkan nilai α. Ingatlah bahwa ketika α bertambah β berkurang. Jadi jika β dikurangi, maka 1 - β akan bertambah, sehingga meningkatkan kekuatan uji. Cara lain untuk meningkatkan kekuatan uji adalah memilih ukuran sampel yang lebih besar. Ukuran sampel yang lebih besar akan membuat kesalahan standar rata-rata lebih kecil dan akibatnya mengurangi β. Empat kemungkinan hasil uji hipotesis terlihat pada gambar berikut.



Gambar 9.1 Kemungkinan Hasil Uji Hipotesis (Bluman, Allan G., 2009)

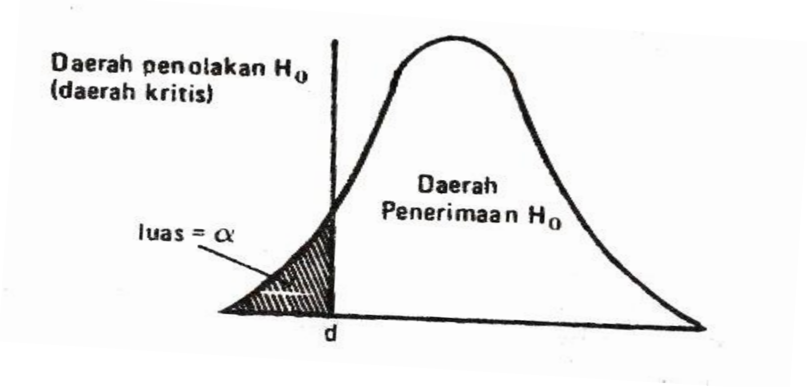
Tipe I : Kekeliruan α (*error* α)/taraf signifikan/taraf arti/taraf nyata
Tipe II : Kekeliruan β (*error* β)

Kekeliruan Tipe I adalah kekeliruan yang dibuat ketika H_0 ditolak atau dianggap salah (*reject* H_0) padahal kenyataannya H_0 seharusnya diterima (H_0 *true*). Hal ini menunjukkan kekeliruan maksimum yang boleh dibuat ketika kita mengambil kesimpulan menolak H_0 dan biasanya besarnya dilambangkan dengan α. Sedangkan kekeliruan Tipe II adalah kekeliruan yang dibuat ketika menerima H_0 (*Do not reject* H_0) yang seharusnya ditolak (H_0 *false*). Kekeliruan tipe II dilambangkan dengan β (beta) dan besaran *correct decision* 1- β yang sering disebut dengan kekuatan uji (*power of test*).

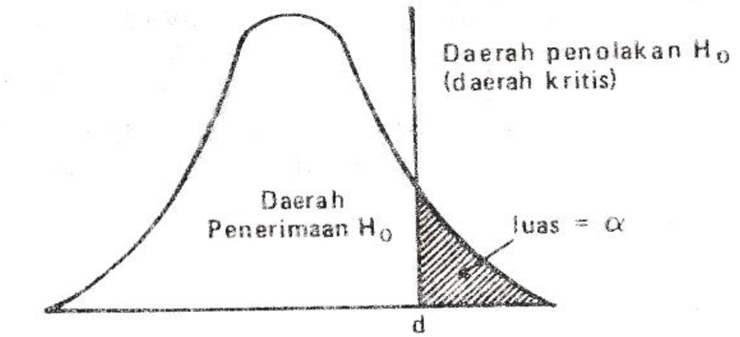
Berikut contoh frase pernyataan hipotesis yang dibuat disertai pengolahan datanya agar mengurangi kekeliruan, baik itu kekeliruan tipe I ataupun kekeliruan tipe II. Misalnya, bila hipotesis kita mengenai adanya perbedaan prestasi belajar, maka data yang kita peroleh harus kita olah kedalam rata-rata. Dan bila saja hipotesis kita mengenai adanya pengaruh sesuatu terhadap prestasi belajar, maka ukuran statistik yang kita peroleh adalah koefisien korelasi. Akan tetapi bila hipotesis itu berbunyi "Tidak ada perbedaan antara . . .", maka satu pun keadaan yang memenuhi hipotesis, secara keseluruhan keadaan itu sudah ditunjukkan kebenarannya. Secara garis besar pengujian hipotesis dibedakan menjadi tiga bagian yaitu; Pengujian hipotesis tentang pengaruh (Hipotesis Deskriptif); Pengujian hipotesis tentang hubungan (Hipotesis Asosiatif/Korelatif); Pengujian hipotesis tentang perbedaan (Hipotesis Komparatif). Berikut contoh frase dari ketiga jenis hipotesis tersebut.

1. Hipotesis Deskriptif:
   H_0 : Tidak semua siswa telah memiliki kemandirian belajar matematika dengan pembelajaran jarak jauh berbasis *WhatsApp group*.
   Ha : Semua siswa telah memiliki kemandirian belajar matematika dengan pembelajaran jarak jauh berbasis *WhatsApp group*.
2. Hipotesis Asosiatif (Korelatif):
   H_0 : Tidak adanya korelasi yang signifikan antara motivasi belajar dengan kemampuan berpikir kreatif siswa.
   Ha : Adanya korelasi yang signifikan antara motivasi belajar dengan kemampuan berpikir kreatif siswa.
3. Hipotesis Komparatif:
   H_0 : Tidak terdapat perbedaan yang signifikan mengenai peningkatan kemampuan komunikasi matematis siswa dengan *accelerated learning* dibandingkan pembelajaran biasa.
   Ha : Terdapat perbedaan yang signifikan mengenai peningkatan kemampuan komunikasi matematis siswa dengan *accelerated learning* dibandingkan pembelajaran biasa.

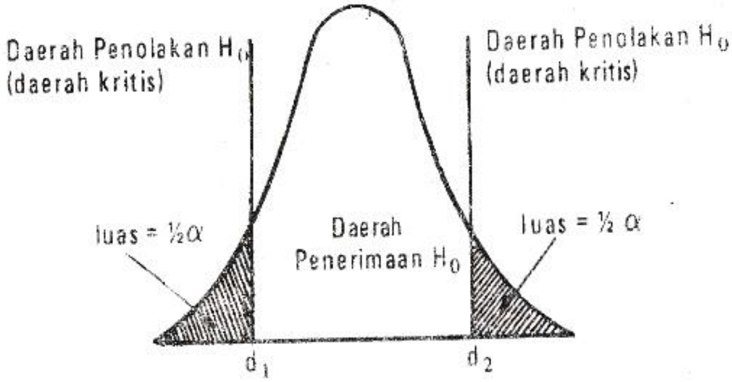
Dalam melihat kebenaran suatu hipotesis itu bahasanya "menguji hipotesis" tidak "membuktikan hipotesis." Hal yang perlu diperhatikan ketika menguji hipotesis yaitu mengenai nilai kritis. Fungsi nilai kritis antara lain memisahkan wilayah kritis dari wilayah *non* kritis. Lambang untuk nilai kritis adalah C.V. Daerah kritis atau daerah penolakan merupakan rentang nilai dari nilai uji yang menunjukkan adanya perbedaan yang signifikan dan hipotesis nol harus ditolak. Wilayah *non* kritis atau *non* penolakan adalah rentang nilai dari nilai uji yang menunjukkan bahwa perbedaan itu mungkin karena kebetulan dan hipotesis nol tidak boleh ditolak. Gambar berikut menunjukkan bagian-bagian yang termasuk daerah penolakan H_0, daerah penerimaan H_0, nilai kritis C.V dan α untuk uji satu pihak *(one tail)* baik itu pihak kiri ataupun pihak kanan, serta uji dua pihak dengan nilai alfa dibagi dua (α /2).



Gambar 9.2 Uji Pihak Kiri (Sudjana, 1989)



Gambar 9.3 Uji Pihak Kanan (Sudjana, 1989)



Gambar 9.4 Uji Dua Pihak (Sudjana, 1989)

Jika ujinya berekor kiri, daerah kritis, dengan luas sama dengan α, akan berada di pihak kiri rata-rata. Jika uji berekor kanan, daerah kritis, dengan luas sama dengan α, akan berada di pihak kanan rata-rata. Jika ujinya dua pihak, α harus dibagi 2; setengah dari luasnya berada di pihak kanan rata-rata, dan setengahnya berada di pihak kiri rata-rata. Ketika nilai uji berada di salah satu dari dua wilayah kritis hipotesis nol harus ditolak jika dalam uji dua pihak.

Suatu populasi berukuran N diambil n sampel dimana n <<< N. Dari peristiwa tersebut, jika ingin dilakukan penarikan kesimpulan mengenai populasi melalui sampel, maka dapat dilakukan apa yang disebut pengujian hipotesis. Karena populasi yang ingin dipelajari hanya satu maka disebut pengujian hipotesis satu populasi. Dari kasus di atas dapat dibuat hipotesis sebagai berikut:

1. Hipotesis dua pihak *(two tails)*
   $H_o : \mu = a$ $H_1 : \mu \neq a$
2. Hipotesis satu pihak *(one tail)*
   $H_o : \mu = a$ $H_1 : \mu > a$ (pihak kanan) $H_1 : \mu < a$ (pihak kiri)

Kata dua pihak ini didasarkan oleh kenyataan dari tanda (≠) dapat berarti < atau >. Untuk menguji hipotesis nol, peneliti menggunakan uji statistik. Banyak uji tes dihitung dengan menggunakan rumus di bawah ini.

$$Test\ value = \frac{Nilai\ yang\ diamati\ -\ Nilai\ yang\ diharapkan}{Standard\ error}$$

Tabel 9.1 berikut ini merupakan beberapa bentuk hipotesis dan macam-macam data sebagai pertimbangan ketika akan memilih uji statistik.

Tabel 9.1 Bentuk Hipotesis (Sugiyono, 2013)

Macam Data	BENTUK HIPOTESIS						
	Deskriptif	Komparatif 2 Sampel		Komparatif > 2 Sampel		Asosiatif	Model Struktural
		Related	Independen	*Related*	Independen		
Nominal	*Binomial Chi Satu Sampel*	*Mc Nemar*	*Fisher Exact Chi dua sampel*	*Chohran Q*	*Chii K sampel*	*Contingency Qoefficient C*	
Ordinal	*Run Test*	*Sign test, Wilcoxon*	*Median test*	*Anova*	*Anova*	*Spearman Korelasi Kendal Tau*	
Interval/Ratio	*t-test satu sampel*	*t-test hubungan*	*t-tes independen*	*Anova*	*Anova*	*Korelasi, Regresi*	*Path, SEM*

Tiga metode yang digunakan untuk menguji hipotesis adalah:

1. Metode tradisional
2. Metode P-*value*
3. Metode selang kepercayaan

Metode tradisional digunakan sejak metode pengujian hipotesis dirumuskan. Metode yang lebih baru, yang disebut metode P-*value*, telah menjadi populer dengan munculnya komputer modern dan kalkulator statistik. Metode ketiga, metode interval kepercayaan mengilustrasikan hubungan antara pengujian hipotesis dan interval kepercayaan. Metode tradisional pengujian hipotesis mengikuti lima langkah berikut; Nyatakan hipotesis dan identifikasi; Temukan nilai kritis; Hitung uji tes; Buat keputusan; Kesimpulan. Metode P-*value* untuk pengujian hipotesis bagi ahli statistik biasanya menguji hipotesis pada tingkat umum 0,05 atau 0,01 dan kadang-kadang pada 0,10 bergantung pada keseriusan kekeliruan tipe I. Selain mencantumkan nilai, banyak paket statistik komputer memberikan P-*value* untuk uji hipotesis. Aturan keputusan saat menggunakan P-*value*, jika P-*value* $\leq \alpha$ maka tolak H_0. Jika P-*value* $> \alpha$ maka terima H_0. Berikut panduan untuk P-*value*:

- Jika P-*value* $\leq$ 0,01, tolak H_0. Perbedaannya sangat signifikan.
- Jika P-*value* $>$ 0.01 tetapi P-*value* $\leq$ 0.05, tolak H_0. Perbedaannya signifikan.
- Jika P-*value* $>$ 0.05 tetapi P-*value* $\leq$ 0.10, pertimbangkan konsekuensi kesalahan tipe I sebelum menolak H_0.
- Jika P-*value* $>$ 0,10, terima H_0. Perbedaannya tidak signifikan.

Semua situasi pengujian hipotesis menggunakan metode P-*value* mengikuti lima langkah seperti metode tradisional. Metode tradisional dan metode P-*value* tidak boleh digunakan atas keinginan peneliti. Sebelum α dapat ditingkatkan, peneliti harus mempertimbangkan konsekuensi melakukan kekeliruan tipe I. Jika akibat ini lebih serius daripada akibat melakukan kekeliruan tipe II, maka α tidak boleh ditambah. Demikian juga, ada konsekuensi untuk meningkatkan ukuran sampel. Konsekuensi ini mungkin termasuk peningkatan jumlah uang yang dibutuhkan untuk melakukan studi dan peningkatan waktu yang dibutuhkan untuk tabulasi data. Ketika peneliti gagal menolak hipotesis nol, bukan hanya tidak ada cukup bukti untuk mendukung hipotesis alternatif. Mungkin hipotesis nol salah, tetapi uji statistik memiliki kekuatan yang terlalu rendah untuk mendeteksi perbedaan yang sebenarnya; karenanya, orang hanya dapat menyimpulkan bahwa dalam penelitian ini, tidak ada cukup bukti untuk menolak hipotesis nol. Hubungan antara α, β, dan kekuatan suatu uji dapat dianalisis lebih rinci namun tidak ada uji statistik yang dapat menjamin hasil yang sangat mudah ketika keputusan dibuat tentang validitas H_0. Apakah keputusannya adalah menolak H_0 atau menerima H_0 kedua kasus tersebut ada kemungkinan salah. Maka, menjaga probabilitas kekeliruan tipe I dan tipe II sekecil mungkin merupakan hal yang sangat penting.

Metode interval kepercayaan untuk menguji hipotesis, singkatnya, ketika H_0 ditolak pada tingkat signifikansi α, interval kepercayaan yang dihitung pada tingkat 1 - α tidak akan mengandung nilai rata-rata yang dinyatakan dalam H_0. Di sisi lain ketika H_0 diterima, interval kepercayaan yang dihitung pada tingkat signifikansi yang sama akan memuat nilai rata-rata yang dinyatakan dalam H_0. Hasil ini berlaku untuk situasi pengujian hipotesis lainnya dan tidak terbatas pada uji rata-rata.

C. UJI t

Setelah dijelaskan mengenai Uji Hipotesis, selanjutnya akan dijelaskan dua uji statistik khusus yang digunakan untuk hipotesis tentang rata-rata yaitu uji z dan uji t. Selain itu, prosedur pengujian hipotesis untuk menguji varian tunggal atau standar deviasi menggunakan distribusi chi-kuadrat akan dijelaskan pada sub bab selanjutnya. Pada bagian ini akan dijelaskan dua uji statistik yaitu uji z yang digunakan jika simpangan baku populasi diketahui, dan uji t digunakan jika simpangan baku populasi tidak diketahui. Bagian uji z akan dijelaskan secara singkat dan bagian mengenai Uji t akan dijelaskan lebih mendalam lagi. Dua uji statistik umum untuk hipotesis tentang rata-rata adalah uji z dan uji t. Uji z dapat digunakan ketika n ≥ 30, atau diasumsikan populasi menyebar normal dan ragam populasi (σ^2) diketahui, maka statistik uji (*Test statistiks*) dari kasus di atas adalah:

$$Z_{hitung} = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

Keterangan:

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

μ = rata-rata populasi yang dihipotesiskan

σ = Standar deviasi populasi

n = ukuran *sample*

Supaya kesimpulannya valid atau sah, jika hipotesisnya menggunakan Uji z, dengan α diberikan, dimana Z_{tabel} diambil dari tabel normal baku maka kesimpulan yang ditarik adalah tolak H_0 jika:

$|Z_{hitung}| > Z_{tabel\,(\frac{\alpha}{2})}$ untuk bentuk hipotesis dua pihak *(two tails)*

$|Z_{hitung}| > Z_{tabel\,(\alpha)}$ untuk bentuk hipotesis satu pihak *(one tail)*

Banyak peneliti tertarik untuk membandingkan dua parameter seperti dua rata-rata, dua proporsi, atau dua varian. Ketika kedua standar deviasi populasi diketahui, uji z dapat digunakan untuk membandingkan dua *mean*. Jika dari populasi tidak ada informasi tentang σ^2 atau ragam populasi tidak diketahui, maka statistik ujinya adalah Uji t, digunakan untuk rata-rata populasi ketika populasi berdistribusi normal atau mendekati normal dimana σ tidak diketahui, rumus untuk uji t adalah :

$$t_{hitung} = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

Keterangan:

s = simpangan baku dari data sampel

$\bar{X}$ = rata-rata sampel

n = banyaknya sampel

μ = nilai yang ditentukan untuk diuji

Rumus uji t mirip dengan rumus uji z. Tetapi karena deviasi standar populasi σ tidak diketahui, deviasi standar sampel s digunakan sebagai gantinya. Nilai kritis untuk uji t diberikan pada tabel distribusi t. Untuk uji satu pihak *(one tail)*, carilah tingkat dengan melihat baris *one tail* α paling atas dari tabel dan temukan kolom yang sesuai. Temukan derajat kebebasan (d.f.) dengan melihat kolom paling kiri. Perhatikan bahwa derajat kebebasan diberikan untuk nilai dari 1 sampai 30, ketika derajat kebebasan di atas 30, beberapa referensi buku akan memberitahu untuk menggunakan nilai tabel terdekat; atau menggunakan pendekatan konservatif yaitu harus selalu membulatkan ke bawah ke nilai tabel terdekat. Misalnya, jika d.f. = 59, karena di tabel distribusi t tidak ada maka gunakan d.f. = 55 untuk menemukan nilai kritis. Ketika derajat kebebasan semakin besar, nilai kritis mendekati nilai z. Oleh karena itu, nilai terbawah (ukuran sampel besar) sama dengan nilai z yang digunakan di bagian sebelumnya.

Hal yang membedakan distribusi t dengan distribusi normal standar yaitu; Variansnya lebih besar dari 1; Distribusi t adalah keluarga kurva berdasarkan derajat kebebasan, yang merupakan angka yang terkait dengan ukuran sampel; Semakin besar ukuran sampel, distribusi t mendekati distribusi normal. Ciri-ciri distribusi variabel mendekati distribusi normal, antara lain:

1. Berbentuk lonceng
2. Simetris terhadap rata-rata
3. Rata-rata, median, dan modus sama dengan 0 dan terletak di pusat distribusi.
4. Kurva tidak pernah menyentuh sumbu x.

<u>Contoh soal 1:</u>

Tentukan nilai t kritis dengan derajat kebebasannya 13 dan α = 0,05 dengan uji pihak kanan, uji pihak kiri dan uji dua pihak.

<u>Solusi</u>

- Uji pihak kanan

  Dari tabel distribusi t dengan nilai derajat kebebasan (kolom paling kiri) = 13 tarik ke sebelah kanan dan *one tail* α (baris kedua dari atas) = 0,05 tarik ke bawah. Nilai kritis terletak diantara perpotongan garis tersebut yaitu 1,771.
- Uji pihak kiri

  Ketika menggunakan uji pihak kiri gunakan tabel uji pihak kiri, dari tabel distribusi t dengan nilai derajat kebebasan (kolom paling kiri) = 13 tarik ke sebelah kanan dan *one tail* α (baris kedua dari atas) = 0,05 tarik ke bawah. Nilai kritis terletak diantara perpotongan garis tersebut yaitu 1,771. Karena menggunakan uji pihak kiri maka nilai kritisnya menjadi 0 – 1,771 = -1,771.
- Uji dua pihak

  dari tabel distribusi dengan nilai derajat kebebasan (kolom paling kiri) = 13 tarik ke sebelah kanan dan *two tail* α (baris ketiga dari atas) = 0,05 tarik ke bawah. Nilai kritis terletak diantara perpotongan garis tersebut yaitu 2,160 dan -2,160.

Supaya kesimpulannya valid atau sah, jika hipotesisnya menggunakan Uji t, dengan α diberikan dimana t_{tabel} diambil dari tabel normal baku atau diambil dari tabel distribusi t dengan derajat kebebasan (d.f.) = n-1, maka kesimpulan yang ditarik adalah tolak H_0 jika:

$$\left|t_{hitung}\right| > t_{tabel\,(\frac{\alpha}{2},df)}$$ untuk bentuk hipotesis dua pihak *(two tails)*

$$\left|t_{hitung}\right| > t_{tabel\,(\alpha,df)}$$ untuk bentuk hipotesis satu pihak *(one tail)*

Saat Anda menguji hipotesis menggunakan uji t dengan metode tradisional, ikuti prosedur yang sama dengan uji z dengan menggunakan tabel distribusi t.

1. Nyatakan hipotesis dan identifikasi klaim.
2. Temukan nilai kritis dari tabel distribusi t.
3. Hitung nilai tes.
4. Buat keputusan untuk menolak atau tidak menolak hipotesis nol.
5. Kesimpulan.

Untuk menginterpretasikan uji t (*t-test*) terlebih dahulu harus ditentukan:

- Nilai signifikansi α
- d.f. (*degree of freedom*) = n - k
- Bandingkan nilai t_{hitung} dengan $t_{tabel\,(\alpha,\,n-1)}$
- Apabila menggunakan uji satu pihak:

  $t_{hitung} > t_{tabel\,(\alpha,\,n-1)}$ → berbeda secara signifikansi (H_0 ditolak)

  $t_{hitung} < t_{tabel\,(\alpha,\,n-1)}$ → tidak berbeda secara signifikansi (H_0 diterima)

Tabel 9.2 Kesimpulan Uji-t

Jenis Uji	Hipotesis	Kesimpulan Uji-t
Uji dua pihak	H_0 : $\mu_1 = \mu_2$ H_1 : $\mu_1 \neq \mu_2$	Jika $\lvert t_{hitung} \rvert \leq t_{tabel}$ Maka H_0 diterima
Uji satu pihak (pihak kanan)	H_0 : $\mu_1 \leq \mu_2$ H_1 : $\mu_1 > \mu_2$	Jika $t_{hitung} \leq t_{tabel}$ Maka H_0 diterima
Uji satu pihak (pihak kiri)	H_0 : $\mu_1 \geq \mu_2$ H_1 : $\mu_1 < \mu_2$	Jika $t_{hitung} \geq t_{tabel}$ Maka H_0 diterima

<u>Contoh soal 2</u>

Suatu penelitian telah dilakukan untuk mengetahui kemampuan pemecahan masalah matematis yang menggunakan pembelajaran jarak jauh berbasis WhatsApp Group dalam bentuk nilai *posttest* siswa, data seperti pada tabel 9.3.

Tabel 9.3 Data Hasil Pengamatan Skor *Posttest*

No. Sampel	Skor *Posttest*	No.Sampel	Skor *Posttest*
1	31	9	45
2	32	10	34
3	30	11	0
4	26	12	0
5	0	13	0
6	21	14	45
7	0	15	0
8	15		

Jika diketahui ambang batas skor *posttest* yaitu 25 dan skor *posttest* rata-rata adalah 18,6. Ujilah apakah skor *posttest* yang didapat sudah melebihi ambang batas yang ditentukan. Asumsikan data berasal dari populasi berdistribusi normal dengan $\alpha = 0{,}05$.

Solusi

Langkah 1 : Nyatakan hipotesis dan identifikasi klaim.

$H_0 : \mu = 10$ $\quad\quad$ $H_1 : \mu > 10$ (klaim)

Langkah 2 : Temukan nilai kritis.

Sesuai klaim hipotesis H_1 maka gunakan uji pihak kanan. Dari tabel distribusi t dengan nilai derajat kebebasan (kolom paling kiri) n – 1 yaitu 15 - 1 = 14 tarik ke sebelah kanan dan α (baris paling atas) = 0,05 tarik ke bawah. Nilai kritis terletak diantara perpotongan garis tersebut yaitu 1,761

Langkah 3 : Hitung nilai tes.

$$|t_{hitung}| = \left|\frac{\bar{X}-\mu}{s/\sqrt{n}}\right|$$

Mencari nilai s dengan menggunakan rumus varians sampel berikut:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1} = \frac{(31-18{,}6)^2+(32-18{,}6)^2+\cdots+(0-18{,}6)^2}{14} = 303{,}114$$

s = 17,41

$$|t_{hitung}| = \left|\frac{\bar{X}-\mu}{s/\sqrt{n}}\right| = \left|\frac{18{,}6-10}{\frac{17{,}41}{\sqrt{15}}}\right| = 1{,}911$$

Langkah 4 : Buat keputusan untuk menolak atau tidak menolak hipotesis nol.

Tolak hipotesis H_0 , karena t_{hitung} yaitu 1,911 berada pada daerah nilai kritis t_{tabel} yaitu 1,761 dengan kata lain $t_{hitung} > t_{tabel}$.

Langkah 5 : Kesimpulan

Ada cukup bukti untuk mendukung pernyataan bahwa skor *posttest* yang didapat sudah melebihi ambang batas yang ditentukan.

Ingatlah bahwa uji t harus digunakan ketika populasi berdistribusi hampir normal dan standar deviasi populasi tidak diketahui. P-*value* untuk uji t dapat ditemukan dengan menggunakan tabel distribusi t; namun, P-*value* spesifik untuk uji t tidak dapat diperoleh dari tabel karena hanya nilai α terpilih (misalnya, 0,01, 0,05) yang diberikan. Untuk menemukan P-*value* spesifik untuk uji t memerlukan tabel yang mirip dengan tabel *Cumulative Standard Normal Distribution* untuk setiap derajat kebebasan. Karena ini tidak praktis, hanya interval yang dapat ditemukan untuk P-*value* maka banyak yang akan menggunakan kalkulator atau program komputer yang memberikan P-*value* spesifik untuk uji t. Untuk menguji hipotesis menggunakan metode P-*value*, mengikuti langkah-langkah yang hampir sama seperti yang dijelaskan metode tradisional.

D. UJI CHI KUADRAT (X^2)

Banyak percobaan yang hasilnya lebih dari dua jenis. Jika dalam eksperimen atau percobaan memiliki lebih dari dua hasil keluaran maka untuk menguji hipotesisnya menggunakan Uji Chi Kuadrat (*Chi Square Testing*) dilambangkan χ^2. Biasanya distribusi Chi Kuadrat digunakan untuk menyusun interval kepercayaan varians tunggal. Ketika akan mencari luas dibawah distribusi Chi Kuadrat maka gunakan tabel distribusi Chi Kuadrat. Hal yang harus dipertimbangkan ketika menggunakan Uji Chi Kuadrat, yaitu: Menentukan nilai kritis Chi Kuadrat dengan α tertentu baik ketika uji hipotesis pihak kanan, pihak kiri ataupun uji dua pihak.

<u>Contoh soal 3</u>

Tentukan nilai kritis Chi Kuadrat dengan nilai derajat kebebasannya 17 dan α = 0,05 dengan uji pihak kanan, uji pihak kiri dan uji dua pihak

Solusi

- Uji pihak kanan
  Dari tabel distribusi Chi Kuadrat dengan nilai derajat kebebasan (kolom paling kiri) = 17 tarik ke sebelah kanan dan α (baris paling atas) = 0,05 tarik ke bawah. Nilai kritis terletak diantara perpotongan garis tersebut yaitu 27,587.
- Uji pihak kiri
  Ketika menggunakan uji pihak kiri gunakan tabel uji pihak kiri, karena tabel Chi Kuadrat memberikan luas di sebelah kanan nilai kritis dan statistik Chi Kuadrat tidak boleh negatif maka tabel diatur sedemikian rupa sehingga memberikan nilai luas di sebelah kanan nilai kritis. Nilai α dikurangkan dulu terhadap 1, sehingga menjadi 1 – 0,05 = 0,95. Dalam hal ini, 95 % area akan berada di sebelah kanan nilainya. Dari tabel distribusi Chi Kuadrat dengan nilai derajat kebebasan (kolom paling kiri) = 17 tarik ke sebelah kanan dan α (baris paling atas) = 0,95 tarik ke bawah. Nilai kritis terletak diantara perpotongan garis tersebut yaitu 8,672.
- Uji dua pihak
  Saat menggunakan uji dua pihak area tersebut harus dibagi. Perlu diperhatikan bahwa luas di sebelah kanan yang lebih besar yaitu: $\alpha/2$ = 0,05/2 = 0,025 dan luas di sebelah kanan yang lebih kecil yaitu: 1 – 0,025 = 0,975. Ingat nilai Chi-Kuadrat tidak boleh negatif, oleh karena itu gunakan 0,025 dan 0,975 dari tabel distribusi Chi Kuadrat dengan nilai derajat kebebasan (kolom paling kiri) = 17 tarik ke sebelah kanan dan α (baris paling atas) = 0,025 dan 0,975 tarik ke bawah. Nilai kritis terletak diantara perpotongan garis tersebut yaitu 30,191 dan 7,564.

Beberapa asumsi Uji Chi Kuadrat untuk varians tunggal, antara lain:

1. Sampel harus dipilih secara acak dari populasi.
2. Populasi harus berdistribusi normal untuk variabel yang diteliti.
3. Pengamatan harus independen satu sama lain.

Saat uji χ^2 adalah uji dua pihak, kedua nilai interval harus digandakan. Misal, uji dua pihak interval adalah 2(0,01) < P-*Value* < 2(0,025), atau 0,02 < P-*Value* < 0,05. Metode P-*Value* pengujian hipotesis untuk Chi Kuadrat mengikuti langkah-langkah yang hampir sama dengan metode tradisional. Rumus Uji Chi Kuadrat (*Chi Square Test*) untuk varians tunggal yaitu:

$\chi 2 = \frac{(n-1)s^2}{\sigma^2}$, dimana derajat kebebasan (d.f.) = n – 1

Keterangan:
n = ukuran sampel
s^2 = varians sampel
σ^2 = varians populasi

Ketika akan melakukan perhitungan praktis gunakan hipotesis awal (H_0) dan Hipotesis alternatif (H_1) pada waktu melakukan uji Hipotesis. Sedangkan untuk melakukan uji statistik, bandingkan nilai χ^2 hasil perhitungan dengan nilai χ^2 dari tabel (nilai χ^2 kritis). Cara untuk mendapatkan nilai χ^2 kritis atau nilai χ^2_{tabel} yaitu dengan menggunakan derajat kebebasan (*degree of freedom*) dan derajat signifikansi (α). Semakin besar derajat kebebasan maka grafik distribusi χ^2 akan mendekati bentuk distribusi normal.

Contoh Soal 4

Seorang dosen ingin melihat apakah variasi nilai ujian akhir semester matakuliah persamaan differensial 21 mahasiswa lebih kecil dari variansi populasi. Varian dari kelas tersebut adalah 45,71. Apakah ada cukup bukti untuk mendukung pernyataan bahwa varians nilai ujian akhir semester matakuliah persamaan differensial mahasiswa kurang dari varians populasi ($\sigma^2 = 64{,}87$) dengan α = 0,05. Asumsikan bahwa nilai ujian akhir semester matakuliah persamaan differensial berdistribusi normal.

Solusi

Langkah 1 : Nyatakan hipotesis dan identifikasi klaim.

$H_0 : \sigma^2 = 64{,}87$ $\qquad$ $H_1 : \sigma^2 < 64{,}87$ (klaim)

Langkah 2 : Temukan nilai kritis.

Ketika menggunakan uji pihak kiri gunakan tabel uji pihak kiri, karena tabel Chi Kuadrat memberikan luas di sebelah kanan nilai kritis dan statistik Chi Kuadrat tidak boleh negatif maka tabel diatur sedemikian rupa sehingga memberikan nilai luas di sebelah kanan nilai kritis. Nilai α dikurangkan dulu terhadap 1, sehingga menjadi 1 – 0,05 = 0,95. Dalam hal ini, 95 % area akan berada di sebelah kanan nilainya. Dari tabel distribusi Chi Kuadrat dengan nilai derajat kebebasan (kolom paling kiri) n - 1 = 21 - 1 = 20 tarik ke sebelah kanan dan α (baris paling atas) = 0,95 tarik ke bawah. Nilai kritis terletak diantara perpotongan garis tersebut yaitu 10,851

Langkah 3 : Hitung nilai tes.

$$\chi 2 = \frac{(n-1)s^2}{\sigma^2} = \frac{20 \; x \; 45{,}71}{64{,}87} = 14{,}09$$

Langkah 4 : Buat keputusan untuk menolak atau tidak menolak hipotesis nol.
Terima hipotesis H_0 , karena χ^2_{hitung} yaitu 14,09 bukan berada pada daerah nilai kritis χ^2_{tabel} yaitu 10,851

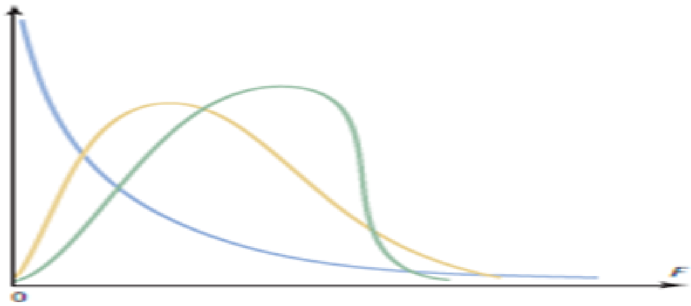
Langkah 5 : Kesimpulan
Tidak ada cukup bukti untuk mendukung pernyataan bahwa varians nilai ujian akhir semester matakuliah persamaan differensial mahasiswa kurang dari varians populasi.

E. UJI-F

Dua varian dapat dibandingkan dengan menggunakan uji F, untuk menentukan apakah dua varian sampel sama, peneliti dapat menggunakan uji F. Bagaimanapun tidak semua ahli statistik setuju untuk menggunakan uji F yang sebelumnya menggunakan uji t. Beberapa percaya bahwa melakukan uji F dan uji t pada tingkat signifikansi yang sama akan mengubah tingkat signifikansi keseluruhan dari uji t. Salah satu contoh uji perbedaan antara dua varians atau standar deviasi yaitu: Apakah perbedaan temperatur pada musim tertentu di dua negara berbeda? Untuk perbandingan dua varians atau standar deviasi, uji F digunakan. Uji F jangan dibingungkan dengan uji chi-kuadrat, yang membandingkan varians sampel tunggal dengan varians populasi tertentu. Jika dua sampel independen dipilih dari dua populasi terdistribusi normal di mana variansnya sama ($\sigma_1^2 = \sigma_2^2$) dan jika varians s_1^2 dan s_2^2 dibandingkan sebagai $\frac{s_1^2}{s_2^2}$, distribusi pengambilan sampel dari varian disebut distribusi F. Karakteristik distribusi F antara lain:

1. Nilai F tidak boleh negatif, karena varian selalu positif atau nol.
2. Distribusi miring positif.

3. Nilai rata-rata F kira-kira sama dengan 1.
4. Distribusi F adalah keluarga kurva berdasarkan derajat kebebasan varian pembilang dan derajat kebebasan varian penyebut



Gambar 9.5 Bentuk-Bentuk Kurva Distribusi F (Bluman, Allan G., 2009)

Pengujian hipotesis untuk uji-F mengikuti langkah-langkah yang hampir sama dengan metode tradisional, adapun rumus uji-F sebagai berikut:

$$F = \frac{s_1^2}{s_2^2}$$

di mana yang lebih besar dari dua varians ditempatkan di pembilang terlepas dari subskripnya. Uji F memiliki dua suku yaitu derajat kebebasan pembilang atau *degrees of freedom numerator* (d.f.N) = n_1-1, dan derajat kebebasan pembilang penyebut atau *degrees of freedom denominator* (d.f.D) = n_2-1, di mana n_1 adalah sampel yang menghasilkan varian lebih besar. Tabel distribusi F menunjukkan nilai kritis F untuk α = 0,005, α = 0,01, α = 0,025, α = 0,05, α = 0,1. Nilai tersebut untuk satu pihak *(one tail)*, jika untuk uji dua pihak maka nilai alfa dibagi dua (α /2). Hal yang harus diperhatikan ketika menggunakan uji F:

1. Varians yang lebih besar selalu ditempatkan di pembilang rumus
   F = $\frac{s_1^2}{s_2^2}$
2. Jika menggunakan uji dua pihak nilai alfa dibagi dua (α/2) dan nilai kritis ditempatkan di pihak kanan kurva F.
3. Jika standar deviasi diberikan dalam soal, maka harus dikuadratkan untuk rumus uji F.
4. Jika pada tabel distribusi F derajat kebebasan tidak ditemukan, maka nilai terdekat pada pihak yang lebih kecil harus digunakan. Misal pada tabel distribusi F tidak ditemukan derajat kebebasan yang tepat, nilai terdekat yang

lebih kecil harus digunakan. Contohnya, jika α = 0,05 (uji pihak kanan), d.f.N = 18, dan d.f.D = 20, gunakan kolom d.f.N = 15 dan baris d.f.D = 20 untuk mendapatkan F = 2,20.

Contoh soal 5

Tentukan nilai kritis (*critical value*) dengan menggunakan uji F pihak kanan, jika diketahui α = 0,05, derajat kebebasan pembilang = 14 dan derajat kebebasan penyebut = 20.

Solusi

Ketika menggunakan uji F pihak kanan dengan α = 0,05 maka gunakan tabel 0,05. d.f.N (baris bagian atas) = 14 derajat kebebasan tidak ditemukan, maka nilai terdekat pada pihak yang lebih kecil harus digunakan yaitu d.f.N (baris bagian atas) = 12 tarik ke bawah dan d.f.D (kolom paling kiri) = 20 tarik ke kanan. Perpotongan garis d.f.N = 12 dengan garis d.f.D = 20 merupakan nilai kritis kasus tersebut yaitu 2,28.

Nilai kritis untuk uji hipotesis pihak kiri dapat ditemukan dengan menukar derajat kebebasan dan mengambil kebalikan dari nilai yang ditemukan di tabel distribusi F. Saat melakukan uji F harus berhati-hati karena data bisa bertentangan dengan hipotesisnya. Misal, jika hipotesisnya $H_0 : \sigma_1^2 \leq \sigma_2^2$ (ditulis $H_0 : \sigma_1^2 = \sigma_2^2$) dan $H_1 : \sigma_1^2 > \sigma_2^2$. Tetapi jika $s_1^2 < s_2^2$, maka uji F tidak boleh dilakukan dan tidak akan menolak hipotesis nol. Hipotesis yang digunakan untuk menguji kesetaraan dua varian:

Uji Pihak Kanan	Uji Pihak Kiri	Uji Dua Pihak
$H_0 : \sigma_1^2 = \sigma_2^2$	$H_0 : \sigma_1^2 = \sigma_2^2$	$H_0 : \sigma_1^2 = \sigma_2^2$
$H_1 : \sigma_1^2 > \sigma_2^2$	$H_1 : \sigma_1^2 < \sigma_2^2$	$H_1 : \sigma_1^2 \neq \sigma_2^2$

Beberapa asumsi ketika menguji perbedaan antara dua varians:

1. Populasi dimana sampel diambil harus berdistribusi normal.
2. Uji tidak boleh digunakan ketika distribusi menyimpang dari normalitas.
3. Sampel harus *independent*.

Contoh Soal 6

Peneliti ingin mengetahui apakah varian *Posttest* siswa yang mendapatkan model pembelajaran *Accelerated Learning* berbeda dengan varian model pembelajaran biasa. Dua sampel dipilih yaitu kelas eksperimen yang menggunakan model pembelajaran *Accelerated Learning* dan kelas kontrol menggunakan model pembelajaran biasa. Data seperti yang ditunjukkan di bawah, dengan α = 0,05 apakah ada cukup bukti untuk mendukung pernyataan tersebut.

Kelas Kontrol	**Kelas Eksperimen**
n_1 = 31	n_2 = 31
s_1^2 = 66,51	s_2^2 = 37,23

Solusi

Langkah 1 : Nyatakan hipotesis dan identifikasi klaim.

$H_0 : \sigma_1^2 = \sigma_2^2$ $\qquad$ $H_1 : \sigma_1^2 \neq \sigma_2^2$ (klaim)

Langkah 2 : Temukan nilai kritis.

(d.f.N) = n_1-1 = 31 – 1 = 30 $\qquad$ (d.f.D) = n_2-1 = 31 – 1 = 30

Karena uji dua pihak yang diklaim maka $\alpha/2$ yaitu 0,05/2 = 0,025

Dari tabel distribusi F didapat nilai kritis 2,07

Langkah 3 : Hitung nilai tes.

$F = \frac{s_1^2}{s_2^2}$ = 66,51/37,23 = 1,7865

Langkah 4 : Buat keputusan untuk menolak atau tidak menolak hipotesis nol.

Terima hipotesis H_0 , karena $F_{hitung} < F_{tabel}$ yaitu 1,7865 < 2,07

Langkah 5 : Kesimpulan

Tidak cukup bukti untuk mendukung klaim bahwa variansi *posttest* kelas kontrol yang menggunakan model pembelajaran biasa dan menggunakan model pembelajaran *Accelerated Learning* berbeda.

F. PENUTUP

Sebagai penutup dari bab ini, ketika pengumpulan data dilakukan dengan cara sensus maka pengambilan keputusan menjadi sah ketika hanya dilakukan secara deskriptif melalui gambaran umum data. Dalam hal ini, statistika inferensial mengambil peranan yang sangat penting dalam proses pengambilan keputusan. Seperti sudah disebutkan, hipotesis dirumuskan setelah studi literatur dilakukan dan sebelum penelitian dilakukan. Logikanya, hipotesis itu lanjutan dari studi literatur dan dibuat berdasarkan kepada penerapan penelitian-penelitian sebelumnya. Hipotesis mendahului studi yang sebenarnya sebab seluruh studi diarahkan oleh hipotesis; subjek, instrumen atau alat-alat ukur, disain dan prosedur penelitian beserta cara menganalisis dan penarikan kesimpulan diarahkan oleh hipotesis. Dalam penelitian kuantitatif, hipotesis memiliki peran yang penting. Karena hipotesis memberi arah yang jelas kepada peneliti dalam rangka melakukan verifikasi menuju terwujudnya suatu kesimpulan. Ciri-ciri hipotesis baik ; sejalan dengan hasil penelitian sebelumnya; *tentative* dan berupa penjelasan yang masuk akal bagi terbentuknya tingkah laku tertentu, gejala (fenomena), atau kejadian. Hipotesis nol menyatakan bahwa tidak ada perbedaan, dan hipotesis alternatif menentukan perbedaan.

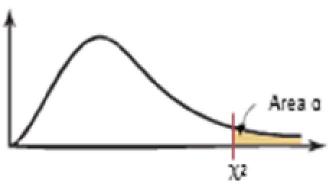
Uji z digunakan baik ketika standar deviasi populasi diketahui dan variabel terdistribusi normal atau ketika σ diketahui dan ukuran sampel lebih besar dari atau sama dengan 30. Ketika standar deviasi populasi tidak diketahui dan variabel terdistribusi normal, standar deviasi sampel digunakan, tetapi uji t harus dilakukan sebagai gantinya. Jika kedua simpangan baku populasi tidak diketahui, simpangan baku sampel dapat digunakan, tetapi uji t digunakan untuk membandingkan dua *mean*. Apabila suatu eksperimen menghasilkan lebih dari dua keluaran maka Uji Chi Kuadrat (*Chi Square Testing*) cocok untuk menguji hipotesisnya. Akhirnya, dua varians dapat dibandingkan menggunakan uji F.

Tabel Distribusi Chi Kuadrat

Degrees of freedom	α									
	0.995	**0.99**	**0.975**	**0.95**	**0.90**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**
1	—	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.299
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289	42.796
23	9.262	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.257	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

Source: Donald B. Owen, *Handbook of Statistics Tables, The Chi-Square Distribution Table*, © 1962 *by Addison-Wesley*

Publishing Company, Inc. Copyright renewal © 1990. Reprinted by permission of Pearson Education, Inc.



Tabel Distribusi F

d.f.D.: degrees of freedom, denominator	α = 0.005																		
	d.f.N.: degrees of freedom, numerator																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	16,211	20,000	21,615	22,500	23,056	23,437	23,715	23,925	24,091	24,224	24,426	24,630	24,836	24,940	25,044	25,148	25,253	25,359	25,465
2	198.5	199.0	199.2	199.2	199.3	199.3	199.4	199.4	199.4	199.4	199.4	199.4	199.4	199.5	199.5	199.5	199.5	199.5	199.5
3	55.55	49.80	47.47	46.19	45.39	44.84	44.43	44.13	43.88	43.69	43.39	43.08	42.78	42.62	42.47	42.31	42.15	41.99	41.83
4	31.33	26.28	24.26	23.15	22.46	21.97	21.62	21.35	21.14	20.97	20.70	20.44	20.17	20.03	19.89	19.75	19.61	19.47	19.32
5	22.78	18.31	16.53	15.56	14.94	14.51	14.20	13.96	13.77	13.62	13.38	13.15	12.90	12.78	12.66	12.53	12.40	12.27	12.14
6	18.63	14.54	12.92	12.03	11.46	11.07	10.79	10.57	10.39	10.25	10.03	9.81	9.59	9.47	9.36	9.24	9.12	9.00	8.88
7	16.24	12.40	10.88	10.05	9.52	9.16	8.89	8.68	8.51	8.38	8.18	7.97	7.75	7.65	7.53	7.42	7.31	7.19	7.08
8	14.69	11.04	9.60	8.81	8.30	7.95	7.69	7.50	7.34	7.21	7.01	6.81	6.61	6.50	6.40	6.29	6.18	6.06	5.95
9	13.61	10.11	8.72	7.96	7.47	7.13	6.88	6.69	6.54	6.42	6.23	6.03	5.83	5.73	5.62	5.52	5.41	5.30	5.19
10	12.83	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97	5.85	5.66	5.47	5.27	5.17	5.07	4.97	4.86	4.75	4.64
11	12.23	8.91	7.60	6.88	6.42	6.10	5.86	5.68	5.54	5.42	5.24	5.05	4.86	4.76	4.65	4.55	4.44	4.34	4.23
12	11.75	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20	5.09	4.91	4.72	4.53	4.43	4.33	4.23	4.12	4.01	3.90
13	11.37	8.19	6.93	6.23	5.79	5.48	5.25	5.08	4.94	4.82	4.64	4.46	4.27	4.17	4.07	3.97	3.87	3.76	3.65
14	11.06	7.92	6.68	6.00	5.56	5.26	5.03	4.86	4.72	4.60	4.43	4.25	4.06	3.96	3.86	3.76	3.66	3.55	3.44
15	10.80	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54	4.42	4.25	4.07	3.88	3.79	3.69	3.58	3.48	3.37	3.26
16	10.58	7.51	6.30	5.64	5.21	4.91	4.69	4.52	4.38	4.27	4.10	3.92	3.73	3.64	3.54	3.44	3.33	3.22	3.11
17	10.38	7.35	6.16	5.50	5.07	4.78	4.56	4.39	4.25	4.14	3.97	3.79	3.61	3.51	3.41	3.31	3.21	3.10	2.98
18	10.22	7.21	6.03	5.37	4.96	4.66	4.44	4.28	4.14	4.03	3.86	3.68	3.50	3.40	3.30	3.20	3.10	2.99	2.87
19	10.07	7.09	5.92	5.27	4.85	4.56	4.34	4.18	4.04	3.93	3.76	3.59	3.40	3.31	3.21	3.11	3.00	2.89	2.78
20	9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96	3.85	3.68	3.50	3.32	3.22	3.12	3.02	2.92	2.81	2.69
21	9.83	6.89	5.73	5.09	4.68	4.39	4.18	4.01	3.88	3.77	3.60	3.43	3.24	3.15	3.05	2.95	2.84	2.73	2.61
22	9.73	6.81	5.65	5.02	4.61	4.32	4.11	3.94	3.81	3.70	3.54	3.36	3.18	3.08	2.98	2.88	2.77	2.66	2.55
23	9.63	6.73	5.58	4.95	4.54	4.26	4.05	3.88	3.75	3.64	3.47	3.30	3.12	3.02	2.92	2.82	2.71	2.60	2.48
24	9.55	6.66	5.52	4.89	4.49	4.20	3.99	3.83	3.69	3.59	3.42	3.25	3.06	2.97	2.87	2.77	2.66	2.55	2.43
25	9.48	6.60	5.46	4.84	4.43	4.15	3.94	3.78	3.64	3.54	3.37	3.20	3.01	2.92	2.82	2.72	2.61	2.50	2.38
26	9.41	6.54	5.41	4.79	4.38	4.10	3.89	3.73	3.60	3.49	3.33	3.15	2.97	2.87	2.77	2.67	2.56	2.45	2.33
27	9.34	6.49	5.36	4.74	4.34	4.06	3.85	3.69	3.56	3.45	3.28	3.11	2.93	2.83	2.73	2.63	2.52	2.41	2.25
28	9.28	6.44	5.32	4.70	4.30	4.02	3.81	3.65	3.52	3.41	3.25	3.07	2.89	2.79	2.69	2.59	2.48	2.37	2.29
29	9.23	6.40	5.28	4.66	4.26	3.98	3.77	3.61	3.48	3.38	3.21	3.04	2.86	2.76	2.66	2.56	2.45	2.33	2.24
30	9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45	3.34	3.18	3.01	2.82	2.73	2.63	2.52	2.42	2.30	2.18
40	8.83	6.07	4.98	4.37	3.99	3.71	3.51	3.35	3.22	3.12	2.95	2.78	2.60	2.50	2.40	2.30	2.18	2.06	1.93

Tabel Distribusi F (Lanjutan)

d.f.D.: degrees of freedom, denominator	α = 0.01 d.f.N.: degrees of freedom, numerator 1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4052	4999.5	5403	5625	5764	5859	5928	5982	6022	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80

Tabel Distribusi F (Lanjutan)

d.f.D.: degrees of freedom, denominator	α = 0.025																		
	d.f.N.: degrees of freedom, numerator																		
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**12**	**15**	**20**	**24**	**30**	**40**	**60**	**120**	**∞**
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3	968.6	976.7	984.9	993.1	997.2	1001	1006	1010	1014	1018
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43	39.45	39.46	39.46	39.47	39.48	39.49	39.50
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	13.90
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4.42	4.36	4.31	4.25	4.20	4.14
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.46	2.40
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	2.32
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.82	2.72	2.62	2.56	2.50	2.44	2.38	2.32	2.25
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.77	2.67	2.56	2.50	2.44	2.38	2.32	2.26	2.19
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	2.13
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	2.04
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	2.00
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	2.67	2.57	2.47	2.36	2.30	2.24	2.18	2.11	2.04	1.97
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	1.91
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65	2.59	2.49	2.39	2.28	2.22	2.16	2.09	2.03	1.95	1.88
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63	2.57	2.47	2.36	2.25	2.19	2.13	2.07	2.00	1.93	1.85
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61	2.55	2.45	2.34	2.23	2.17	2.11	2.05	1.98	1.91	1.83
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59	2.53	2.43	2.32	2.21	2.15	2.09	2.03	1.96	1.89	1.81
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64

Tabel Distribusi F (Lanjutan)

d.f.D.: degrees of freedom, denominator	α = 0.05																		
	d.f.N.: degrees of freedom, numerator																		
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**12**	**15**	**20**	**24**	**30**	**40**	**60**	**120**	**∞**
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

Tabel Distribusi F (Lanjutan)

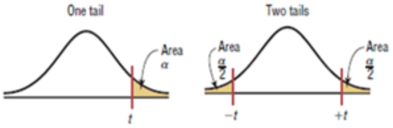
d.f.D.: degrees of freedom, denominator	α = 0.10																		
	d.f.N.: degrees of freedom, numerator																		
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**12**	**15**	**20**	**24**	**30**	**40**	**60**	**120**	**∞**
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38

From M. Merrington and C. M. Thompson (1943). Table of Percentage Points of the Inverted Beta (F) Distribution. Biometrika 33, pp. 74–87. Reprinted with permission from Biometrika.

Tabel Distribusi t

d.f.	*Confidence intervals*	80%	90%	95%	98%	99%
	One tail, α	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**
	Two tails, α	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**
1		3.078	6.314	12.706	31.821	63.657
2		1.886	2.920	4.303	6.965	9.925
3		1.638	2.353	3.182	4.541	5.841
4		1.533	2.132	2.776	3.747	4.604
5		1.476	2.015	2.571	3.365	4.032
6		1.440	1.943	2.447	3.143	3.707
7		1.415	1.895	2.365	2.998	3.499
8		1.397	1.860	2.306	2.896	3.355
9		1.383	1.833	2.262	2.821	3.250
10		1.372	1.812	2.228	2.764	3.169
11		1.363	1.796	2.201	2.718	3.106
12		1.356	1.782	2.179	2.681	3.055
13		1.350	1.771	2.160	2.650	3.012
14		1.345	1.761	2.145	2.624	2.977
15		1.341	1.753	2.131	2.602	2.947
16		1.337	1.746	2.120	2.583	2.921
17		1.333	1.740	2.110	2.567	2.898
18		1.330	1.734	2.101	2.552	2.878
19		1.328	1.729	2.093	2.539	2.861
20		1.325	1.725	2.086	2.528	2.845
21		1.323	1.721	2.080	2.518	2.831
22		1.321	1.717	2.074	2.508	2.819
23		1.319	1.714	2.069	2.500	2.807
24		1.318	1.711	2.064	2.492	2.797
25		1.316	1.708	2.060	2.485	2.787
26		1.315	1.706	2.056	2.479	2.779
27		1.314	1.703	2.052	2.473	2.771
28		1.313	1.701	2.048	2.467	2.763
29		1.311	1.699	2.045	2.462	2.756
30		1.310	1.697	2.042	2.457	2.750
32		1.309	1.694	2.037	2.449	2.738
34		1.307	1.691	2.032	2.441	2.728
36		1.306	1.688	2.028	2.434	2.719
38		1.304	1.686	2.024	2.429	2.712
40		1.303	1.684	2.021	2.423	2.704
45		1.301	1.679	2.014	2.412	2.690
50		1.299	1.676	2.009	2.403	2.678
55		1.297	1.673	2.004	2.396	2.668
60		1.296	1.671	2.000	2.390	2.660
65		1.295	1.669	1.997	2.385	2.654
70		1.294	1.667	1.994	2.381	2.648
75		1.293	1.665	1.992	2.377	2.643
80		1.292	1.664	1.990	2.374	2.639
90		1.291	1.662	1.987	2.368	2.632
100		1.290	1.660	1.984	2.364	2.626
500		1.283	1.648	1.965	2.334	2.586
1000		1.282	1.646	1.962	2.330	2.581
(z) ∞		1.282*a*	1.645*b*	1.960	2.326*c*	2.576*d*

a This value has been rounded to 1.28 in the textbook.
b This value has been rounded to 1.65 in the textbook.
c This value has been rounded to 2.33 in the textbook.
d This value has been rounded to 2.58 in the textbook.



Source: Adapted from W. H. Beyer, Handbook of Tables for Probability and Statistics, 2nd ed., CRC Press, Boca Raton, Fla., 1986. Reprinted with permission.

BAB

10

UJI NORMALITAS DAN HOMOGENITAS, UJI BARLETT, ANOVA SATU JALUR

A. PENGANTAR

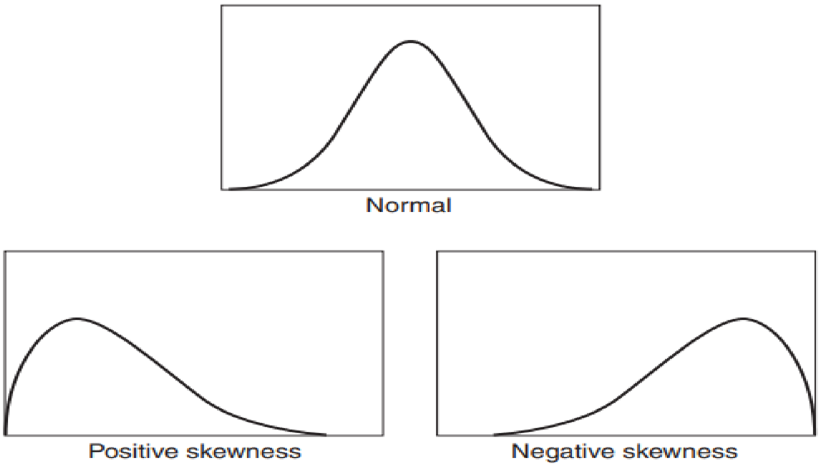
Uji normalitas dan homogenitas merupakan prosedur statistik yang digunakan untuk menilai asumsi normalitas dan homogenitas varians pada suatu kumpulan data. Kedua asumsi tersebut sangat penting pada penggunaan metode statistik parametrik. Sebelum melakukan analisis parametrik, sangat penting untuk mengeksplorasi data dan memvalidasi bahwa asumsi-asumsi tertentu telah terpenuhi. Uji parametrik mencakup beberapa alat analisis yang paling umum digunakan untuk membandingkan kelompok data dengan variabel kontinu, seperti t-*Student* dan Analisis Varians (ANOVA). Asumsi kedua uji tersebut harus dipenuhi agar hasil analisis menjadi lebih akurat dan valid.

Pengujian normalitas bertujuan untuk mengetahui apakah sebaran suatu sampel atau data terpilih bersumber dari sebaran data populasi yang normal atau tidak normal. Ada beberapa teknik analisis statistik memerlukan asumsi menggunakan distribusi normal. Persyaratan lain yang sering digunakan adalah melakukan uji homogenitas seperti uji F dan uji Bartlett. Uji ini penting dilakukan untuk membandingkan pada uji hipotesis tentang perbedaan rata-rata pada dua atau lebih kelompok data. Salah satu contohnya adalah perhitungan Anova yang harus memenuhi uji normalitas dan homogenitas agar hasilnya lebih valid.

B. UJI NORMALITAS

Pada statistik, distribusi normal menjadi salah satu syarat penting dalam melakukan analisis data. Distribusi ini sering diasumsikan dalam prosedur statistik inferensial, yaitu metode untuk menarik serta mengukur andal dan validnya suatu kesimpulan berdasarkan hasil informasi yang didapatkan dari sampel populasi. Statistik inferensial mencakup metode estimasi dan pengujian hipotesis yang semuanya didasarkan pada teori probabilitas (Isotalo, 2010). Seluruh kerangka kerja statistik didasarkan pada asumsi normalitas. Jika asumsi ini tidak terpenuhi atau dilanggar maka hasil kesimpulannya tidak valid terutama pada ukuran sampel kecil.

Uji statistik sebagian besar dikembangkan berdasarkan asumsi bahwa variabel hasil atau model residual berdistribusi normal. Normalitas dikenal juga sebagai distribusi Gaussian atau kurva yang berbentuk lonceng simetris dengan nilai μ= 0 serta σ= 1. Gambar 10.1 memperlihatkan bentuk kurva pada data yang memiliki distribusi normal dan bentuk kurva pada data yang memiliki distribusi tidak normal. Data yang tidak berdistribusi normal akan terlihat dari tingkat kemiringannya *(skewness*). Jika data lebih miring ke arah kiri maka disebut *positive skewness*. Sedangkan, data lebih miring ke kanan disebut *negative skewness*.



Gambar 10.1 Distribusi Normal dan Distribusi *Skewness*. Tabachnick and Fidell, 2013

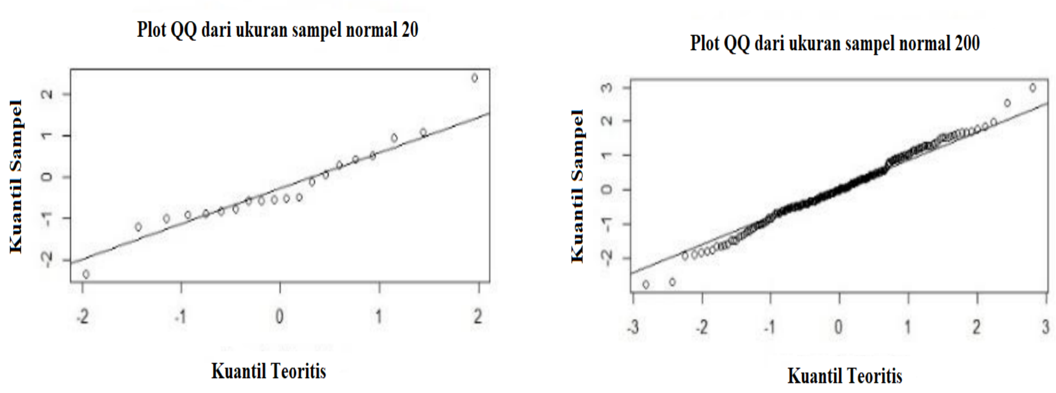
Sangat penting diketahui bahwa data pada setiap kelompok yang dibandingkan menampilkan fitur normalitas karena uji parametrik membandingkan rata-rata antar kelompok data. Hasil rata-rata tersebut harus secara akurat mewakili representasi data. Terdapat dua metode utama untuk mengevaluasi atau menilai normalitas data, yaitu metode grafik dan metode statistika atau tes numerik (Yap and Sim, 2011; Hays, 1983). Metode grafik dapat dilakukan dengan beberapa teknik, diantaranya adalah dengan melihat grafik histogram dan plot Q-Q. Sedangkan, uji normalitas menggunakan metode statistika atau tes numerik yang sering banyak digunakan atau diterapkan, yaitu uji normalitas *Kolmogorov-Smirnov* dan uji normalitas *Shapiro-Wilk*.

a) Metode Grafik

Salah satu metode untuk mengevaluasi atau menguji normalitas adalah dengan menganalisis grafik. Cara yang ampuh dan efektif untuk mengeksplorasi data adalah dengan cara memplotnya secara grafis untuk memeriksa karakteristik data secara visual. Grafik yang sering digunakan untuk menguji normalitas, yaitu grafik histogram dan Plot Q-Q. Meskipun metode grafik berguna dalam menguji normalitas pada data sampel, metode ini tidak dapat memberikan bukti formal bahwa asumsi normal berlaku. Metode ini bersifat subjektif karena tergantung dari sudut pandang orang lain, pengalaman, dan pengetahuan mengenai statistik sangat diperlukan dalam menginterpretasikan grafik dengan benar (Yap and Sim, 2011; Huang *et al.*, 2019).

1) *Plot Q-Q*

Plot Q-Q adalah alat visualisasi yang banyak digunakan dan efektif untuk menilai distribusi probabilitas empiris dari variabel acak terhadap distribusi teoritis yang dihipotesiskan. Plot Q-Q membandingkan dua distribusi probabilitas dengan memplot kuantil teoretis (sumbu horizontal) terhadap kuantil empiris (sumbu vertikal) (Huang *et al.*, 2019). Pada plot Q-Q, kuantil sampel diplot terhadap kuantil yang diharapkan jika sampel berasal dari distribusi normal. Data yang memiliki distribusi normal, hasil yang didapatkan berupa garis diagonal yang lurus. Artinya, sebaran titik-titik data berada disekitar garis atau mengikuti garis diagonal. Sebaliknya, semakin banyak titik-titik dalam plot yang menyimpang secara signifikan dari garis diagonal lurus, semakin kecil kemungkinan kumpulan data tersebut mengikuti distribusi normal. Plot Q-Q lebih mudah diinterpretasikan dalam kasus ukuran sampel yang besar

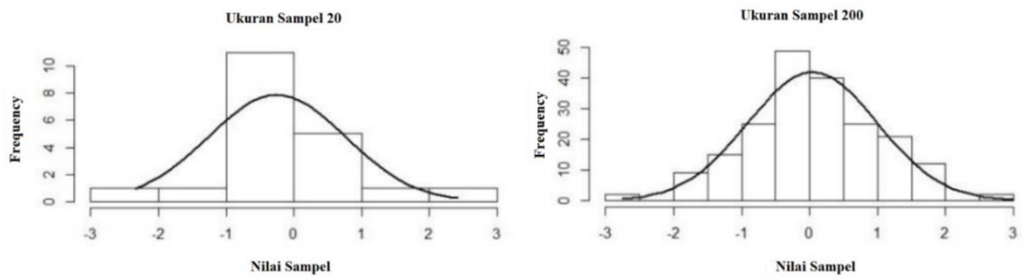


Gambar 10.2 Plot Q-Q ukuran sampel 20 dan 200. Oppong and Agbedra, 2016

Pada Gambar 10.2 diketahui bahwa plot dengan ukuran sampel 20 dan 200 memiliki distribusi normal. Namun, hasil metode ini tidak dapat memberikan bukti formal bahwa asumsi normal berlaku. Oleh karena itu, uji statistik formal diperlukan untuk memastikan kesimpulan dari metode grafis.

2) *Histogram*

Metode grafik berguna untuk memvisualisasikan distribusi normal dari variabel tertentu dengan memplot data pada grafik yang dikenal sebagai plot distribusi frekuensi atau histogram. Histogram harus mendekati bentuk lonceng (simetri) dari distribusi normal untuk memperlihatkan bahwa data terdistribusi secara normal. Membuat histogram adalah sering dilakukan untuk memeriksa secara visual distribusi data dan kemungkinan adanya pencilan sebelum melakukan uji statistik formal.



Gambar 10.3 Histogram untuk data normal ukuran 20 dan 200.
Oppong and Agbedra, 2016.

Pada Gambar 10.3 dapat diketahui histogram untuk dua data dengan ukuran berbeda, yaitu sejumlah 20 dan 200 sampel. Hasil plot memperlihatkan bahwa sampel berdistribusi normal. Akan tetapi, ditinjau dari histogram pada sampel ukuran 20 tidak mendekati lonceng dibandingkan dengan sampel ukuran 200 yang menghasilkan lebih simetris.

b) Metode Statistika/ Tes Numerik

Terdapat lebih dari 40 uji normalitas secara numerik yang diusulkan dalam literatur sejak tahun 1900 (Bayoud, 2021). Ada dua uji normalitas utama yang biasanya digunakan oleh para peneliti secara numerik, yaitu uji normalitas *Kolmogorov-Smirnov* pada ukuran sampel yang lebih besar dari 50 dan uji normalitas *Shapiro-Wilk* menggunakan data sampel berjumlah kurang dari 50 (Fein *et al.*, 2021). Kedua uji tersebut mengasumsikan bahwa distribusi adalah normal. Oleh karena itu, jika pengujian ini signifikan (yaitu nilai $p < 0.05$), hal ini memberi arti data bervariasi dari model normal dan harus dianggap tidak normal.

1) *Uji Kolmogorov-Smirnov*

Uji *Kolmogorov-Smirnov* atau uji K-S merupakan uji normalitas dengan cara melakukan perbandingan antara distribusi data diuji normalitasnya dengan distribusi Z (distribusi normal baku). Uji ini sangat efektif untuk sampel yang lebih besar dari 50. Berikut contoh dan langkah-langkah dalam melakukan perhitungan uji *Kolmogorov-Smirnov*.

Contoh soal:

Sebanyak 100 orang peserta mengikuti lomba kuis. Diberikan soal sebanyak 10 pertanyaan dan didapatkan data sebagai berikut:

Jumlah Jawaban Tepat	Frekuensi
4	10
5	15
6	18
7	20
8	17
9	12
10	8

Dengan $\alpha = 2\%$, ujilah normalitas data berdasarkan jumlah jawaban yang tepat!

Solusi soal atau kasus di atas dilakukan dengan prosedur sebagai berikut:

1. Hipotesis

   H_0: Data jumlah jawaban tepat berdistribusi normal

   H_1: Data jumlah jawaban tepat tidak berdistribusi normal

2. Statistik Uji

   N= 100

   $\bar{x}$= 6.87

   s≈1.7503

X	Frekuensi	f(X)	F(X)	Z	F(Z)	d = ∣F(X)-F(Z)∣
4	10	0.1	0.1	-1.639668787	0.050537026	0.049462974
5	15	0.15	0.25	-1.068355621	0.142680064	0.107319936
6	18	0.18	0.43	-0.497042455	0.309579557	**0.120420443**
7	20	0.2	0.63	0.074270712	0.529602509	0.100397491
8	17	0.17	0.8	0.645583878	0.740725559	0.059274441
9	12	0.12	0.92	1.216897044	0.888178305	0.031821695
10	8	0.08	1	1.78821021	0.963128950	0.03687105

Keterangan:

f(X) = hasil bagi frekuensi dengan ukuran sampel (N)

F(X)= frekuensi kumulatif dari f(X)

Z= nilai normal baku dengan persamaan sebagai berikut:

$Z = \frac{X-\bar{x}}{s}$ ....(1)

F(Z)= Nilai peluang dengan menggunakan tabel z untuk menentukan nilai F(Z)

Hasil perhitungan dapat diketahui bahwa d maksimal memiliki nilai sebesar (d_{maks})≈0.1204

3. Daerah kritis:

   Nilai kritis dengan α=2%=0,02 dan ukuran sampel N=100, merujuk pada tabel K-S, didapatkan $d_{0.02;100}$≈ 0,151.

4. Keputusan:

   $d_{maks} \leq d_{tabel}$ maka H0 diterima

   $d_{maks} > d_{tabel}$ maka H0 ditolak

   d_{maks} =0,12042 < $d_{0.02;100} \approx$ 0,151 dapat diberi kesimpulan bahwa statistik uji tidak jatuh pada daerah kritis. Dengan demikian, H_0 diterima.

5. Kesimpulan:

   Pada taraf signifikansi dengan α=2%, data untuk jumlah jawaban yang tepat memiliki data berdistribusi normal.

2) *Uji Shapiro-Wilk*

Uji normalitas seperti uji *Shapiro-Wilk* dapat digunakan untuk memverifikasi apakah asumsi normalitas terpenuhi atau dilanggar. Namun, akurasi uji bergantung pada jumlah sampel pengamatan yang representatif dalam uji coba. Uji ini sangat efektif digunakan pada sampel dengan ukuran/jumlahnya kecil, yaitu < 50 (Souza *et al.*, 2023). Berikut contoh dan langkah-langkah dalam melakukan perhitungan uji *Shapiro-Wilk*.

Contoh soal:

Data usia anak usia bawah lima tahun (dalam satuan bulan) berdasarkan kartu berobat di suatu rumah sakit yang diambil secara acak adalah sebagai berikut: 33, 23, 58, 36, 55, 33, 24, 56, 19, 58, 26, 36, 46, 40, 37, 36, 35, 30, 18, 48, 32, 27, 41, 34, Berdasarkan data tersebut dengan tingkat α = 5%, apakah data usia anak usia bawah lima tahun diambil berdasarkan dari populasi berdistribusi normal?

Langkah-langkah menyelesaikan contoh kasus di atas adalah:

1. Hipotesis

   H_0: Data usia anak bawah lima tahun berdistribusi normal

   H_1: Data usia anak bawah lima tahun tidak berdistribusi normal

2. Statistik Uji

- Hitung nilai D menggunakan persamaan (2)

No.	X_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	No.	X_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
1	18	-18,7083	350,0005	14	36	-0,7083	0,5017
2	19	-17,7083	313,5839	15	36	-0,7083	0,5017
3	23	-13,7083	187,9175	16	37	0,2917	0,0851
4	24	-12,7083	161,5009	17	40	3,2917	10,8353
5	26	-10,7083	114,6677	18	41	4,2917	18,4187
6	27	-9,7083	94,2511	19	46	9,2917	86,3357
7	30	-6,7083	45,0013	20	48	11,2917	127,5025
8	32	-4,7083	22,1681	21	55	18,2917	334,5863
9	33	-3,7083	13,7515	22	56	19,2917	372,1697
10	33	-3,7083	13,7515	23	58	21,2917	453,3365
11	34	-2,7083	7,3349	24	58	21,2917	453,3365
12	35	-1,7083	2,9183	$\sum$	881		
13	36	-0,7083	0,5017	$\bar{X}$	36,7083		

$D = \sum_{i=1}^{n}(X_i - \bar{X})^2$....................................(2)

Keterangan:

X_i= data ke i

$\bar{X}$= nilai rata-rata usia anak bawah lima tahun

$$D = \sum_{i=1}^{n}(18 - 36{,}7083)^2 + (19 - 36{,}7083)^2 + (23 - 36{,}7083)^2 + \cdots + (58 - 36{,}7083)^2 + (58 - 36{,}7083)^2$$

$D = 184{,}9583$

- Hitung nilai T menggunakan persamaan (3)

No.	a_i	$\left(X_{(n-i+1)} - X_i\right)$	$a_i\left(X_{(n-i+1)} - X_i\right)$
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
		$\sum_{i=1}^{n} a_i \left(X_{(n-i+1)} - X_i\right) =$	54,69

$T_3 = \frac{1}{D}\left[\sum_{i=1}^{n} a_i \left(X_{(n-i+1)} - X_i\right)\right]$................................(3)

Keterangan:

a_i = Koefisien tes (Shapiro Wilk)

$X_{(n-i+1)}$ = data ke n – i + 1

$$T_3 = \frac{1}{184{,}9583}[54{,}69]$$

$$T_3 = 0{,}9391$$

3. Daerah kritis

   Nilai T_3 didapatkan sebesar 0,939 dan nilai tersebut berada pada α(0,10)= 0,930 dan α(0,50)=0,963. Hal ini menyatakan bahwa p berada pada nilai 0,10 dan 0,50. Gunakan tabel *percentage point of W test* untuk mendapatkan nilai tersebut.

4. Keputusan

   Jika nilai p > 5% maka diputuskan menerima H_o atau menolak H_1

   Jika nilai p < 5% maka diputuskan menolak H_o atau menerima H_1

   Nilai p > α=0,05 maka menerima Ho

   Untuk mendapatkan keputusan bisa dilakukan dengan cara lain, yaitu mencari nilai G pada persamaan (4) di bawah ini:

   $G = b_n + c_n + In\left(\frac{T_3 - d_n}{1 - T_3}\right)$...................................(4)

   Keterangan:

   G = nilai Z pada distribusi normal

   b_n , c_n, d_n = nilai konversi statistik Shapiro-Wilk pendekatan distribusi normal

$$G = b_{24} + c_{24} + In\left(\frac{T_3 - d_{24}}{1 - T_3}\right)$$
$$G = -5{,}605 + 1{,}862 + In\left(\frac{0{,}9391 - 0{,}2106}{1 - 0{,}9391}\right)$$
$$G = -1{,}2617$$

Hasil perhitungan didapatkan nilai G=−1,2617, sehingga didapatkan juga nilai p sebesar 0,1038. Nilai $p > \alpha$=0,05, maka H_0 diterima.

5. Kesimpulan:
   Pada taraf signifikansi dengan α=5%, data usia anak usia bawah lima tahun diambil dari populasi yang berdistribusi normal.

C. UJI HOMOGENITAS

Seperti halnya menguji asumsi normalitas, ada beberapa uji statistik yang digunakan untuk menguji asumsi kesamaan varians pada statistik inferensial parametrik. Pada pengujian hipotesis statistik, sering kali diasumsikan bahwa varians kelompok data sampel yang digunakan pada dasarnya sama (Fernholz, 2011). Begitu juga ketika menguji perbedaan rata-rata, kebutuhan untuk memiliki varians yang sama masih merupakan asumsi yang harus dipenuhi. Uji Homogenitas digunakan untuk memenuhi asumsi tersebut. Jika asumsi ini tidak terpenuhi, distribusi bersyarat dari statistik uji dalam hal validitas hipotesis yang diuji sering berubah secara signifikan (Lemeshko and Novikova, 2018). Uji homogenitas merupakan metode uji statistik yang bertujuan untuk menguji kesamaan varians pada dua buah distribusi atau lebih dari data yang digunakan sebagai sampel. Sebagai contoh, asumsi ini digunakan pada uji-t dua sampel dan ANOVA. Uji homogenitas dapat dilakukan dengan berbagai teknik , seperti: *uji Bartlett, uji F, Cochran's test, Lavene, Brown-Forsythe, O'Brien test, Z-variance test*, dan *Modified Levene's Test* (Jiamwattanapong and Ingadapa, 2020). Pada buku ini hanya akan dijelaskan beberapa teknik popular yang banyak digunakan untuk melakukan uji homogenitas.

a) Uji F

Uji Fisher atau uji F merupakan uji homogenitas untuk membandingkan variansi dari dua kelompok data atau sampel. Perbandingan dilakukan dengan menghitung variansi kelompok pada data/sampel pertama dengan variansi kelompok pada data/sampel kedua. Hasil perhitungan dibandingkan dengan nilai F_{tabel} sesuai dengan tingkat keyakinan dan derajat kebebasan pada kelompok satu dan kedua. Contoh soal dan langkah-langkah uji F adalah sebagai berikut:

Contoh soal:

Suatu perusahaan memproduksi dua model produk, yaitu produk X dan produk Y. Perusahaan ingin menguji pada $\alpha = 5\%$, apakah variansi kedua model produk tersebut sama. Data didapatkan secara random sebanyak 4 buah untuk produk X dan 6 buah untuk produk Y yang terdapat pada tabel berikut:

Model	
Produk X	**Produk Y**
4	5
7	1
6	3
6	5
	3
	4

Berikut tahapan untuk penyelesaian soal di atas:

1. Hipotesis
   $H_0: \sigma_1^2 = \sigma_2^2$ (Variansi produk X sama dengan variansi produk Y)
   $H_0: \sigma_1^2 \neq \sigma_2^2$ (Variansi produk X tidak sama dengan variansi produk Y)

2. Tingkat Signifikansi
   $\alpha = 5\%$

3. Statistik Uji

Produk X		**Produk Y**	
X	$\boldsymbol{X - \overline{X}}$	**Y**	$\boldsymbol{Y - \overline{Y}}$
4	-1.75	5	1.5
7	1.25	1	-2.5
6	0.25	3	-0.5
6	0.25	5	1.5
		3	-0.5
		4	0.5

$$\bar{X} = \frac{4+7+6+6}{4} = 5{,}75 \text{ dan } \bar{Y} = \frac{5+1+3+5+3+4}{6} = 3{,}5$$

- Hitung nilai variansi masing-masing produk dengan menggunakan rumus (5) dan (6):

  $S_X^2 = \frac{\sum(X_i - \bar{X})^2}{n_x - 1}$......................................(5)

  $S_Y^2 = \frac{\sum(Y_i - \bar{Y})^2}{n_Y - 1}$......................................(6)

  Keterangan:
  S_X^2, S_Y^2 = variansi X/ variansi Y
  X_i, Y_i= data $X\ atau\ Y\ ke$ i
  $\bar{X}, \bar{Y}$ = nilai rata-rata dari data X / nilai rata-rata dari data Y
  n_x, n_Y = jumlah data X/ jumlah data Y

  Variansi untuk produk X
  $$S_X^2 = \frac{(-1.75^2 + 1{,}25^2 + 1{,}5^2 + 0{,}25^2)}{4-1}$$
  $S_X^2 = 1{,}5883$
  nilai $S_X = 1{,}258306$
  Variansi untuk produk Y

  $S_Y^2 = \frac{(1.5^2 - 2{,}5^2 - 0{,}5^2 + 1{,}5^2 - 0{,}5^2 + 0{,}5^2)}{6-1}$ = 2,3 dan nilai $S_Y = 1{,}516575$

- Hitung nilai F menggunakan rumus (7):

  $F_{hitung} = \frac{Variansi\ data\ terbesar}{Variansi\ data\ terkecil}$................................(7)

  $F_{hitung} = \frac{S_Y^2}{S_X^2} = \frac{2{,}3}{1{,}5883} = 1{,}452632$

4. Daerah Kritis
   dk $_{pembilang}$= (6-1)= 5
   dk$_{penyebut}$= (4-1)= 3
   F $_{tabel}$=F$_{(0,05;5;3)}$ =9,013 (lihat pada tabel distribusi F)

5. Keputusan

   F hitung < F table, maka menerima Ho

   F hitung > Ftabel, maka menolak Ho

   Fhitung=1,452632 < F tabel=9,013

6. Kesimpulan

   Hasil F hitung < Ftabel maka H0 diterima, yaitu Variansi produk X sama dengan variansi produk Y atau homogen.

b) Uji Bartlett

Uji Bartlett adalah uji homogenitas untuk mengetahui keseragaman sejumlah k sampel dari populasi. Uji Bartlett digunakan untuk melakukan uji homogenitas pada sampel atau kelompok yang lebih dari 2. Uji Bartlett dapat digunakan apabila telah memenuhi uji normalitas. Berikut contoh soal dan langkah-langkah penyelesaian pada uji Bartlett.

Contoh soal:

Seorang profesor ingin mengetahui dengan taraf signifikansi 5%, apakah perbedaan tiga teknik belajar mahasiswa dapat menyebabkan nilai rata-rata ujian berbeda. Secara acak ditugaskan 10 orang mahasiswa untuk menggunakan setiap teknik selama satu minggu, kemudian setiap mahasiswa mengerjakan ujian dengan tingkat kesulitan yang sama. Nilai ujian dari 21 mahasiswa ditunjukkan di bawah ini:

Sampel	**Nilai Ujian**		
	Teknik A	**Teknik B**	**Teknik C**
1	73	88	76
2	89	48	64
3	82	51	86
4	43	76	72
5	80	81	68
6	73	92	88
7		56	71
8			21

Penyelesaian untuk kasus di atas adalah sebagai berikut:

1. Hipotesis

   $H_0: \sigma_A = \sigma_B = \sigma_C$ (Ketiga teknik belajar memiliki variansi yang sama)

   $H_0: \sigma_A \neq \sigma_B \neq \sigma_C$ (Ada teknik belajar tidak memiliki variansi yang sama)

2. Tingkat Signifikansi dengan $\alpha = 5\%$
3. Hitung variansi masing-masing kelompok dengan menggunakan rumus (8):

   $S_i^2 = \frac{n\sum X_i^2 - (\sum X_i)^2}{n(n-1)}$..(8)

Teknik Belajar					
A	**A²**	**B**	**B²**	**C**	**C²**
73	5329	88	7744	76	5776
89	7921	48	2304	64	4096
82	6724	51	2601	86	7396
43	1849	76	5776	72	5184
80	6400	81	6561	68	4624
73	5329	92	8464	88	7744
		56	3136	71	5041
				21	441
$\sum A_i = 440$	$\sum A_i^2 = 33552$	$\sum B_i = 492$	$\sum B_i^2 = 36586$	$\sum C_i = 546$	$\sum C_i^2 = 40302$

Variansi Teknik Belajar A

$$S_A^2 = \frac{6\ x\ 33552 - 440^2}{6(6-1)} = 257{,}0667$$

Variansi Teknik Belajar B

$$S_B^2 = \frac{7\ x\ 36586 - 492^2}{7(7-1)} = 334{,}2381$$

Variansi Teknik Belajar B

$$S_C^2 = \frac{8\ x\ 40302 - 546^2}{8(8-1)} = 433{,}9286$$

4. Susun kelompok data ke *table* berikut:

Sampel	**db=(n-1)**	**Variansi (S^2)**	**db S^2**	**Log S^2**	**db log S^2**
A	5	257.07	1285.35	2.41	12.05
B	6	334.24	2005.44	2.52	15.14
C	7	433.93	3037.51	2.64	18.46
Jumlah	18	1025.24	6328.30	7.57	45.65

5. Hitung variansi gabungan dengan menggunakan rumus (9):

$S_g^2 = \frac{\sum(db)S_g^2}{\sum(db)}$ ....(9)

$$S_g^2 = \frac{6328,30}{18} = 351,57$$

$$Log\ S_g^2 = Log(351,57) = 2,55$$

6. Hitung nilai satuan Bartlett (B) menggunakan rumus (10):

$B = (\sum db)(\log S_g^2)$ ....(10)

$$B = \left(\sum db\right)(\log S_g^2) = 18\ x\ 2,55 = 45,9$$

7. Hitung nilai Chi Square $_{hitung}$ menggunakan rumus (11):

$\chi^2_{hitung} = (In\ 10)\{B - (dbLog\ S_i^2)\}$ ....(11)

$$\chi^2_{hitung} = (In\ 10)\{45,9 - 45,65\} = 0,575$$

8. Menentukan nilai Chi Kuadrat $_{tabel}$

$\chi_t^2(0,05; k-1)$
$\chi_t^2(0,05; 3-1)$
$\chi^2_{tabel}(0,05; 2)$= 5,991

9. Keputusan

Jika $\chi^2_{hitung} \leq \chi^2_{tabel}$ maka terima H_o
Jika $\chi^2_{hitung} \geq \chi^2_{tabel}$ maka tolak H_o

$\chi^2_{hitung} = 0{,}575 \leq \chi^2_{tabel} = 5{,}991$

10. Kesimpulan
    $\chi^2_{hitung} \leq \chi^2_{tabel}$ maka terima H_o, yaitu Variansi ketiga teknik belajar sama atau homogen

D. ANOVA SATU JALUR

Anova atau *Analysis of Variance* dan biasa disebut juga dengan Anava. Anova merupakan metode statistika yang digunakan untuk mengetahui apakah ada beda rata-rata dua atau lebih pada suatu kelompok atau perlakuan. Metode ini membandingkan variasi antara rata-rata kelompok dengan variasi di dalam kelompok. Jika variasi antara rata-rata kelompok secara signifikan lebih besar daripada variasi di dalam kelompok, hal tersebut memperlihatkan adanya perbedaan signifikan antara rata-rata kelompok.

Anova terbagi menjadi dua jenis, yaitu *one way Anova* atau lebih dikenal anova satu jalur dan *two way Anova* atau dikenal dengan sebutan Anova dua jalur. Anova satu jalur digunakan untuk mengetahui perbedaan rata-rata dari beberapa kelompok, metode ini terdiri dari satu variabel bebas/variabel independen dan satu variabel terikat/*variable* dependen. Jenis data yang digunakan adalah skala rasio atau interval (Riadi, 2014). Metode ini memiliki asumsi bahwa data yang digunakan berdistribusi normal, homogen dan independen (Dahman, 2018). Berikut contoh soal dan langkah-langkah dalam penyelesaian menggunakan Anova satu jalur.

Contoh soal:

Seorang guru sedang melakukan penelitian mengenai tiga program pembelajaran untuk persiapan ujian. 30 siswa direkrut pada penelitian ini dan dibagi dalam tiga kelompok. Siswa di setiap kelompok dipilih secara acak untuk ditugaskan menggunakan salah satu dari tiga program pembelajaran tersebut selama periode tiga minggu mendatang. Setelah tiga minggu, semua siswa mengikuti tes yang sama. Hasil tes untuk setiap kelompok direkap pada tabel di bawah ini. Pada tingkat signifikansi 5%, apakah tiga program pembelajaran tersebut akan menghasilkan nilai rata-rata berbeda?

Kelompok 1	Kelompok 2	Kelompok 3
85	91	79
86	92	78
88	93	88
75	85	94
78	87	92
94	84	85
98	82	83
79	88	85
71	95	82
80	96	81

Contoh soal di atas dapat diselesaikan dengan tahapan sebagai berikut:

1. Hipotesis

   $H_0: \mu_1 = \mu_2 = \mu_3$ (tidak ada perbedaan nilai ujian antar kelompok dari tiga program persiapan ujian yang digunakan)

   $H_0: \mu_1 = \mu_2 = \mu_3$ (Ada perbedaan nilai ujian antar kelompok dari tiga program persiapan ujian yang digunakan)

2. Tingkat signifikansi

   $\alpha = 5\%$

3. Statistik Uji dengan menghitung jumlah kuadrat dan derajat bebas menggunakan persamaan pada tabel di bawah ini.

Sumber Variansi	Derajat Bebas (db)	Jumlah Kuadrat (JK)	*Mean* Kuadrat (MK)	F
Kelompok (K)	$db_k = K - 1$	$JK_k = \sum \frac{(\sum X_K)^2}{n_K} - \frac{(\sum X_T)^2}{N}$	$MK_k = \frac{JK_K}{db_K}$	$F = \frac{MK_K}{MK_d}$
Dalam (d)	$db_d = N - K$	$JK_D = JK_T - JK_K$	$MK_d = \frac{JK_d}{db_d}$	
Total (T)	$db_T = N - 1$	$JK_T = \sum X_T^2 - \frac{(\sum X_T)^2}{N}$	MK_T	

Keterangan:

F = nilai *F test*

X = nilai observasi/data

n_k = jumlah anggota pada kelompok k

K = jumlah kelompok

N = jumlah seluruh anggota

K_1	K_1^2	K_2	K_2^2	K_3	K_3^2	Total
85	7225	91	8281	79	6241	
86	7396	92	8464	78	6084	
88	7744	93	8649	88	7744	
75	5625	85	7225	94	8836	
78	6084	87	7569	92	8464	
94	8836	84	7056	85	7225	
98	9604	82	6724	83	6889	
79	6241	88	7744	85	7225	
71	5041	95	9025	82	6724	
80	6400	96	9216	81	6561	
N=10		N=10		N=10		30
$\bar{K}_1 = 83{,}4$		$\bar{K}_2 = 89{,}3$		$\bar{K}_3 = 84{,}7$		257,4
$\sum = 834$		$\sum = 893$		$\sum = 847$		2574
	$\sum = 70196$		$\sum = 79953$		$\sum = 71993$	150149

$$db_T = 30 - 1 = 29$$

$$db_K = 3 - 1 = 2$$

$$db_d = 30 - 3 = 27$$

$$JK_T = 222142 - \frac{2574^2}{30} = 1292{,}8$$

$$JK_K = \frac{834^2}{10} + \frac{893^2}{10} + \frac{847^2}{10} - \frac{2574^2}{30} = 192{,}2$$

$$JK_d = 1292{,}8 \text{ - } 192{,}2 = 1100{,}6$$

$$MK_K = \frac{192{,}2}{2} = 96{,}1$$

$$MK_d = \frac{1100{,}6}{27} = 40{,}8$$

$F = \frac{96,1}{40.8} = 2,358$

4. Daerah Kritis

   $F_{tabel}\left(\frac{db_k}{db_d};\alpha\right)$ maka $F_{tabel}\left(\frac{2}{27};0,05\right) = 3,35$

5. Keputusan

   F hitung < F tabel, maka menolak Ho atau menerima H_1

   F hitung > Ftabel, maka menerima Ho atau menolak H_1

   F hitung=2,358 < F tabel=3,35

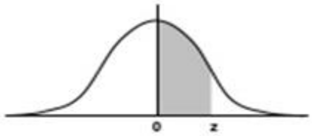
6. Kesimpulan

   Hasil F hitung < Ftabel maka H_0 ditolak, yaitu Ada perbedaan nilai ujian antara kelompok 1, kelompok 2 dan kelompok 3 dari tiga program persiapan ujian yang digunakan.

Berikut adalah *table* Z:

Distribusi Z

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



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.7	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.8	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.9	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000

ade_statistik@yahoo.com

Dibuat dengan menggunakan Software MS Excel 2003

BAB
11

PENGUJIAN RERATA (UJI T)

Dalam perjalanan ilmu pengetahuan, kita seringkali dihadapkan pada kebutuhan untuk membandingkan dan menilai - apakah perbedaan yang kita lihat dalam data hanyalah kebetulan, ataukah mencerminkan perbedaan yang signifikan dan berarti? Salah satu alat yang paling umum dan kuat dalam perbendaharaan statistika untuk menjawab pertanyaan ini adalah "Uji T" atau "Penguji Rerata". Melalui uji ini, peneliti dapat memahami dan menguji perbedaan antar kelompok, antar kondisi, atau bahkan perubahan seiring waktu. Bab "Pengujian Rerata (Uji T)" akan memaparkan Anda mengenai prinsip, prosedur, dan praktek dari Uji T.

A. UJI T SATU SAMPEL (*ONE SAMPLE T-TEST*)

Uji T satu sampel adalah analisis statistik yang digunakan untuk menguji apakah rata-rata dari satu sampel berbeda secara signifikan dari nilai populasi tertentu atau nilai teoretis. Uji T satu sampel biasanya digunakan untuk mengetahui apakah suatu sampel yang diperoleh dari penelitian memiliki rata-rata yang berbeda dari ekspektasi atau nilai yang sudah diketahui sebelumnya dari populasi. Uji T satu sampel memiliki syarat yang harus dipenuhi, seperti: data yang digunakan dalam uji T satu sampel merupakan data yang bertipe numerik (< 30 sampel), data sampel diambil dari populasi dengan cara acak, dan data berdistribusi normal (Al-Ziadi and Ramadhan, 2022). Data berdistibusi normal ini dapat diperiksa dengan berbagai tes, seperti *Shapiro-Wilk* atau plot Q-Q.

Langkah-Langkah Uji T Satu Sampel

- Penentuan Hipotesis

  Hipotesis satu sisi:

  $H_0 : \mu = \mu_0$

  $H_1 : \mu < \mu_0$

  atau

  $H_0 : \mu = \mu_0$

  $H_1 : \mu > \mu_0$

  Hipotesis dua sisi:

  $H_0 : \mu = \mu_0$

  $H_1 : \mu \neq \mu_0$

- Pemilihan Tingkat Signifikansi: Biasanya 0,05 namun bisa berbeda tergantung konteks penelitian dan df = n – 1
- Penghitungan Statistik Uji: Menggunakan rumus t-statistik.

  Rumus:

$$t_{hit} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

  Ket:

  t = nilai statistik uji T

  $\bar{x}$ = rata-rata dari sampel yang dianalisis

  μ_0 = rata-rata yang dihipotesiskan atau nilai teoritis yang ingin dibandingkan dengan rata-rata sampel

  s = standar deviasi dari sampel

  n = ukuran sampel atau jumlah pengamatan dalam sampel

- Penarikan Kesimpulan: Dibandingkan dengan nilai t tabel atau menggunakan p-*value*.

  Kriteria Uji:

  $\mu < \mu_0$: Jika t_{hitung} < - t_{tabel}, maka H_0 ditolak

  Jika t_{hitung} > - t_{tabel}, maka H_0 diterima

$\mu > \mu_0$: Jika $t_{hitung} > t_{tabel}$, maka H_0 ditolak

Jika $t_{hitung} < t_{tabel}$, maka H_0 diterima

$\mu \neq \mu_0$: Jika $t_{hitung} < -t_{tabel}$ atau $t_{hitung} > t_{tabel}$, maka H_0 ditolak

Jika $t_{hitung} > -t_{tabel}$ atau $t_{hitung} < t_{tabel}$, maka H_0 diterima

- Interpretasi Hasil
  - Memahami Nilai t-statistik: ada atau tidak ada perbedaan signifikan antara rata-rata sampel dan nilai populasi.
  - Memahami p-*value*: Jika p-*value* kurang dari tingkat signifikansi (biasanya 0,05), tolak H_0.

Contoh Kasus:

Komisi Pemilihan Umum (KPU) melakukan perhitungan partisipasi politik masyarakat dalam Pemilukada serentak tahun 2023 yang dilakukan di 12 kota di Indonesia. Berikut pada Tabel 11.1 adalah data rekapitulasi partisipasi politik (dalam %):

Tabel 11.1. Persentase partisipasi politik pada pemilukada 2023

Kota	**% Partisipasi Politik**
A	84
B	88
C	65
D	77
E	71
F	50
G	91
H	82
I	74
J	78
K	80
L	65

Dengan taraf signifikan 0.05, ujilah hipotesis bahwa rata-rata partisipasi politik dalam pemilukada di Indonesia lebih dari 75%?

Diketahui:

n = 12

$\mu_0 = 75$

$\alpha = 5\% = 0.05$

Penyelesaian:

$\bar{x}$ = ? $\Rightarrow$ $\bar{x} = \frac{84+88+\cdots+80+65}{12} = \frac{905}{12} = 75.42$

s = ? $\Rightarrow$ $s^2 = \frac{\sum x_1 - \bar{x}_1}{n-1} = \frac{(84-75.42)^2 + (88-75.42)^2 + \cdots + (65-75.42)^2}{12-1} = 130.27$

$\Rightarrow$ $s = \sqrt{s^2} = \sqrt{130.27} = 11.41$

1. Hipotesis satu sisi
   H_0 : μ = 75 : Rata-rata partisipasi politik dalam pilkada serentak tidak lebih dari 75%
   H_1 : μ > 75 : Rata-rata partisipasi politik dalam pilkada serentak lebih dari 75%

2. Tentukan t_{tabel} dengan taraf signifikansi α dan df = n -1
   $\alpha = 5\% = 0.05$
   df = n – 1 = 12 – 1 = 11 $\Rightarrow$ Lihat Tabel t $\Rightarrow$ $t_{tabel} = 1.796$

3. Statistik Uji

$$t_{hitung} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{75.42 - 75}{11.41/\sqrt{12}} = \frac{0.42}{11.41/3.46} = \frac{0.42}{3.297} = 0.127$$

4. Kriteria Uji $\Rightarrow$ $\mu > \mu_0$
   Jika $t_{hitung} > t_{tabel}$, maka H_0 ditolak
   Jika $t_{hitung} < t_{tabel}$, maka H_0 diterima

5. Kesimpulan
   Karena 0.127 < 1.796 $\Rightarrow$ H_0 diterima
   " H_0 " Rata-rata partisipasi politik dalam pilkada serentak tidak lebih dari 75%

B. UJI T DUA SAMPEL BEBAS (*INDEPENDENT SAMPLE T-TEST*)

Uji T dua sampel bebas (sering disebut *independent t-test*) adalah uji statistik yang digunakan untuk membandingkan rata-rata dari dua kelompok independen untuk mengetahui apakah ada perbedaan yang signifikan di antara kelompok tersebut. Uji T dua sampel bebas digunakan untuk menentukan apakah perbedaan rata-rata antara dua kelompok berbeda secara statistik. Uji T dua sampel bebas memiliki syarat yang harus dipenuhi, seperti: sampel yang digunakan dari kedua kelompok harus independen dan diambil secara acak (< 30 sampel), data bertipe numerik, serta data dalam kedua kelompok harus berdistribusi normal (Kim, 2019). Variansi (simpangan baku) dari kedua kelompok ada dua tipe, homogen (sama) atau tidak homogen (berbeda). Contoh kasus yang diberikan dalam sub bab ini adalah dari tipe dengan variansi yang sama.

Langkah-Langkah Uji T Dua Sampel

- Penentuan Hipotesis

  Hipotesis satu sisi

  $H_0 : \mu_1 = \mu_2$

  $H_0 : \mu_1 < \mu_2$

  atau

  $H_0 : \mu_1 = \mu_2$

  $H_0 : \mu_1 > \mu_2$

  Hipotesis satu sisi

  $H_0 : \mu_1 = \mu_2$

  $H_0 : \mu_1 \neq \mu_2$
- Pemilihan Tingkat Signifikansi: Biasanya 0,05 namun bisa berbeda tergantung konteks penelitian dan df = $n_1 + n_2 - 2$
- Penghitungan Statistik Uji: Menggunakan rumus t-statistik untuk dua sampel independen.

  Rumus:

$$t_{hit} = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{S_p\sqrt{\left(\frac{1}{n_1}\right) + \left(\frac{1}{n_2}\right)}}$$

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

Ket:

t_{hit} = nilai statistik uji T

$\bar{x}_1$ = rata-rata dari sampel kelompok pertama

$\bar{x}_2$ = rata-rata dari sampel kelompok kedua

d_0 = asumsi selisih dari μ_1 - μ_2 yang dihipotesiskan jika tidak disebutkan maka diasumsikan $d_0 = 0$

S_p = simpangan baku gabungan dari kelompok 1 dan kelompok 2

S_1 = standar deviasi kelompok 1

S_2 = standar deviasi kelompok 2

n_1 = ukuran sampel pertama, atau jumlah pengamatan dalam sampel kelompok pertama

n_2 = ukuran sampel pertama, atau jumlah pengamatan dalam sampel kelompok kedua

- Penarikan Kesimpulan: Dibandingkan dengan nilai t tabel atau menggunakan p-*value*.

  Kriteria Uji:

  $\mu_1 < \mu_2$: Jika $t_{hitung} < -t_{tabel}$, maka H_0 ditolak

  Jika $t_{hitung} > -t_{tabel}$, maka H_0 diterima

  $\mu_1 > \mu_2$: Jika $t_{hitung} > t_{tabel}$, maka H_0 ditolak

  Jika $t_{hitung} < t_{tabel}$, maka H_0 diterima

  $\mu_1 \neq \mu_2$: Jika $t_{hitung} < -t_{tabel}$ atau $t_{hitung} > t_{tabel}$, maka H_0 ditolak

  Jika $t_{hitung} > -t_{tabel}$ atau $t_{hitung} < t_{tabel}$, maka H_0 diterima

- Interpretasi Hasil
  - Memahami nilai t-statistik: ada atau tidak ada perbedaan signifikan antara rata-rata kedua kelompok.
  - Memahami p-*value*: Jika p-*value* kurang dari tingkat signifikansi, tolak H_0.

Contoh Kasus:

Suatu penelitian dilakukan untuk menguji apakah nilai mata kuliah ekologi perairan yang diberikan pada mahasiswa Prodi Manajemen Sumberdaya Perairan dan Teknologi Hasil Perairan memiliki perbedaan atau tidak. Untuk itu, diambil sampel sebanyak 10 nilai mahasiswa Prodi Manajemen Sumberdaya Perairan dan 10 nilai mahasiswa Teknologi Hasil Perairan. Berikut pada Tabel 11.2 disajikan datanya:

Tabel 11.2. Nilai mata kuliah ekologi perairan pada dua program studi

No.	Prodi Manajemen Sumberdaya Perairan	Prodi Teknologi Hasil Perairan
1	63	69
2	78	56
3	71	67
4	82	72
5	93	59
6	72	71
7	61	55
8	63	88
9	56	79
10	82	49

Dengan taraf signifikan 5%, lakukan pengujian apakah terdapat perbedaan nilai mata kuliah ekologi perairan antara mahasiswa Prodi Manajemen Sumberdaya Perairan dan Teknologi Hasil Perairan? (dengan asumsi variansi sama)

Diketahui:

Misal: μ_1 = Prodi Manajemen Sumberdaya Perairan → $\bar{x}_1$

μ_2 = Prodi Teknologi Hasil Perairan → $\bar{x}_2$

α = 5% = 0.05 $\quad d_0 = 0$

Penyelesaian:

$$\bar{x}_1 = \frac{63 + 78 + \ldots + 82}{10} = \frac{721}{10} = 72.1$$

$$\bar{x}_2 = \frac{69 + 56 + \ldots + 49}{10} = \frac{665}{10} = 66.5$$

$$S_1^2 = \frac{\sum (x_1 - \bar{x}_1)^2}{n-1}$$

$$= \frac{(63-72.1)^2 + (78-72.1)^2 + \cdots + (82-72.1)^2}{10-1} = 135.211 \Longrightarrow s_1 = \sqrt{s_1^2} = \sqrt{135.211} = 11.63$$

$$S_2^2 = \frac{\sum (x_1 - \bar{x}_2)^2}{n-1}$$

$$= \frac{(69-66.5)^2 + (56-66.5)^2 + \cdots + (49-66.5)^2}{10-1} = 142.278 \Rightarrow s_2 = \sqrt{s_2^2} = \sqrt{142.278} = 11.93$$

1. Hipotesis dua sisi

   H0 : $\mu_1 = \mu_2$: Tidak ada perbedaan rata-rata nilai mahasiswa Prodi Manajemen Sumberdaya Perairan dan Teknologi Hasil Perairan

   H0 : $\mu_1 \neq \mu_2$: Ada perbedaan rata-rata nilai mahasiswa Prodi Manajemen Sumberdaya Perairan dan Teknologi Hasil Perairan

2. Tentukan t_{tabel} dengan taraf signifikansi α dan df = $n_1 + n_2 - 2$

   $$\left.\begin{array}{l}\alpha = 5\% = 0.05 / 2 = 0.025 \\ df = n_1 + n_2 - 2 = 10 + 10 - 2 = 18\end{array}\right\} \Rightarrow \text{Lihat Tabel t} \Rightarrow t_{tabel} = 2.101$$

3. Statistik Uji

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

$$= \frac{(10-1)(135.211) + (10-1)(142.278)}{10 + 10 - 2}$$

$$= \frac{(1216.899) + (1280.502)}{18}$$

$$= \frac{2497.401}{18}$$

$$= 138.745$$

$$S_p = \sqrt{S_p^2} = \sqrt{138.745} = 11.779$$

$$t_{hitung} = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{S_p\sqrt{\left(\frac{1}{n_1}\right) + \left(\frac{1}{n_2}\right)}}$$

$$= \frac{(72.1 - 66.5) - 0}{(11.779)\sqrt{\left(\frac{1}{10}\right) + \left(\frac{1}{10}\right)}}$$

$$= \frac{(72.1 - 66.5) - 0}{(11.779)\sqrt{\left(\frac{1}{10}\right) + \left(\frac{1}{10}\right)}}$$

$$= \frac{(72.1 - 66.5) - 0}{(11.779)\sqrt{\left(\frac{1}{10}\right) + \left(\frac{1}{10}\right)}}$$

$$= \frac{5.6}{(11.779)\sqrt{\left(\frac{2}{10}\right)}} = \frac{5.6}{(11.779)\sqrt{0.2}}$$

$$= \frac{5.6}{(11.779)(0.477)} = \frac{5.6}{5.265} = 1.064$$

4. Kriteria Uji $\Rightarrow$ $\mu_1 \neq \mu_2$
   Jika $t_{hitung} > t_{tabel}$, maka H_0 ditolak
   Jika $t_{hitung} < t_{tabel}$, maka H_0 diterima

5. Kesimpulan
   Karena 1.064 < 2.101 $\Rightarrow$ H_0 diterima
   "H_0" Tidak ada perbedaan rata-rata nilai mahasiswa Prodi Manajemen Sumberdaya Perairan dan Teknologi Hasil Perairan.

C. UJI T DUA SAMPEL BERPASANGAN (*PAIRED SAMPLE T-TEST*)

Uji T dua sampel berpasangan digunakan untuk membandingkan rata-rata dari dua pengukuran atau pengamatan yang berhubungan atau berpasangan, seperti pengukuran berulang, pada subjek yang sama. Uji T dua sampel berpasangan bertujuan untuk menilai apakah ada perbedaan signifikan antara dua waktu pengukuran pada kelompok yang sama, seperti sebelum dan sesudah intervensi atau membandingkan dua metode pengukuran pada subjek yang sama. Asumsi Uji T dua sampel berpasangan adalah sebagai berikut: Distribusi data berpasangan harus normal, data yang digunakan harus bertipe numerik (< 30 sampel), dan pengamatan harus berpasangan dan berhubungan (Nuryadi *et al.*, 2017). Uji ini banyak digunakan dalam penelitian eksperimen.

Langkah-Langkah Uji T Berpasangan

- Penentuan Hipotesis
  Hipotesis satu sisi:
  $H_0 : \mu_1 = \mu_2 : \mu_d = 0$
  $H_1 : \mu_1 < \mu_2 : \mu_d < 0$

atau

$H_0 : \mu_1 = \mu_2 : \mu_d = 0$

$H_1 : \mu_1 > \mu_2 : \mu_d > 0$

Hipotesis dua sisi:

$H_0 : \mu_1 = \mu_2 : \mu_d = 0$

$H_1 : \mu_1 \neq \mu_2 : \mu_d \neq 0$

- Pemilihan Tingkat Signifikansi: Biasanya 0,05, tetapi bisa berbeda tergantung konteks penelitian, dan dengan df = n – 1
- Penghitungan Statistik Uji: Menggunakan rumus t-statistik untuk sampel berpasangan.

Rumus:

$$t = \frac{\bar{d} - d_0}{s_d / \sqrt{n}} \qquad \bar{d} = \frac{\sum d_i}{n} \qquad S_d^2 = \frac{n \sum d_1^2 - (\sum d_1)^2}{n\,(n-1)}$$

Ket.:

t = nilai t hitung

d_i = selisih dua perlakuan atau dua variabel

d_0 = asumsi selisih dari μ_1 - μ_2 yang dihipotesiskan, jika tidak disebutkan maka diasumsikan $d_0 = 0$

S_d = simpangan baku dari selisih perlakuan

- Penarikan Kesimpulan: Dibandingkan dengan nilai t tabel atau menggunakan p-*value*.

Kriteria Uji:

$\mu_1 < \mu_2$: Jika t_{hitung} < - t_{tabel}, maka H_0 ditolak
Jika t_{hitung} > - t_{tabel}, maka H_0 diterima

$\mu_1 > \mu_2$: Jika $t_{hitung} > t_{tabel}$, maka H_0 ditolak
Jika $t_{hitung} < t_{tabel}$, maka H_0 diterima

$\mu_1 \neq \mu_2$: Jika t_{hitung} < - t_{tabel} atau $t_{hitung} > t_{tabel}$, maka H_0 ditolak
Jika t_{hitung} > - t_{tabel} atau $t_{hitung} < t_{tabel}$, maka H_0 diterima

- Interpretasi Hasil
  - Memahami Nilai t-statistik: ada atau tidak ada perbedaan signifikan antara dua pengukuran.
  - Memahami p-*value*: Jika p-*value* lebih kecil dari tingkat signifikansi, tolak H_0.

Contoh kasus:

Seorang guru SMA di Kota Tarakan akan memberikan kursus pematangan materi kepada 10 siswa yang akan mengikuti perlombaan cerdas cermat mengenai pelajaran biologi. Untuk mengetahui apakah pemberian kursus tersebut efektif atau tidak, maka dicatat nilai pengetahuan mereka sebelum pemberian kursus maupun setelah kursus. Berikut adalah nilai yang diperoleh: (angka dengan kisaran 0 hingga 100)

Tabel 11.3. Nilai siswa sebelum dan setelah kursus pematangan materi

Siswa	Nilai Sebelum Kursus	Nilai Setelah Kursus
1	70	68
2	72	75
3	74	72
4	76	68
5	70	74
6	74	79
7	73	73
8	71	71
9	72	80
10	70	68

Dengan tingkat kepercayaan 95%, apakah dapat disimpulkan bahwa pemberian kursus efektif untuk meningkatkan pengetahuan siswa tentang pelajaran biologi?

Diketahui:

Misal: μ_1 = Setelah kursus

μ_2 = Sebelum kursus

α = 5% = 0.05 $d_0 = 0$

Penyelesaian:

1. Hipotesis satu sisi

H0 : $\mu_1 = \mu_2 : \mu_d = 0$: Pemberian kursus tidak efektif meningkatkan pengetahuan siswa tentang pelajaran biologi

H0 : $\mu_1 > \mu_2 : \mu_d > 0$: Pemberian kursus efektif meningkatkan pengetahuan siswa tentang pelajaran biologi

2. Tentukan t_{tabel} dengan taraf signifikansi α dan df = $n-1$

$$\left.\begin{array}{l}\alpha = 5\% = 0.05 \\ df = n-1 = 10-1 = 9\end{array}\right\} \Rightarrow \text{Lihat Tabel t} \Rightarrow t_{tabel} = 1.833$$

3. Statistik Uji

No.	**Setelah (x_1)**	**Sebelum (x_2)**	d_i	d_i^2
1	68	70	-2	4
2	75	72	3	9
3	72	74	-2	4
4	68	76	-8	64
5	74	70	4	16
6	79	74	5	25
7	73	73	0	0
8	71	71	0	0
9	80	72	8	64
10	68	70	-2	4
			$\sum d_i = 6$	$\sum d_i^2 = 190$

$$\bar{d} = \frac{\sum d_i}{n} = \frac{6}{10} = 0.6$$

$$S_d^2 = \frac{n \sum d_1^2 - (\sum d_1)^2}{n\ (n-1)}$$

$$= \frac{(10)(190) - (6)^2}{10\ (10-1)} = \frac{1900 - 36}{90} = \frac{1864}{90} = 20.71$$

$$S_d = \sqrt{S_d^2} = \sqrt{20.71} = 4.55$$

$$t_{hitung} = \frac{\bar{d} - d_0}{s_d/\sqrt{n}}$$

$$= \frac{0.6 - 0}{4.55/\sqrt{10}} = \frac{0.6}{4.55/3.162} = \frac{0.6}{1.439} = 0.417$$

4. Kriteria Uji $\Rightarrow$ $\mu_1 > \mu_2$
   Jika $t_{hitung} > t_{tabel}$, maka H_0 ditolak
   Jika $t_{hitung} < t_{tabel}$, maka H_0 diterima

5. Kesimpulan
   Karena 0.417 < 1.833 $\Rightarrow$ H_0 diterima
   "H_0" Pemberian kursus tidak efektif meningkatkan pengetahuan siswa tentang pelajaran biologi

BAB
12

PENGUJIAN PROPORSI

A. PENGANTAR

Proporsi dalam istilah statistika dapat diartikan sebagai presentase suatu kejadian dari seluruh kejadian yang ada. Perkiraan atau pengujian hipotesis terkait proporsi seringkali sangat diperlukan pada berbagai bidang. Misalnya pada bidang ekonomi, industri, kesehatan, atau politik. Pada bidang ekonomi dapat digunakan untuk mengetahui proporsi laba perusahaan yang dapat diperoleh dalam satu tahun atau proporsi produk yang telah dipromosikan dapat laku di pasaran. Pada bidang manufaktur dan industri dapat digunakan untuk mengetahui proporsi suatu produk cacat yang dihasilkan selama proses produksi. Pada bidang kesehatan dapat dilakukan perkiraan mengenai proporsi kesembuhan bagi pasien penyakit kangker setelah dilakukan beberapa metode perawatan. Pada bidang politik dapat digunakan untuk mengetahui proporsi masyarakat yang puas terhadap pemerintahan presiden pada periode tertentu.

B. PENGUJIAN PROPORSI SATU SAMPEL

Pengujian hipotesis pada bab ini adalah proporsi keberhasilan suatu kejadian dalam suatu percobaan binomial yang dinyatakan dalam suatu nilai tertentu (p_0). Artinya, kita akan menguji hipotesis dengan hipotesis nolnya adalah $H_0: p = p_0$ dengan p merupakan parameter dari distribusi binomial. Sementara itu, hipotesis alternatifnya (H_1) dapat berupa hipotesis alternatif satu arah atau hipotesis alternatif dua arah.

Hipotesis Satu Arah (Arah Kiri)	Hipotesis Satu Arah (Arah Kiri)	Hipotesis Dua Arah
$p < p_0$	$p > p_0$	$p \neq p_0$

Variabel random binomial X dapat kita gunakan untuk kriteria penolakan atau penerimaan hipotesis pengujian proporsi tersebut. Hipotesis nol ditolak apabila nilai X jauh dari nilai tengah $\mu = n \cdot p_0$. Variabel random binomial X merupakan variabel diskrit sehingga tidak mungkin dapat benar-benar tepat dalam menentukan daerah kritis dengan tingkat signifikansi α yang digunakan. Dengan demikian, untuk kasus sampel berukuran kecil, keputusan penolakan hipotesis nol didasarkan pada nilai *P-value*. Nilai *P-value* dihitung menggunakan distribusi binomial kumulatif yang telah dibahas pada Bab 8.

Langkah-langkah pengujian proporsi satu sampel secara ringkas dapat diurutkan sebagai berikut.

1. Nyatakan hipotesis nolnya, $H_0: p = p_0$
2. Pilihlah salah satu hipotesis alternatifnya:
   $H_1: p < p_0$ atau $H_1: p > p_0$ atau $H_1: p \neq p_0$
3. Pilih tingkat signifikansi: α
4. Hitunglah nilai *P-value*, dengan menentukan peluang distribusi binomial kumulatif berdasarkan jumlah keberhasilan suatu kejadian dan jumlah sampel.
5. Tentukan daerah kritisnya. Daerah kritis merupakan daerah penolakan hipotesis nol dan penerimaan hipotesis alternatif.
6. Tentukan keputusan:
   - H_0 ditolak atau H_1 diterima
   - H_0 diterima atau H_1 ditolak

Perhitungan *P-value* dan penentuan daerah kritis berdasarkan pada hipotesis yang akan digunakan.

a. Hipotesis Arah Kiri

Hipotesis:

$$H_0: p = p_0$$
$$H_1: p < p_0$$

Tingkat signifikansi: α

Perhitungan nilai *P-value*:

$P\text{-}value = P(X \leq x; p = p_0)$ .... (13.1)

Nilai x merupakan jumlah keberhasilan suatu kejadian pada jumlah sampel n.

Daerah kritis: H_0 ditolak apabila $P\text{-}value < \alpha$

b. Hipotesis Arah Kanan

Hipotesis:

$$H_0: p = p_0$$
$$H_1: p > p_0$$

Tingkat signifikansi: α

Perhitungan nilai *P-value*:

$$P\text{-}value = P(X \geq x; p = p_0) \text{ .... } (13.2)$$

atau

$$P\text{-}value = 1 - P(X < x; p = p_0) \text{ .... } (13.3)$$

Nilai x merupakan jumlah keberhasilan suatu kejadian pada jumlah sampel n.

Daerah kritis: H_0 ditolak apabila $P\text{-}value < \alpha$

c. Hipotesis Dua Arah

Hipotesis:

$$H_0: p = p_0$$
$$H_1: p \neq p_0$$

Tingkat signifikansi: α

Perhitungan nilai *P-value*:

- Jika $x < n \cdot p_0$

  $P\text{-}value = 2 \cdot P(X \leq x; p = p_0)$ .... (13.4)
- Jika $x > n \cdot p_0$

  $P\text{-}value = 2 \cdot P(X \geq x; p = p_0)$ .... (13.5)

  atau

  $P\text{-}value = 2 - 2 \cdot P(X < x; p = p_0)$ .... (13.6)

  Nilai x merupakan jumlah keberhasilan suatu kejadian pada jumlah sampel n.

  Daerah kritis: H_0 ditolak apabila $P\text{-}value < \alpha$

Table 12.1 Hipotesis dan Daerah Kritis Pengujian Proporsi Satu Sampel untuk Jumlah Sampel Kecil

Hipotesis Arah Kiri	**Hipotesis Arah Kanan**	**Hipotesis Dua Arah**
$H_0: p = p_0$ $H_1: p < p_0$	$H_0: p = p_0$ $H_1: p > p_0$	$H_0: p = p_0$ $H_1: p \neq p_0$
P-value: $P(X \leq x; p = p_0)$	*P-value*: $P(X \geq x; p = p_0)$	*P-value*: $2 \cdot P(X \leq x; p = p_0)$ untuk $x < n \cdot p_0$ $2 \cdot P(X \geq x; p = p_0)$ untuk $x < n \cdot p_0$
Daerah kritis: *P-value* $< \alpha$		

Contoh 12.1

Suatu perusahaan logistik menyatakan bahwa sebanyak 65% *seller e-commerce* di Kota Jakarta Selatan telah menggunakan jasanya untuk melakukan pengiriman barang ke konsumen. Untuk membuktikan pernyataan perusahaan tersebut diambil 25 sampel secara acak dan terdapat 12 *seller* yang menggunakan jasa perusahaan logistik tersebut. Gunakanlah tingkat signifikansi 5% untuk menguji apakah pernyataan tersebut benar atau tidak.

Jawab:

Hipotesis:

$$H_0: p = 0{,}65$$
$$H_1: p \neq 0{,}65$$

Tingkat Signifikansi:

$$\alpha = 0{,}05$$

Nilai P-*value*:

$x = 12$; $n = 25$; dan $p_0 = 0{,}65$

$$n \cdot p_0 = 25 \cdot 0{,}65 = 16{,}25$$

Karena $x = 12 < n \cdot p_0 = 16{,}25$, maka

P-value $= 2 \cdot P(X \leq 12; p = 0{,}65)$

$$= 2 \cdot \sum_{x=0}^{12} b(x; 25; 0{,}65)$$

$= 2 \cdot 0{,}06044$

$= 0{,}12088$

Daerah kritis:

H_0 ditolak apabila *P-value* $< 0{,}05$

Kesimpulan:

Karena *P-value* $= 0{,}12088 > 0{,}05$, maka H_0 tidak ditolak atau H_0 diterima. Dengan demikian dapat disimpulkan bahwa pernyataan perusahaan logistik terkait terdapat sebanyak 65% *seller e-commerce* di Kota Jakarta Selatan telah menggunakan jasanya untuk melakukan pengiriman barang ke konsumen adalah **benar**.

Nilai probabilitas binomial kumulatif dapat dilakukan dengan perhitungan rumus binomial sebenarnya atau dengan tabel binomial kumulatif selama sampel n berukuran kecil. Apabila sampel n berukuran besar, maka diperlukan prosedur pendekatan lain. Pendekatan distribusi Poisson dapat digunakan ketika nilai p_0 sangat dekat dengan 0 atau 1 dengan parameter $\mu = n \cdot p_0$. Namun ketika nilai p_0 tidak terlalu dekat dengan 0 atau 1, pendekatan distribusi normal dengan parameter $\mu = n \cdot p_0$ dan $\sigma^2 = n \cdot p_0 \cdot q_0$ lebih disukai.

Apabila digunakan pendekatan distribusi normal, maka perlu dilakukan perhitungan nilai z yang merupakan nilai variabel normal standar Z. Perhitungan nilai z untuk pengujian $p = p_0$ menggunakan rumus berikut.

$$z = \frac{x - n \cdot p_0}{\sqrt{n \cdot p_0 \cdot q_0}} \qquad \text{.. (13.7)}$$

atau

$$z = \frac{\hat{p} - n \cdot p_0}{\sqrt{\frac{p_0 \cdot q_0}{n}}} \qquad \text{.. (13.8)}$$

dengan $\hat{p} = \frac{x}{n}$.

Daerah kritis nilai z dengan tingkat signifikansi α pada hipotesis satu arah adalah $z < -z_\alpha$ untuk $p < p_0$ dan $z > z_\alpha$ untuk $p > p_0$. Sementara daerah kritis pada hipotesis dua arah adalah $z < -z_{\frac{\alpha}{2}}$ atau $z > z_{\frac{\alpha}{2}}$.

Table 12.2 Hipotesis dan Daerah Kritis Pengujian
Proporsi Satu Sampel untuk Jumlah Sampel Besar

<table>
<tr><th>Hipotesis Arah Kiri</th><th>Hipotesis Arah Kanan</th><th>Hipotesis Dua Arah</th></tr>
<tr><td>$H_0: p = p_0$
$H_1: p < p_0$</td><td>$H_0: p = p_0$
$H_1: p > p_0$</td><td>$H_0: p = p_0$
$H_1: p \neq p_0$</td></tr>
<tr><td colspan="3">Nilai z :
$$z = \frac{x - n \cdot p_0}{\sqrt{n \cdot p_0 \cdot q_0}} = \frac{\hat{p} - n \cdot p_0}{\sqrt{\frac{p_0 \cdot q_0}{n}}}$$</td></tr>
<tr><td>Daerah kritis:
$z < -z_\alpha$</td><td>Daerah kritis:
$z > z_\alpha$</td><td>Daerah kritis:
$z < -z_{\frac{\alpha}{2}}$ atau $z > z_{\frac{\alpha}{2}}$</td></tr>
</table>

Contoh 12.2

Salah satu layanan pada aplikasi ojek *online* diduga mendapatkan kurang dari 75% kepuasan dari penggunanya. *Rating* 4 dan *rating* 5 menunjukkan bahwa pengguna puas akan layanan tersebut. Perusahaan aplikasi ojek *online* tersebut ingin membertimbangkan apakah akan mempertahankan atau mengganti layanan tersebut. Layanan akan digantikan dengan layanan lain apabila mendapatkan kurang dari 75% kepuasan dari pengguna.

Percobaan dilakukan dengan melihat *rating* yang diberikan pada layanan tersebut oleh 150 akun pengguna aplikasi yang diambil secara acak. Hasil percobaan menunjukkan bahwa 80% dari jumlah sampel menyatakan puas terhadap layanan pada aplikasi ojek *online* tersebut. Dari data tersebut apakah kita perlu mengganti layanan tersebut? Gunakan tingkat signifikansi 5%.

Jawab:

Hipotesis:

$$H_0: p = 0{,}75$$
$$H_1: p < 0{,}75$$

Tingkat Signifikansi:

$$\alpha = 0{,}05$$

Nilai z :

- $x = 80\% \cdot 150 = 120$
- $n = 150$
- $p_0 = 0{,}75$
- $q_0 = 1 - p_0 = 0{,}25$

$$z = \frac{x - n \cdot p_0}{\sqrt{n \cdot p_0 \cdot q_0}} = \frac{120 - (150 \cdot 0{,}75)}{\sqrt{150 \cdot 0{,}75 \cdot 0{,}25}}$$

$= 1{,}4142$

Daerah kritis:

H_0 ditolak apabila $z < -z_\alpha$. Nilai $z_{0,05} = 1{,}645$ sehingga H_0 ditolak apabila $z < -1{,}645$

Kesimpulan:

Karena $z = 1{,}4142 > -1{,}96$, maka H_0 tidak ditolak atau H_0 diterima. Dengan demikian dapat disimpulkan bahwa tingkat kepuasan pengguna terhadap salah satu layanan pada aplikasi ojek *online* tidak kurang dari 75%. Dengan demikian layanan pada aplikasi ojek *online* dapat dipertahankan.

C. PENGUJIAN SELISIH DUA PROPORSI

Pengujian selisih dua populasi digunakan ketika berhadapan dengan permasalahan yang mengharuskan kita membandingkan proporsi dari dua sampel kelompok yang berbeda. Misalkan untuk membandingkan proporsi penjualan dari dua kelompok produk yang berbeda di pasaran, seperti produk baju dan produk sepatu. Contoh lainnya apabila ingin membandingkan proporsi kemajuan pembangunan wilayah di dua provinsi di Indonesia atau membandingkan proporsi penderita diabetes antara kelompok laki-laki dan perempuan.

Secara umum, permasalahan tersebut dapat diselesaikan dengan menyusun hipotesis nol dari dua parameter binomial p_1 dan p_2 yang bernilai sama.

$$H_0: p_1 = p_2 = p \quad \text{atau} \quad H_0: p_1 - p_2 = 0$$

Sementara untuk hipotesis alternatifnya dapat berupa satu arah atau dua arah.

Hipotesis Satu Arah (Arah Kiri)	Hipotesis Satu Arah (Arah Kiri)	Hipotesis Dua Arah
$H_1: p_1 < p_2$ atau $H_1: p_1 - p_2 < 0$	$H_1: p_1 > p_2$ atau $H_1: p_1 - p_2 > 0$	$H_1: p_1 \neq p_2$ atau $H_1: p_1 - p_2 \neq 0$

Parameter p_1 dan p_2 merupakan dua proporsi populasi yang diselidiki. Untuk menentukan proporsi populasi dari beberapa sampel diperlukan proposi estimasi yang dijadikan landasan bagi kriteria pengambilan keputusan yaitu berupa variabel acak $\hat{p}_1 - \hat{p}_2$. Dimisalkan n_1 dan n_2 merupakan jumlah sampel dari dua populasi binomial independen yang diambil secara acak, serta x_1 dan x_2 merupakan jumlah keberhasilan dalam masing-masing sampel dari dua populasi. Proporsi sukses dari kedua sampel adalah $\hat{p}_1$ dan $\hat{p}_2$.

$$\hat{p}_1 = \frac{x_1}{n_1}$$

$$\hat{p}_2 = \frac{x_2}{n_2}$$

Untuk membangun interval kepercayaan p_1 dan p_2 dengan nilai n_1 dan n_2 yang cukup besar, titik estimasi variabel acak $\hat{p}_1 - \hat{p}_2$ mendekati distribusi norma dengan nilai rataan

$$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$$

dan nilai variansi

$$\sigma^2_{\hat{p}_1 - \hat{p}_2} = \frac{p_1 \cdot q_1}{n_1} + \frac{p_2 \cdot q_2}{n_2}$$

untuk $q_1 = 1 - p_1$ dan $q_2 = 1 - p_2$.

Berdasarkan nilai rataan dan variansi tersebut, dapat dibangun daerah kritis berdasarkan variabel normal standar berikut.

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 \cdot q_1}{n_1} + \frac{p_2 \cdot q_2}{n_2}}} \qquad \text{................................ (13.9)}$$

Ketika asumsi H_0 benar, maka dapat kita substitusikan $p_1 = p_2 = p$ dan $q_1 = q_2 = q$ pada persamaan 13.9 sehingga menjadi

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{p \cdot q \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad \text{................................ (13.10)}$$

Untuk menghitung nilai z pada persamaan 13.10, kita perlu menduga nilai p dan nilai q yang berada di bawah tanda akar. Dengan menggabungkan jumlah kedua sampel, diperoleh estimasi gabungan proposi p yaitu

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

Sementara nilai $\hat{q} = 1 - \hat{p}$. Dengan demikian, nilai normal standar z yang digunakan untuk menguji hipotesis nol $p_1 = p_2$ atau $p_1 - p_2 = 0$ menjadi

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p} \cdot \hat{q} \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad \text{................................ (13.11)}$$

Daerah kritis nilai z dengan tingkat signifikansi α pada hipotesis satu arah adalah $z < -z_\alpha$ untuk $p < p_0$ dan $z > z_\alpha$ untuk $p > p_0$. Sementara daerah kritis pada hipotesis dua arah adalah $z < -z_{\frac{\alpha}{2}}$ atau $z > z_{\frac{\alpha}{2}}$.

Table 12.3 Hipotesis dan Daerah Kritis Pengujian Selisih Dua Proporsi

<table>
<tr><th>Hipotesis Arah Kiri</th><th>Hipotesis Arah Kanan</th><th>Hipotesis Dua Arah</th></tr>
<tr><td>$H_0: p_1 = p_2$
$H_1: p_1 < p_2$
atau
$H_0: p_1 - p_2 = 0$
$H_1: p_1 - p_2 < 0$</td><td>$H_0: p_1 = p_2$
$H_1: p_1 > p_2$
atau
$H_0: p_1 - p_2 = 0$
$H_1: p_1 - p_2 > 0$</td><td>$H_0: p_1 = p_2$
$H_1: p_1 \neq p_2$
atau
$H_0: p_1 - p_2 = 0$
$H_1: p_1 - p_2 \neq 0$</td></tr>
<tr><td colspan="3">Nilai z :
$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p} \cdot \hat{q} \cdot \left[\left(\frac{1}{n_1}\right) + \left(\frac{1}{n_2}\right)\right]}}$$
$\hat{p}_1 = \frac{x_1}{n_1}$; $\hat{p}_2 = \frac{x_2}{n_2}$; $\hat{p} = \frac{x_1+x_2}{n_1+n_2}$; dan $\hat{q} = 1 - \hat{p}$</td></tr>
<tr><td>Daerah kritis:
$z < -z_\alpha$</td><td>Daerah kritis:
$z > z_\alpha$</td><td>Daerah kritis:
$z < -z_{\frac{\alpha}{2}}$ atau $z > z_{\frac{\alpha}{2}}$</td></tr>
</table>

Contoh 12.3

Perusahaan parfum mengeluarkan dua produk parfum yang berbeda, yaitu parfum dengan aroma bunga dan aroma vanila. Setelah dilakukan tester, ternyata 87 dari 100 orang menyukai parfum aroma bunga dan 56 dari 80 orang menyukai parfum aroma vanila. Dari hasil tersebut parfum mana yang mungkin terjual lebih banyak? Ujilah menggunakan $\alpha = 2{,}5\%$

Jawab:

Hipotesis:

$$H_0: p_1 - p_2 = 0$$
$$H_1: p_1 - p_2 > 0$$

p_1 untuk proporsi orang yang menyukai parfum dengan aroma bunga.

p_2 untuk proporsi orang yang menyukai parfum dengan aroma vanilla.

Tingkat Signifikansi:

$$\alpha = 0{,}025$$

Nilai z :

- $x_1 = 87$
- $n_1 = 100$
- $x_2 = 56$
- $n_1 = 80$

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{87}{100} = 0{,}87$$
$$\hat{p}_2 = \frac{x_2}{n_2} = \frac{56}{80} = 0{,}7$$
$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{87 + 56}{100 + 80} = \frac{143}{180} = 0{,}794$$
$$\hat{q} = 1 - \hat{p} = 1 - 0{,}794 = 0{,}206$$

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p} \cdot \hat{q} \cdot \left[\left(\frac{1}{n_1}\right) + \left(\frac{1}{n_2}\right)\right]}} = \frac{0{,}87 - 0{,}7}{\sqrt{0{,}794 \cdot 0{,}206 \cdot \left[\left(\frac{1}{100}\right) + \left(\frac{1}{80}\right)\right]}}$$
$$= 2{,}809121$$

Daerah kritis:

H_0 ditolak apabila $z > z_\alpha$. Nilai $z_{0,025} = 1{,}96$ sehingga H_0 ditolak apabila $z > 1{,}96$

Kesimpulan:

Karena $z = 2{,}809121 > 1{,}96$, maka H_0 ditolak atau H_1 diterima. Dengan demikian dapat disimpulkan bahwa proporsi orang yang menyukai parfum dengan aroma bunga lebih banyak daripada proporsi orang yang menyukai parfum aroma vanila. Artinya, parfum dengan aroma bunga mungkin terjual lebih banyak.

Contoh 12.4

Suatu *survey* hendak dilakukan di antara masyarakat kota A dan kota B untuk mengetahui pendapat mereka mengenai rasa produk makanan baru yang akan dijual di pasaran. Survei dilakukan untuk melihat apakah ada perbedaan yang nyata terkait selera masyarakat kota A dan kota B. Untuk itu diambil sampel acak dari kedua kota tersebut, ternyata 140 dari 250 masyarakat kota A dan 360 dari 600 masyarakat kota B menyatakan bahwa produk tersebut sesuai selera mereka dan akan membeli produk tersebut jika dijual di pasaran. Dari data tersebut apakah Anda setuju apabila dikatakan bahwa proporsi masyarakat kota A yang akan membeli produk lebih rendah daripada proporsi masyarakat B? Gunakan $\alpha = 5\%$

Jawab:

Hipotesis:

$$H_0: p_1 = p_2$$
$$H_1: p_1 < p_2$$

p_1 untuk proporsi masyarakat kota A yang akan membeli produk.

p_2 untuk proporsi masyarakat kota B yang akan membeli produk.

Tingkat Signifikansi:

$$\alpha = 0{,}05$$

Nilai z:

- $x_1 = 140$
- $n_1 = 250$
- $x_2 = 360$
- $n_1 = 600$

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{140}{250} = 0{,}56$$
$$\hat{p}_2 = \frac{x_2}{n_2} = \frac{360}{600} = 0{,}6$$
$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{140 + 360}{250 + 600} = \frac{500}{850} = 0{,}588$$
$$\hat{q} = 1 - \hat{p} = 1 - 0{,}588 = 0{,}412$$

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p} \cdot \hat{q} \cdot \left[\left(\frac{1}{n_1}\right) + \left(\frac{1}{n_2}\right)\right]}} = \frac{0{,}56 - 0{,}6}{\sqrt{0{,}588 \cdot 0{,}412 \cdot \left[\left(\frac{1}{250}\right) + \left(\frac{1}{600}\right)\right]}}$$
$$= -1{,}07959$$

Daerah kritis:

H_0 ditolak apabila $z < -z_\alpha$. Nilai $z_{0,05} = 1{,}646$ sehingga H_0 ditolak apabila $z > 1{,}96$

Kesimpulan:

Karena $z = -1{,}07959 > -1{,}96,$ maka H_0 tidak ditolak atau H_0 diterima. Dengan demikian dapat disimpulkan bahwa proporsi masyarakat kota A dan proporsi masyarakat B yang akan membeli produk sama besar.

BAB
13

UJI KETERKAITAN DAN *NON* PARAMETRIK LAINNYA

A. PENGANTAR

Pada bab sebelumnya kita telah membahas tentang berbagai dasar statistika dan juga beberapa teknik statistika lainnya. Pada bab ini akan dipaparkan beberapa uji keterkaitan dan *non* parametrik lain yang mungkin cukup sering digunakan dalam pengaplikasian statistika sehari-hari. Uji keterkaitan yang akan dibahas pada bab ini merupakan uji korelasi dan uji keterkaitan *non* parametrik berdasarkan perbedaan rerata.

Seperti yang kita ketahui uji korelasi digunakan untuk mengukur sejauh mana keterkaitan di antara dua variabel ataupun lebih. Uji korelasi ini banyak digunakan dalam melakukan analisis untuk membantu memahami hubungan antara variabel-variabel yang relevan dalam sebuah penelitian atau fenomena. Tapi sama halnya dengan teknik analisis statistik lainnya, sangatlah penting menginterpretasikan hasil analisis statistik dengan baik. Korelasi tidak menunjukkan sebab-akibat dalam interpretasinya, perlu digaris bawahi bahwa hubungan yang kuat pada suatu variabel tidak selalu mengartikan bahwa salah satu variabel akan selalu menyebabkan perubahan pada variabel lainnya. Pada bab ini sedikit banyak akan diperkenalkan beberapa uji korelasi lain yang cukup sering digunakan seperti Uji Korelasi Eta, Korelasi Parsial, Korelasi Darab, Korelasi *Spearman,* dan Uji Korelasi Kendall.

Selain membahas terkait uji korelasi, pada bab ini akan dibahas juga mengenai uji *non* parametrik lainnya yang cukup sering digunakan pengaplikasiannya. Uji *non* parametrik biasanya digunakan ketika syarat atau asumsi dari data yang kita teliti tidak memenuhi distribusi tertentu. Uji *non* parametrik ini sering menjadi alternatif pilihan penyelesaian statistika dikarenakan fleksibilitasnya dalam mengolah berbagai jenis data, baik data yang tidak berdistribusi normal, data ordinal ataupun data dengan ukuran sampel yang kecil. Dalam bab ini akan diperkenalkan beberapa uji *non* parametrik lainnya dalam menguji perbedaan rerata, yaitu Uji Mann- Whitney, Median Test, Uji Wilcoxon dan Uji Tanda.

B. KORELASI ETA

Eta atau dikenal dengan simbol η merupakan salah satu ukuran korelasi yang dapat mengukur seberapa besar hubungan antara dua variabel berdasarkan perbedaan antara distribusi observasi aktual dan distribusi yang diharapkan. Eta ini biasanya digunakan pada data kategori seperti ordinal, nominal, atau disebut juga data kontingensi. Korelasi eta ini cukup jarang digunakan dan belum umum digunakan. Korelasi eta biasanya digunakan apabila data antara kedua variabel berbeda, salah satu data bersifat interval dan lainnya bersifat nominal. Karena sifatnya yang spesifik menyebabkan korelasi eta ini cukup jarang digunakan. Berikut ini adalah rumus untuk menentukan besar nilai eta (η):

$$\eta = \sqrt{1 - \frac{\sum Y_T^2 - (n_1)(\overline{Y_1})^2 - (n_2)(\overline{Y_2})^2}{\sum Y_T^2 - (n_1 + n_2)(\overline{Y_T})^2}} \qquad \ldots(1)$$

Keterangan:

Y_T = Rata-rata gabungan kelompok 1 dan kelompok 2

$\Sigma\, {Y_T}^2$ = Jumlah Kuadrat kedua kelompok sampel

$\bar{Y}_1$ dan $\bar{Y}_2$ = Rata-rata tiap kelompok

n_1 dan n_2 = Jumlah sampel 1 dan sampel 2

η` = Koefisien eta

Setelah menentukan besar nilai koefisien korelasi eta, kemudian dilakukan pengujian nilai signifikansinya menggunakan uji F. Rumusan yang digunakan sebagai berikut:

$$F = \frac{\eta^2 (N-K)}{(1- \eta^2)(K-1)} \qquad \ldots(2)$$

Keterangan:

N = Jumlah sampel

K = Jumlah subkelas variabel nominal.

F = Nilai signifikansi

Hasilnya kemudian dibandingkan dengan nilai F tabel dengan df atas atau numerator K-1 dan df bawah atau denumerator N-K. Koefisien korelasi dapat ditentukan berdasarkan probabilitas batas kritis taraf keberartian atau perbandingan antara F hitung dengan F tabel. Korelasi dinyatakan berarti ketika nilai F hitung lebih besar dari F tabel atau probabilitasnya berada dibawah batas kritis taraf keberartian.

C. KORELASI PARSIAL

Korelasi parsial merupakan salah satu bagian dari uji korelasi yang digunakan untuk melihat kuat hubungan antara variabel. Yang membedakan korelasi parsial ini dengan korelasi lainnya adalah korelasi parsial umumnya digunakan apabila ingin membuktikan adanya hubungan antara masing-masing variabel bebas secara parsial atau sendiri-sendiri terhadap variabel terikatnya (Muhid, 2019). Korelasi parsial bertujuan melihat korelasi yang lebih murni dari suatu variabel bebas dari variabel bebas lainnya terhadap variabel terikatnya, karena dalam korelasi parsial terdapat variabel yang dikontrol (Mundir, 2012). Adapun rumusan untuk mencari koefisien korelasi parsial sebagai berikut:

$$r_{y12} = \frac{r_{y1} - r_{y2}.r_{12}}{\sqrt{(1-r_{y2}^2)} - \sqrt{(1-r_{12}^2)}} \qquad \ldots(3)$$

Keterangan:

r_{y12} = Korelasi antara variabel Y dengan X1 dikontrol oleh X2

r_{y1} = Korelasi antara variabel Y dengan variabel X1

r_{y2} = Korelasi antara variabel Y dengan variabel X2

r_{12} = Korelasi antara variabel X1 dengan X2

Rumus tersebut diatas mensyaratkan terlebih dahulu telah diketahui besar korelasi antara variabel, seperti korelasi antara Y terhadap X1, Y terhadap X2 dan X1 terhadap X2 atau dapat dikatakan perlu diketahui terlebih dahulu korelasi antara setiap dua variabel tanpa variabel kontrol. Oleh karena itu terlebih dahulu diperlukan perhitungan korelasi tersebut dengan rumusan sebagai berikut:

$$r_{y1} = \frac{\sum x_1 y}{\sqrt{(\sum x_1^2)(\sum y^2)}} \quad \ldots(4)$$

$$r_{y2} = \frac{\sum x_2 y}{\sqrt{(\sum x_2^2)(\sum y)}} \quad \ldots(5)$$

$$r_{12} = \frac{\sum x_1 x_2}{\sqrt{(\sum x_1^2)(\sum x_2^2)}} \quad \ldots(6)$$

Keterangan:

r_{y1} = Korelasi antara variabel Y dengan variabel X1

r_{y2} = Korelasi antara variabel Y dengan variabel X2

r_{12} = Korelasi antara variabel X1 dengan X2

x_1, x_2 = Besar nilai variabel x1, dan x2

y = Besar nilai variabel y

Selain dari besarnya koefisien korelasi parsialnya, sebagai dasar pengambilan keputusan dilakukan juga uji signifikansi pada koefisien korelasi parsial tersebut (Kadir, 2010). Dalam uji signifikansi ini menggunakan uji t dengan hipotesis:

Hipotesis Parsial 1

Ho : Tidak adanya korelasi antara variabel x1 dengan variabel y secara parsial

Ha : Terdapat korelasi antara variabel x1 dengan variabel y secara parsial

Hipotesis Parsial 2

Ho : Tidak adanya korelasi antara variabel x2 dengan variabel y secara parsial

Ha : Terdapat korelasi antara variabel x2 dengan variabel y secara parsial

Dasar pengambilan keputusan, dikatakan Ho diterima jika nilai *t hitung* < dari *t tabel* atau probabilitasnya > dari taraf keberartian. Berlaku sebaliknya Ho ditolak jika nilai *t hitung* ≥ *t tabel* atau probabilitasnya ≤ taraf keberartian. Adapun rumus yang digunakan dalam menghitung nilai t adalah sebagai berikut:

$$t = \frac{r_{y12}\sqrt{n-3}}{\sqrt{(1-r_{y12}^2)}} \qquad \ldots(7)$$

Keterangan:

r_{y12} = Korelasi antara variabel Y dengan X1 dikontrol oleh X2

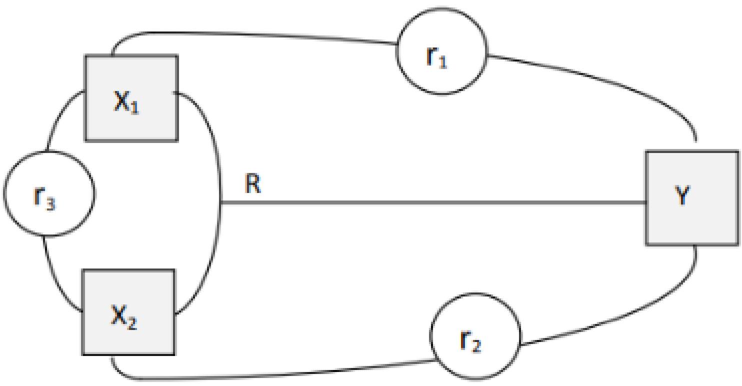
t = Besar t hitung

n = Jumlah sampel

Saat ini untuk melakukan perhitungan koefisien parsial beserta signifikansinya sudah dapat dibantu melalui banyak alat bantu statistika seperti SPSS. Sehingga dalam proses analisisnya pun akan lebih cepat dan tepat. Namun tetap saja kita tetap harus hati-hati dan teliti dalam menentukan teknik statistika yang tepat dan asumsi penggunaannya dengan benar.

D. KORELASI DARAB

Korelasi darab yang lebih sering didengar dengan istilah *multiple correlation* atau korelasi ganda dapat dikatakan sebagai kelanjutan dari korelasi parsial. Berbeda halnya dengan korelasi parsial yang ingin mengetahui kuatnya hubungan antara setiap variabel bebas secara terpisah-pisah terhadap variabel terikatnya, korelasi darab ini bertujuan untuk melihat kuat lemahnya hubungan antara seluruh variabel bebas terhadap variabel terikatnya secara simultan bersamaan. Korelasi darab ini bisa dilakukan bersamaan dengan korelasi parsial terlebih lagi jika jumlah variabel bebas adalah dua.



Gambar 13.1 Korelasi darab dengan dua variabel bebas. (Mundir, 2012)

Pada Gambar 13.1 diperlihatkan bagaimana hubungan korelasi darab yang dihasilkan dari dua variabel bebas. Dimana korelasi darab ini umumnya disimbolkan dengan huruf *R* (besar) dan korelasi parsial antar variabelnya disimbolkan dengan huruf *r* (kecil). Rumus korelasi darab dengan dua variabel bebas adalah sebagai berikut:

$$R_{y-12} = \sqrt{\frac{(r_{y1}^2 + r_{y2}^2) - 2(r_{y1})(r_{y2})(r_{12})}{1 - r_{12}^2}} \qquad \ldots(8)$$

Keterangan:

R_{y-12} = Korelasi antara variabel Y dengan variabel X1 dan X2

r_{y1} = Korelasi antara variabel Y dengan variabel X1

r_{y2} = Korelasi antara variabel Y dengan variabel X2

r_{12} = Korelasi antara variabel X1 dengan X2

Untuk mencari besar korelasi setiap 2 variabel terpisah tanpa variabel kontrol dapat dilihat pada bagian korelasi parsial. Secara matematis nilai korelasi darab ini menunjukkan kuatnya hubungan dari variabel, namun perlu di ingatkan kembali bahwa nilai korelasi ini tidak mengimplikasikan hubungan sebab akibat. Pengujian signifikasi pada korelasi darab ini dapat menggunakan hipotesis uji F.

Ho : Tidak adanya korelasi antara variabel x1, x2 dengan variabel y secara bersama-sama

Ha : Terdapat korelasi antara variabel x1, x2 dengan variabel y secara bersama-sama

Adapun rumusan menghitung besarnya nilai F dapat menggunakan rumus sebagai berikut:

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)} \qquad \ldots(9)$$

Keterangan:

R = Koefisien korelasi darab

k = jumlah variabel terikat

n = jumlah sampel

Besar nilai F hitung ini kemudian dibandingkan dengan besar nilai F tabel. Ketentuan besar F tabel diambil dengan numerator besar *k*, dan denumerator dengan besar *n-k-1*. Dikatakan Ho diterima jika F hitung ≤ F tabel atau probabilitasnya > taraf keberartian yang ditentukan, serta berlaku sebaliknya.

E. KORELASI SPEARMAN

Korelasi spearman merupakan ukuran tertua yang disusun berdasarkan peringkat (Krishnainah and Sen, 1984). Charles Edward Spearman pada tahun 1904 seorang ahli psikologi, memprakarsai ukuran korelasi spearman ini sebelum disiplin ilmu statistika itu menjadi tren (Kvam and Vidakovic, 2007). Beberapa peneliti menyatakan bahwa korelasi spearman ini cocok digunakan untuk mencari kuat hubungan antara dua variabel bivariat yang bersifat ordinal, bahkan ada yang mensyaratkan jumlah sampel harus ≥ 4 (Corder and Foreman, 2014). Korelasi spearman ini dikenal dengan simbol ρ (rho).

Bentuk data pada korelasi spearman ini dapat berupa data sampel acak yang bersifat *non* numerik berpasangan. Pada prosesnya korelasi ini menggunakan sistem peringkat, sehingga ketika terdapat data yang sama atau kembar maka pemeringkatannya akan menggunakan nilai rata-rata (Nugroho, 2008). Selain data murni berskala ordinal, data skala *non* numerik ini dapat berupa skala nominal yang dikonversi menjadi ordinal dan diperingkatkan (Stolp, Dowdy and Wearden, 1984), atau dapat berupa data interval maupun rasio yang diordinalkan (Anwar, 2009). Secara keseluruhan dapat dikatakan bahwa korelasi spearman ini dapat digunakan ketika data tidak berdistribusi normal, tidak homogen, dan berskala ordinal (Riadi, 2014). Adapun rumus yang digunakan untuk mencari besar koefisien korelasi spearman ini adalah sebagai berikut:

$$\rho = 1 - \frac{6 \sum_i^n d_i^2}{n^3 - n} \qquad \ldots(10)$$

Keterangan:

ρ = Koefisien korelasi Spearman

d = selisih rangking antara kedua variabel

n = jumlah sampel

Setelah diketahui besar koefisien korelasi, kemudian dilanjutkan menguji tingkat signifikansi dari korelasi ini. Untuk menguji signifikasi pada korelasi spearman pada sampel kecil dapat menggunakan bantuan tabel r dengan panduan besar sampel dan nilai koefisien korelasinya dan juga taraf keberartiannya. Karena tabel r terbatas pada jumlah sampel tertentu, jika sampel yang dimiliki lebih besar lagi dapat menggunakan *z-score* dan tabel distribusi normal dalam penentuannya (Corder and Foreman, 2014). Berikut ini adalah rumus mencari *z-score*:

$$z = \rho \left| \sqrt{n-1} \right| \qquad \ldots(11)$$

Keterangan:

ρ = Koefisien korelasi Spearman

n = jumlah pasangan sampel

Seiring dengan perkembangan waktu karena keterbatasan dalam ukuran sampel menggunakan *z-score,* beberapa peneliti lebih merekomendasikan untuk menguji signifikansi dari korelasi spearman ini menggunakan formula *t-student*. Hipotesis yang digunakan dalam uji signifikansi ini adalah:

Ho : Terdapat hubungan signifikan antara kedua variabel

Ha : Tidak terdapat hubungan signifikan antara kedua variabel

Rumus yang digunakan dalam mencari besar nilai t hitung, dapat menggunakan rumus berikut:

$$t = \rho \sqrt{\frac{n-2}{1-\rho^2}} \qquad \ldots(12)$$

Keterangan:

ρ = Koefisien korelasi Spearman

n = Jumlah pasangan sampel

Setelah didapatkan nilai t hitung, kemudian diputuskan Ho diterima jika nilai t hitung < t tabel atau jika probabilitasnya > taraf keberartian, berlaku pula sebaliknya. Karena korelasi spearman ini membutuhkan proses pemeringkatan terlebih dahulu makan ada baiknya untuk dapat melihat contoh proses pemeringkatan dibawah ini:

Tabel 13.1 Contoh Data Korelasi Spearman

No	X	Y
1	111	94
2	102	94
3	102	100
4	115	96
5	122	103
6	98	105
7	92	120
8	97	108
9	85	109
10	112	112

Tabel 13.1 memperlihatkan contoh data mentah yang akan dilakukan uji korelasi spearman. Langkah pertama adalah melakukan pemeringkatan dengan melakukan pengurutan data terlebih dahulu dari data terkecil ke terbesar. Perlu diingat kembali ketika terdapat data kembar atau sama, maka peringkatnya merupakan hasil rata-rata. Sebagai contoh pada Tabel 14.2 terdapat data kembar yaitu pada x dengan nilai 102 pada peringkat 5 dan 6 sehingga peringkat keduanya menjadi rata-rata dari 5 dan 6 yaitu 5,5.

Tabel 13.2 Contoh Pemeringkatan Data Korelasi Spearman

X yang diurutkan	Peringkat X	Y yang diurutkan	Peringkat Y
85	1	94	1,5
92	2	94	1,5
97	3	96	3
98	4	100	4
102	5,5	103	5
102	5,5	105	6
111	7	108	7
112	8	109	8
115	9	112	9
122	10	120	10

Setelah dilakukan pemeringkatan barulah data dapat diolah ke tahap yang selanjutnya untuk mencari nilai selisih peringkat (d), menentukan nilai koefisien korelasi, hingga melakukan uji signifikansi korelasi. Saat ini untuk menentukan hasil korelasi spearman sudah dapat menggunakan bantuan berbagai alat bantu statistika seperti SPSS.

F. KORELASI KENDALL'S TAU

Sama seperti Korelasi Spearman, Kendall's Tau juga merupakan korelasi dengan sistem peringkat yang digunakan untuk mengukur derajat kedekatan atau relasional antara variabel bebas dan variabel terikat ketika data tidak berdistribusi normal (Riadi, 2014). Umumnya korelasi kendall's tau ini digunakan sebagai alternatif pilihan analisis ketika asumsi analisis *product-moment* tidak terpenuhi, oleh karena itu korelasi kendall's tau juga dikenal sebagai korelasi *product moment sign of concordance* (Krishnainah and Sen, 1984). (Kvam and Vidakovic, 2007) menyatakan bahwa kendall's tau merupakan alternatif pengukuran dependen bivariat yang melihat seberapa banyak pasangan yang konkordan, artinya adalah seberapa banyak yang menandakan bahwa variabel bebas dan variabel terikat adalah benar-benar berpasangan.

Meskipun korelasi kendall's tau dan korelasi spearman sama-sama melakukan pemeringkatan biasanya korelasi kendall's tau lebih banyak dipilih untuk data-data ordinal dengan jumlah yang besar, sedangkan korelasi spearman untuk data-data dengan jumlah sampel yang lebih kecil. Asumsi pada korelasi ini pun sama dengan spearman, yaitu bebas distribusi, data berbentuk ordinal, dan merupakan bivariat. Korelasi Kendall's tau ini terdapat terbagi menjadi dua kategori, berdasarkan variabel yang peringkatnya sama dan peringkatnya tidak sama (Riadi, 2104). Rumus mencari koefisien korelasi kendall's tau untuk peringkat yang tidak sama sebagai berikut:

$$\tau = \frac{S_y}{S_x} = \frac{S}{\frac{1}{2}N(N-1)} \qquad \ldots(13)$$

Keterangan:

τ = Koefisien korelasi kendall

S = Jumlah total statistik untuk konkordansi dan diskonkordansi

N = Jumlah sampel

Jika peringkat yang sama, akan menggunakan rumusan sebagai berikut:

$$\tau_s = \frac{S}{\sqrt{\frac{1}{2}n\,(n-1) - T_x} - \sqrt{\frac{1}{2}n\,(n-1) - T_y}} \qquad ...(14)$$

Dimana $T_x = \frac{1}{2}\sum t\,(t-1)$ …(15)

$T_y = \frac{1}{2}\sum t\,(t-1)$ …(16)

Keterangan:

τ = Koefisien korelasi kendall

S = Jumlah total statistik untuk konkordansi dan diskonkordansi

t = banyaknya peringkat yang sama

n = Jumlah sampel

Untuk menguji signifikansi dari korelasi kendall ini menggunakan nilai z, dimana hipotesis yang diuji sebagai berikut:

Ho : Tidak Terdapat hubungan signifikan antara kedua variabel

Ha : Terdapat hubungan signifikan antara kedua variabel

Keputusan Ho diterima jika nilai z hitung < z tabel atau nilai probabilitasnya > dari taraf keberartian yang ditentukan. Begitupun sebaliknya Ho ditolak ketika z hitung ≥ z tabel atau probabilitasnya ≤ taraf keberartiannya. Adapun rumus z hitung yang digunakan untuk menguji signifikansi dari korelasi Kendal yaitu:

$$z = \frac{\tau}{\sqrt{\frac{2\,(2n+5)}{9n\,(n-1)}}} \qquad …(17)$$

Keterangan:

τ = Koefisien korelasi kendall

n = Jumlah sampel

G. MANN-WHITNEY

Uji Mann-Whitney lebih dikenal dengan sebutan *U-test*. Mann-Whitney merupakan uji *non* parametrik yang cukup kuat menjadi alternatif pengganti uji t ketika asumsi normalitas dan homogenitas tidak terpenuhi (Kadir, 2010). Uji ini bertujuan untuk mengetahui apakah rataan dua buah populasi bebas berasal dari populasi yang homogen (Riadi, 2014). Uji ini juga dapat melihat ada tidaknya perbedaan rataan antara dua populasi (Kvam and Vidakovic, 2007).

Jenis data yang dapat menggunakan uji Mann-Whitney ini dapat berupa data ordinal, interval maupun rasio. Nugroho (2008) juga memaparkan bahwa terdapat beberapa asumsi yang diperlukan dalam melakukan uji ini, seperti halnya sampel harus bersifat acak, kedua *sample* harus saling bebas, dan secara *mutual* pun bebas, data sampel sebaiknya bersifat kontinu, dan minimal data bersifat ordinal. Untuk mencari besarnya hasil uji dapat menggunakan rumus berikut:

Untuk ukuran sampel ≤ 20

$$U_1 = n_1.n_2 + \frac{n_2(n_2+1)}{2} - \sum R_2 \qquad \ldots(18)$$

atau

$$U_2 = n_1.n_2 + \frac{n_1(n_1+1)}{2} - \sum R_1 \qquad \ldots(19)$$

Untuk memastikan nilai U $U_{terkecil} = n_1.n_2 - U_{terbesar}$ …(20)

Keterangan:

U_1, U_2 = Penguji U_1 dan Penguji U_2

R_1, R_2 = Rangking sampel 1 dan rangking sampel 2

$n_1.n_2$ = Jumlah sampel 1 dan jumlah sampel 2

Untuk ukuran sampel > 20

$$Z = \frac{U - \frac{n_1.n_2}{2}}{\sqrt{\frac{n_1.n_2\ (n_1+n_2+1)}{12}}} \qquad \ldots(21)$$

Keterangan:

Z = Nilai U Test

$n_1.n_2$ = Jumlah sampel 1 dan jumlah sampel 2

Sebagai contoh, apabila seorang pengusaha jasa sablon ingin melihat efektivitas penggunaan mesin otomatis terbaru sejumlah 5, dibandingkan dengan mesin lama yang berjumlah 8 mesin. Pengusaha tersebut ingin melihat apakah mesin otomatis terbaru ini bisa menjadi alternatif pilihan atau sama saja. Tabel 14.3 merupakan data yang diterima berupa rata-rata jumlah produk yang dihasilkan dalam 1 jam pemakaian.

Tabel 13.3 Contoh Kasus Mesin Mann-Whitney

Mesin Lama	**Jumlah**	**Mesin Baru**	**Jumlah**
1	95	9	100
2	84	10	120
3	96	11	110
4	81	12	95
5	83	13	99
6	92		
7	88		
8	82		

Untuk menyelesaikan kasus diatas maka yang pertama tentukan terlebih dahulu hipotesis statistik yang akan diujikan, sebagai berikut:

Ho : $\mu 1 \geq \mu 2$

H1 : $\mu 1 < \mu 2$

Kemudian dilanjutkan dengan menentukan taraf keberartian yaitu 5%, dan menentukan besar titik kritis U dengan n1 adalah 8 dan n2 adalah 5, maka berdasarkan tabel didapatkan nilai kritis U adalah 8. Langkah selanjutnya adalah melakukan pemeringkatan seperti pada Tabel 13.4.

Tabel 13.4 Pemeringkatan Mann-Whitney

Mesin Lama	**Jumlah**	**Peringkat**	**Mesin Baru**	**Jumlah**	**Peringkat**
4	81	1	12	95	7,5
8	82	2	13	99	10
5	83	3	9	100	11
2	84	4	11	110	12
7	88	5	10	120	13
6	92	6			
1	95	7,5			
3	96	9			
R1		**37,5**	**R2**		**53,5**

Setelah didapatkan besar R1 = 37,5 dan R2 = 53,5 dari hasil pemeringkatan kemudian dilakukan perhitungan nilai U dengan besar U1 = 38,5 dan U2 = 1,5. Nilai U ditentukan berdasarkan nilai terkecil dari nilai U yang dihitung yaitu 1.5. Sehingga dapat disimpulkan bahwa nilai U lebih kecil dari U kritis yaitu 1,5 < 8, maka Ho ditolak, artinya mesin otomatis baru lebih besar kapasitas produksi dari mesin lama.

H. UJI MEDIAN

Sesuai dengan sebutannya uji ini menggunakan median sebagai patokannya. Uji median digunakan untuk mengkomparasikan beberapa populasi bebas yang datanya minimal bertipe ordinal (Anwar, 2009). Uji median ini juga diyakini dapat menjadi pelengkap dari uji Kruskal-Wallis, dimana ini menguji beberapa populasi apakah memiliki median yang sama atau tidak (Subandriyo, 2018). Secara teknis uji median ini merekam data numerik ke dalam skala nominal, dengan catatan hanya menguji apakah setiap nilai tersebut lebih besar atau lebih kecil dari median (Stolp, Dowdy and Wearden, 1984).

Prosedur pengujian signifikansi uji median ini mengikuti prosedur uji kontingensi. Nugroho (2008) memaparkan beberapa asumsi yang harus dipenuhi dalam melakukan uji median ini , diantaranya adalah setiap sampelnya berasal dari sampel acak dan saling bebas. Kemudian skala pengukuran data minimal ordinal sehingga dapat dikonversi ke dalam nominal lebih dari dan kurang dari median. Hipotesis yang digunakan pada uji median ini adalah:

Ho : Tidak Terdapat perbedaan pada populasi diamati

Ha : Terdapat perbedaan pada populasi yang diamati

Dari hipotesis tersebut kemudian dihitung besar nilai uji berdasarkan prosedur kontingensi chi-kuadrat. Ho dinyatakan diterima apabila nila chi kuadrat hitung < chi kuadrat tabel atau probabilitasnya melebihi taraf keberartian yang ditentukan. Rumus mencari nilai Chi-kuadrat hitung adalah:

$$X^2 = \sum_{i=1}^{c} \frac{(O_{1i} - O_{2i})^2}{n_i} \qquad \ldots(22)$$

Keterangan:

X^2 = nilai uji median

O_{1i} = jumlah nilai median populasi pertama yang ke i

O_{2i} = jumlah nilai median populasi kedua yang ke i

n_i = jumlah sampel i

I. UJI TANDA

Sama halnya dengan uji-uji yang telah diperkenalkan sebelumnya, uji tanda merupakan bagian uji *non* parametrik, yang digunakan jika jumlah populasinya relatif kecil kurang dari 30 atau data tidak berdistribusi normal. Uji tanda menguji hipotesis menggunakan median populasi, dalam beberapa kasus rataan biasanya digantikan oleh median sebagai parameter lokasi yang relevan (Walpole *et al.*, 2016). Sesuai dengan namanya uji tanda ini menggunakan tanda "+" dan "-" dalam pengaplikasiannya.

Uji tanda ini biasanya digunakan untuk membandingkan dua sampel yang saling berpasangan. Satu-satunya prasyarat yang harus dipenuhi dalam Uji tanda ini adalah antar pasangan memiliki hubungan yang relevan meskipun berasal dari populasi yang berbeda-beda (Riadi, 2016). Misalnya sampel acak diperoleh dari distribusi kontinu atau tidak normal kemudian akan diuji median populasinya menggunakan uji tanda. Umumnya menggunakan hipotesis sebagai berikut, dimana $\tilde{\mu}$ adalah padanan populasinya:

$H_0: \tilde{\mu} = \widetilde{\mu_0}$

$H_1: \tilde{\mu} > \widetilde{\mu_0}\ atau\ \ \tilde{\mu} < \widetilde{\mu_0}\quad atau\quad \tilde{\mu} \neq \widetilde{\mu_0}$

Dalam praktiknya kita mengganti setiap pengamatan diatas $\widetilde{\mu_0}$ dengan tanda positif dan dibawah $\widetilde{\mu_0}$ mengunakan tanda negatif. H_0 diterima atau dinyatakan benar jika jumlah tanda positif dan negatif *hampir* sama. Bagaimana jika dalam pengamatan ternyata nilai $\widetilde{\mu_0}$ persis sama, secara dasar distribusi yang bersifat kontinu probabilitasnya untuk mencapai nilai persis sama adalah nol. Namun jika hal tersebut terjadi maka setiap pengamatan yang sama dengan $\widetilde{\mu_0}$ dikeluarkan dari analisis (Kokoska, 2015).

Selain melihat berdasarkan besar nilai $\widetilde{\mu_0}$, penentuan uji tanda ini dapat menggunakan peubah acak binomial X, yang menyatakan banyaknya tanda positif dalam sampel acak (Walpole *et al.*, 2016). H_0dikatakan diterima apabila apabila nilai sampel menghasilkan tanda postif dan negatife sama dengan 0,5 (P = *0,5*). H_1diterima jika nilai sampel berada lebih kecil atau lebih besar dari peubah acah binomial yang ditentukan diawal, dan juga dibandingkan dengan taraf keberartiannya. Untuk mempersingkat dan memperjelas penggunaannya pada uji tanda ini menggunakan rumusan sebagai berikut (Riadi, 2016):

Jika Sampel Kecil (≤ 25)

$p\ (X_A > X_B) = p\ (X_A < X_B) = 0{,}5$ ....*(23)*

Keterangan:

Jika arah p $(X_A > X_B)$ diberi tanda +

Jika arah p $(X_A < X_B)$ diberi tanda –

Jika arah p $(X_A > X_B)$ sama dengan p $(X_A < X_B)$ dibuang

Jika Sampel Besar (> 25)

$Z = \frac{(x \pm 0{,}5) - 0{,}5\ N}{0{,}5\ \sqrt{N}}$ ...*(24)*

Keterangan:

X + 0,5 digunakan bila x < 0,5 N

X – 0,5 digunakan bila x > 0,5 N

N = banyaknya pasangan yang berbeda (tidak sama)

X = banyaknya tanda (+ atau -) yang paling sedikit

Untuk dapat lebih memahami ada baiknya kita gunakan studi kasus. Seorang penjahit memiliki mesin jahit *portable* yang beroperasi dengan tenaga baterai. Selama proses pemakaian diukur lama pemakaian mesin jahit hingga baterainya diisi kembali dalam satuan jam. Data yang didapatkan adalah 1.5; 2,2; 1,3; 0,9: 1,7; 1,3: 1,8; 1,6; 2.0; 1,9; 1,5. Lakukanlah uji tanda pada taraf keberartian 0,05 untuk melihat kinerja mesin dengan median 1,8 jam.

1. Diketahui
   $H_0: \tilde{\mu} = 1{,}8$
   $H_1: \tilde{\mu} \neq 1{,}8$
   $\alpha = 0{,}05$
2. Karena data kurang jumlahnya kecil maka menggunakan rumusan peubah binomial X dengan p = 0,5.
3. Langkah selanjutnya adalah mengganti data dengan simbol positif negatif. Karena terdapat data persis sama dengan median yang dicari yaitu 1,8 maka data tersebut dibuang, sehingga didapatkan:

-,+, -, -, -, -, -, +, +, -

Dimana x adalah banyaknya tanda tersedikit yaitu positif yang jumlahnya adalah tiga. Sehingga didapatkan *x = 3* dan *N = 10.*

4. Kemudian dilakukan perhitungan nilai P menggunakan bantuan tabel binomial dengan N = 10 dan x = 3 diperoleh nilai p = 0,3438.
5. Nilai P tersebut berada di atas taraf keberartian 0,05 sehingga terima Ho, yang artinya tidak cukup bukti untuk menerima bawah median waktu berbeda secara berarti dari 1,8 jam.

J. UJI WILCOXON

Uji Wilcoxon dapat dikatakan sebagai penyempurnaan uji tanda (Anwar, 2009). Tujuan dari uji ini sama seperti uji tanda, menguji relevansi antara 2 sampel yang berhubungan, dapat berupa sampel beruntun atau berpasangan (Kadir, 2010). Uji ini dikatakan sebagai penyempurnaan dari uji tanda karena uji Wilcoxon menguji perbedaan *treatment* dengan mempertimbangkan arah, *magnitude* relatif dan besar selisih nilai positif dan negatif dari sampel tidak hanya nilai positif dan negatifnya saja (Riadi, 2014). Untuk menguji signifikansi uji Wilcoxon ini menggunakan rumusan berikut:

$$Z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \qquad \ldots(25)$$

Keterangan:

Z = Nilai uji Wilcoxon

T = Jumlah peringkat positif atau jumlah negatif terkecil

n = Banyaknya pasangan yang tidak sama nilainya

Adapun hipotesis yang digunakan dalam pengujian Wilcoxon ini adalah:

Ho : $\tilde{\mu} = \widetilde{\mu_0}$ *(Tidak Terdapat perbedaan antara populasi 1 dan 2)*

Ha : $\tilde{\mu} \neq \widetilde{\mu_0}$ *(Terdapat perbedaan antara populasi 1 dan 2)*

Penentuan Ho diterima atau tidak berdasarkan hasil nilai z apakah terletak di daerah penerimaan hipotesis Ho atau tidak.

K. PENUTUP

Dalam bab ini, kita telah menjelajahi analisis korelasi dan uji keterkaitan *non*parametrik sebagai alat penting dalam memahami hubungan antara variabel-variabel dalam data. Kita telah membahas metode korelasi klasik seperti korelasi eta, Spearman, Darab dan Kendall, yang berguna untuk mengukur sejauh mana hubungan linear atau monotonik antara variabel-variabel kontinu atau ordinal. Selain itu, kita juga membahas uji keterkaitan *non*parametrik seperti Uji Tanda dan Uji Mann-Whitney, yang sangat berguna ketika data tidak memenuhi asumsi distribusi normal.

Melalui selintas pemahaman tentang analisis korelasi dan uji keterkaitan *non*parametrik ini, diharapkan dapat membantu pembaca untuk menerapkan alat-alat ini secara tepat dan cerdas dalam eksplorasi data mereka, memberikan wawasan berharga tentang hubungan dan pola dalam fenomena yang mereka teliti. Kesadaran akan kekuatan dan batasan dari masing-masing metode ini dapat meningkatkan kepercayaan diri dalam pengambilan keputusan statistik yang lebih akurat dan relevan dalam berbagai konteks penelitian dan analisis data.

BAB

14

PENERAPAN METODE STATISTIKA DALAM PENELITIAN

A. PENGANTAR

Mahasiswa mendapatkan kemampuan dalam mengetahui, memahami, dan mengaplikasikan penerapan metode statistika dalam penelitian utamanya dalam hal ini berfokus pada penelitian kuantitatif dengan diberbentukan alat olah data Aplikasi SPSS (*Statistical Package for the Social Sciences*), untuk pembahasan ini menggunakan SPSS V.20 kemudian disertakan langkah, *output* hasil olah data hingga pada interpretasinya.

B. CONTOH SOAL OLAH DATA REGRESI LINIER SEDERHANA

a. Siapkan data dalam bentuk Ms. Excel

Berikut contoh data pertumbuhan ekonomi (X) terhadap Tingkat Kemiskinan (Y) yang di ambil dari penelitian terdahulu bersumber dari Mudawanah, (2022), sebagai berikut:

Tabel 14.1 Data Pertumbuhan Ekonomi dan Tingkat Kemiskinan

KABUPATEN	TAHUN	Y	X
Kab Lebak	2011	9,2	1,13
Kab Lebak	2012	8,63	1,05
Kab Lebak	2013	9,5	0,98
Kab Lebak	2014	9,17	0,91
Kab Lebak	2015	9,97	0,83
Kab Lebak	2016	8,71	0,76

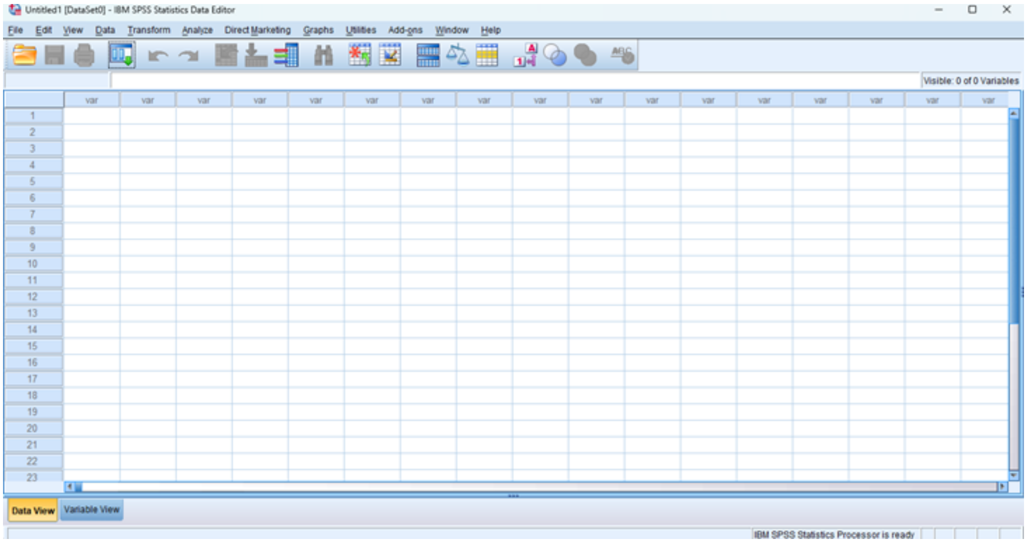
KABUPATEN	TAHUN	Y	X
Kab Lebak	2017	8,64	0,68
Kab Lebak	2018	8,41	0,6
Kab Lebak	2019	8,3	0,52
Kab Pandeglang	2011	9,8	0,84
Kab Pandeglang	2012	9,28	0,77
Kab Pandeglang	2013	10,25	0,86
Kab Pandeglang	2014	9,5	0,46
Kab Pandeglang	2015	10,43	0,55
Kab Pandeglang	2016	9,67	0,47
Kab Pandeglang	2017	9,74	0,39
Kab Pandeglang	2018	9,61	0,32
Kab Pandeglang	2019	9,42	0,24
Kab Serang	2011	5,63	1,06
Kab Serang	2012	5,28	0,98
Kab Serang	2013	5,02	0,92
Kab Serang	2014	4,87	0,84
Kab Serang	2015	5,09	0,77
Kab Serang	2016	4,58	0,69
Kab Serang	2017	4,63	0,61
Kab Serang	2018	4,3	0,53
Kab Serang	2019	4,08	0,46
Kab Tangerang	2011	6,42	3,54
Kab Tangerang	2012	5,71	3,47
Kab Tangerang	2013	5,78	3,34
Kab Tangerang	2014	5,26	3,39
Kab Tangerang	2015	5,71	3,24
Kab Tangerang	2016	5,29	3,17
Kab Tangerang	2017	5,39	3,08
Kab Tangerang	2018	5,18	3,01
Kab Tangerang	2019	5,14	2,93
Kota Cilegon	2011	3,98	1,99
Kota Cilegon	2012	3,82	1,9
Kota Cilegon	2013	3,99	1,82
Kota Cilegon	2014	3,81	1,76
Kota Cilegon	2015	4,1	1,68
Kota Cilegon	2016	3,57	1,6

KABUPATEN	TAHUN	Y	X
Kota Cilegon	2017	3,52	1,53
Kota Cilegon	2018	3,25	1,46
Kota Cilegon	2019	3,03	1,37
Kota Serang	2011	6,25	2,2
Kota Serang	2012	5,7	2,14
Kota Serang	2013	5,92	2,06
Kota Serang	2014	5,7	1,99
Kota Serang	2015	6,28	1,92
Kota Serang	2016	5,58	1,83
Kota Serang	2017	5,57	1,77
Kota Serang	2018	5,36	1,68
Kota Serang	2019	5,28	1,59
Kota Tangerang	2011	6,14	2,66
Kota Tangerang	2012	5,56	2,59
Kota Tangerang	2013	5,26	2,51
Kota Tangerang	2014	4,91	2,43
Kota Tangerang	2015	5,04	2,36
Kota Tangerang	2016	4,94	2,28
Kota Tangerang	2017	4,95	2,21
Kota Tangerang	2018	4,76	2,12
Kota Tangerang	2019	4,43	2,04
Kota Tangerang Selatan	2011	1,5	3,67
Kota Tangerang Selatan	2012	1,33	3,59
Kota Tangerang Selatan	2013	1,75	3,51
Kota Tangerang Selatan	2014	1,68	3,44
Kota Tangerang Selatan	2015	1,69	3,36
Kota Tangerang Selatan	2016	1,67	3,28
Kota Tangerang Selatan	2017	1,76	3,21
Kota Tangerang Selatan	2018	1,68	3,13
Kota Tangerang Selatan	2019	1,68	3,04

Sumber: Mudawanah, (2022)

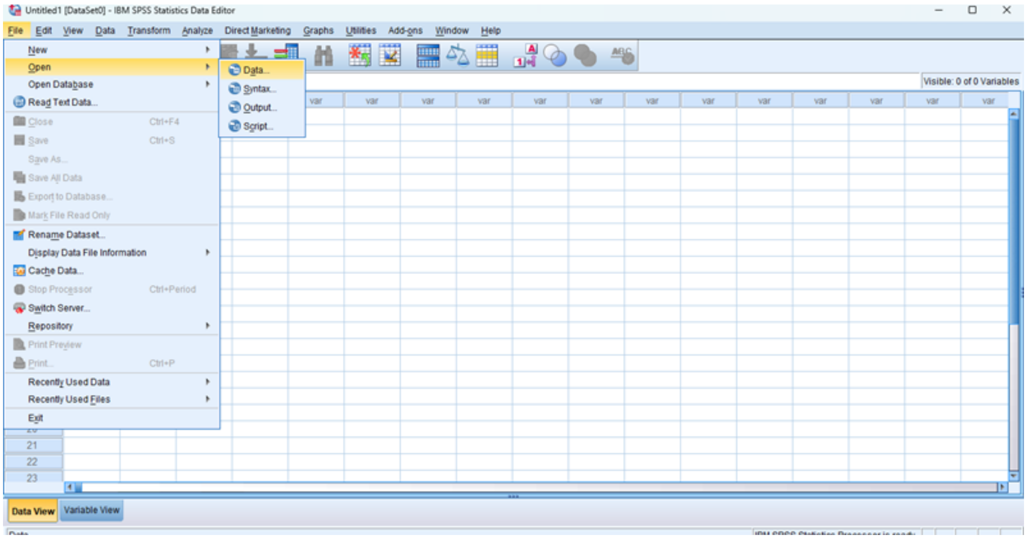
b. Masukkan data ke dalam aplikasi SPSS

1. Buka aplikasi SPSS

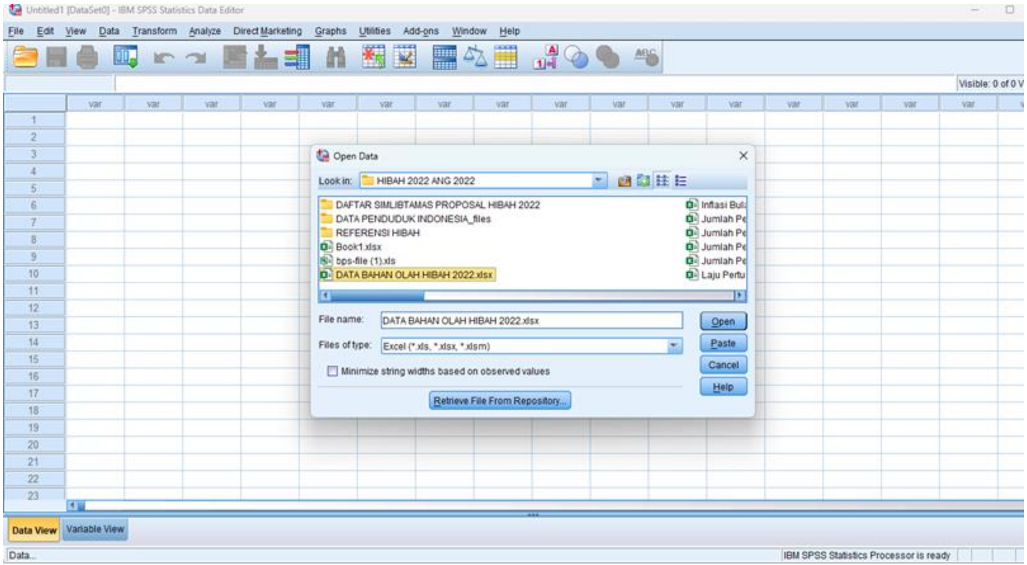


2. *Open file*

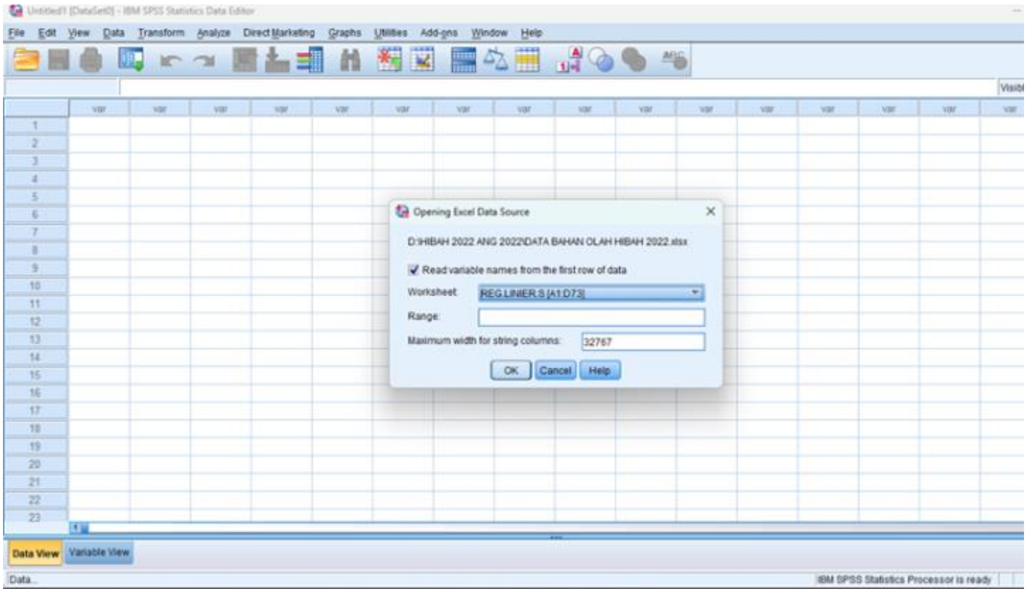
   Klik *file-open-data*



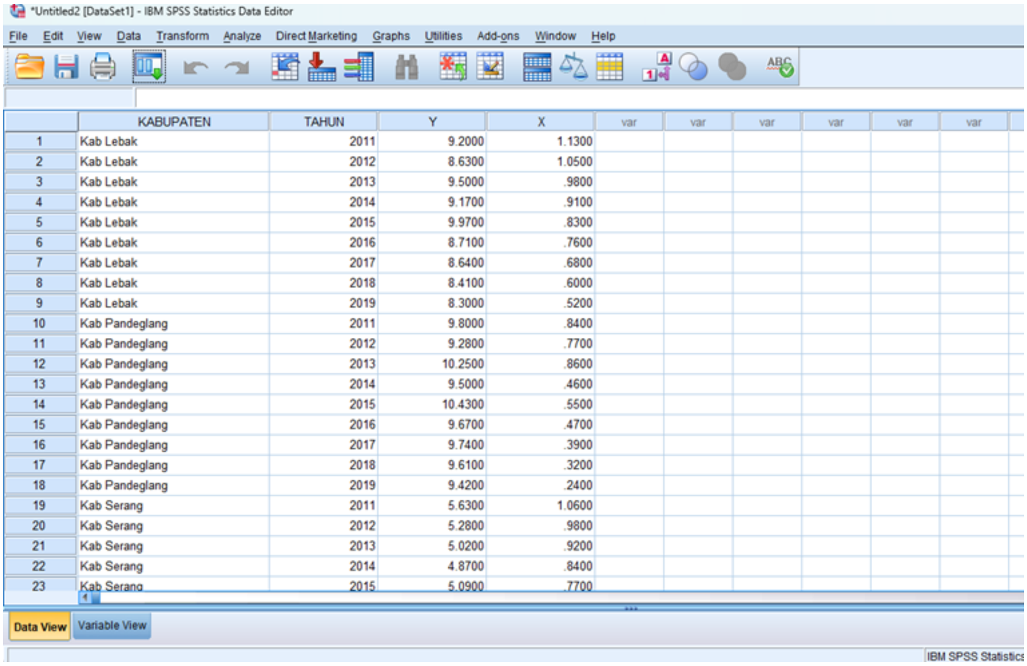
Pilih *file* yang telah kita buat dalam bentuk ms.excel



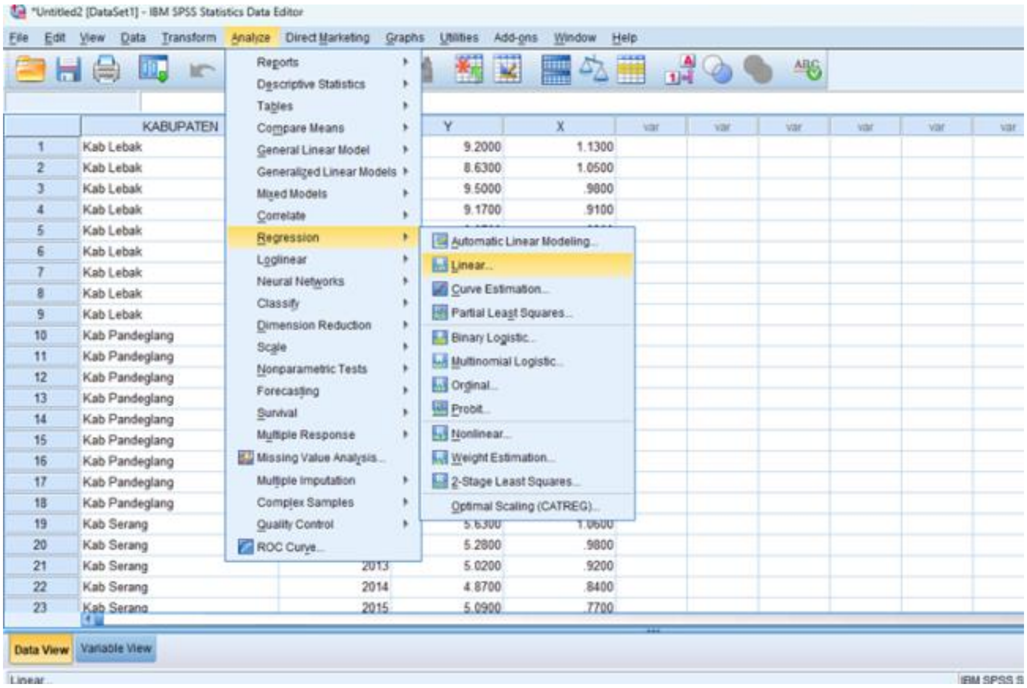
Pastikan *sheet* pada ms.excel yang akan kita masukkan telah sesuai



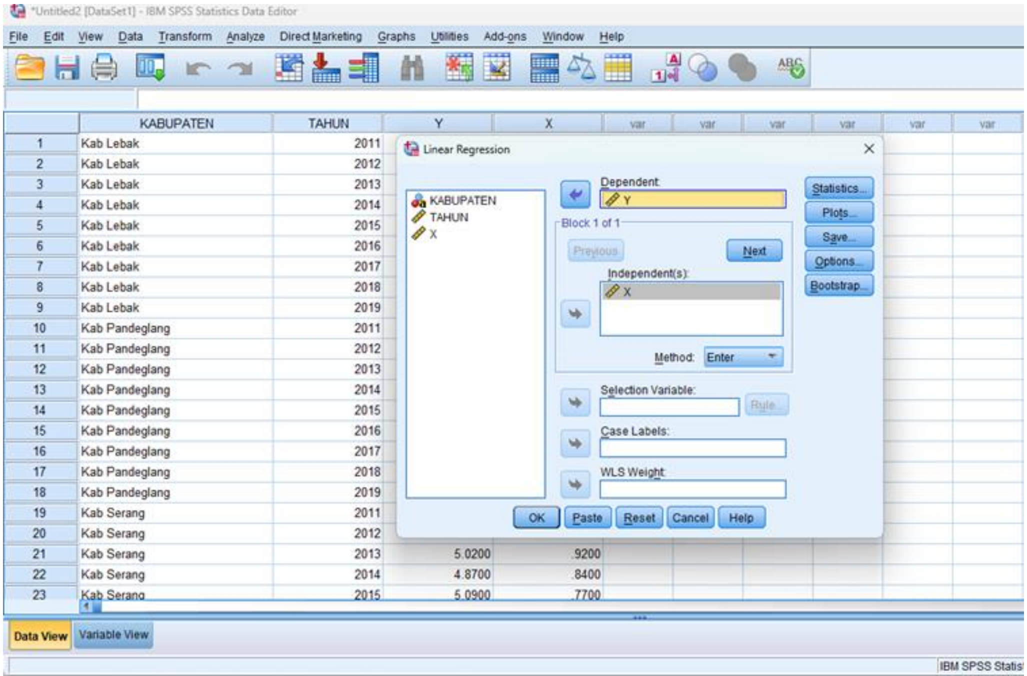
Akan muncul tampilan berikut ini



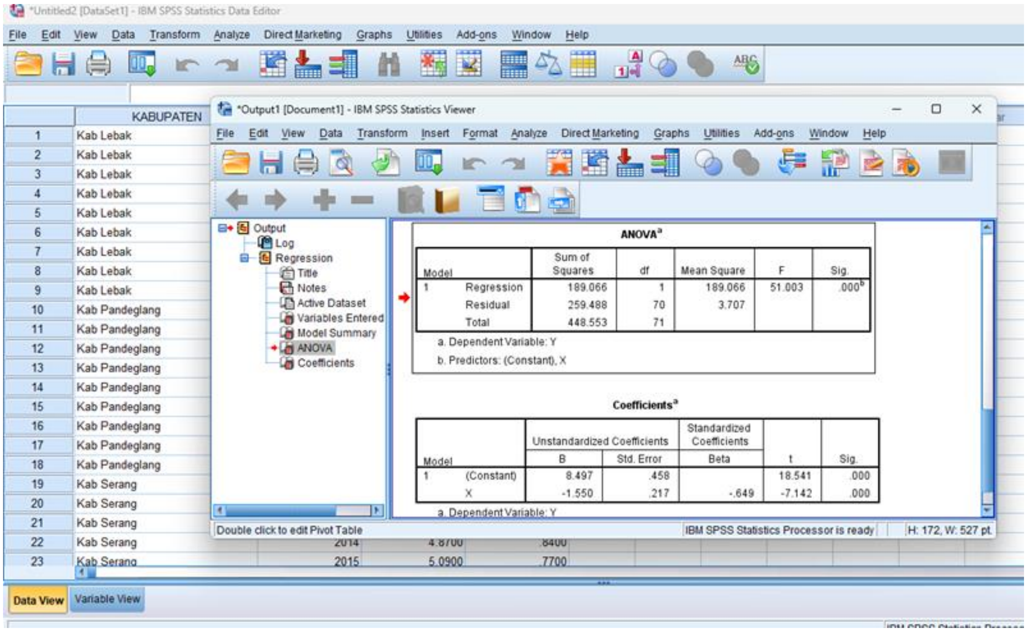
Untuk mengetahui hasil uji t (uji pengaruh secara parsial) maka dapat dilakukan yaitu pilih menu *analyze-regression-linier*



Pada menu *regression liner* maka masukkan variabel Y ke *dependent* dan *variabel* X ke *independent*, lalu tekan ok

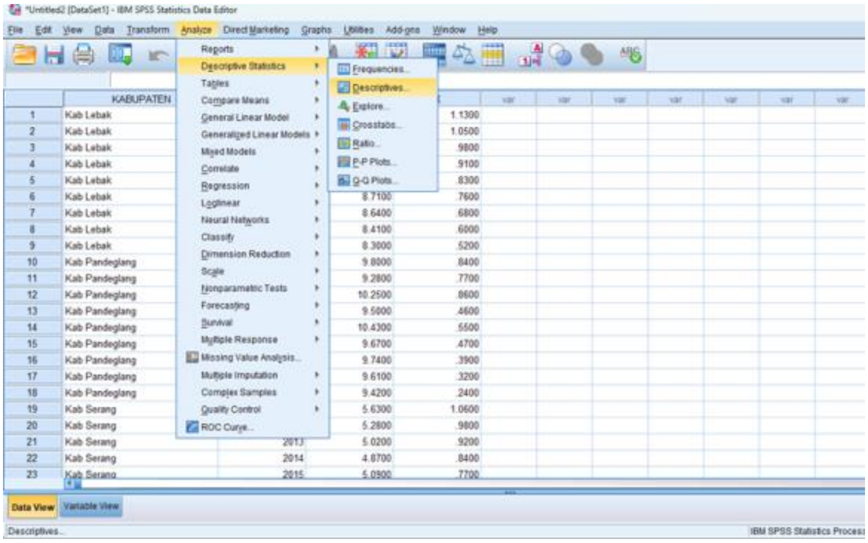


Akan muncul hasil *output* spss sebagai berikut

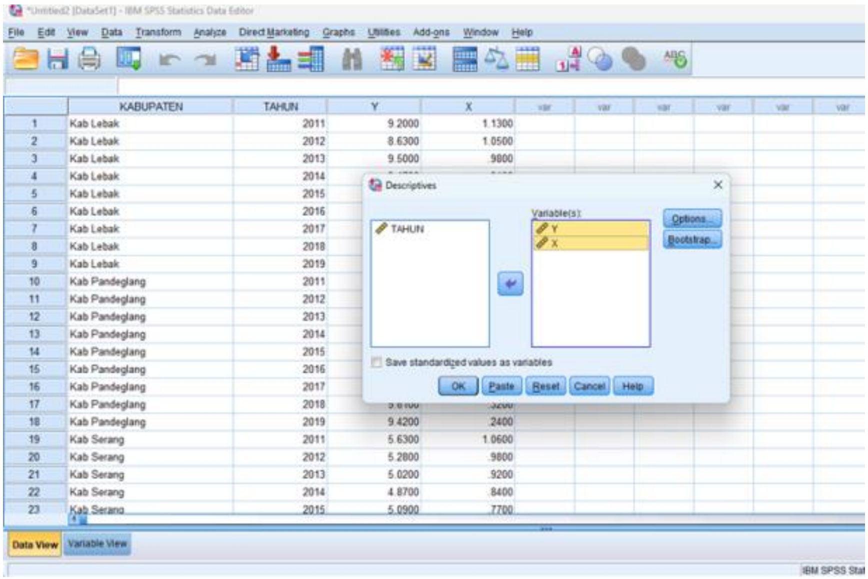


Terlihat bahwa hasil *output* spss pada tabel *Coefficients* menggambarkan bahwa dengan menggunakan alat analisis spss telah menghasilkan uji t dengan nilai signifikan 0,000 dibawah nilai *probability* 0,05 maka artinya pertumbuhan ekonomi (X) memiliki pengaruh terhadap tingkat kemiskinan (Y). Adapun deskripsi penelitian dapat dilakukan dengan cara sebagai berikut:

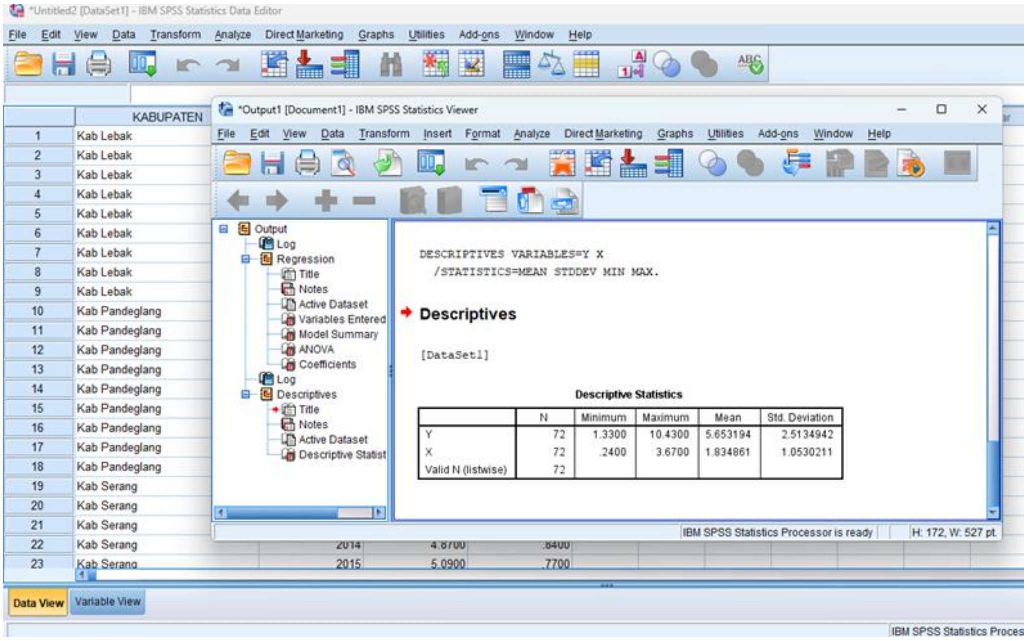
Klik *Analyze-Descriptive statistics-Descriptives*



Kemudian, akan muncul kotak dialog yang perlu dilakukan adalah memindahkan X dan Y dengan memblok X dan Y kemudian klik tanda panah untuk memindahkan ke kolom *Variable* (s) lalu klik ok yaitu dapat dilihat di bawah ini



Maka, akan muncul tabel hasil di *output* spss tabel *descriptive statistics* sebagai berikut



Terlihat bahwa nilai minimum dan maksimum serta nilai rata-rata dari masing-masing variabel yang diteliti.

C. CONTOH SOAL OLAH DATA REGRESI LINIER BERGANDA

a. Siapkan data dalam bentuk Ms. Excel

Berikut contoh data pertumbuhan ekonomi (X1) dan Upah Minimum (X2) terhadap Tingkat Kemiskinan (Y) yang di ambil dari penelitian terdahulu bersumber dari Mudawanah, (2022), sebagai berikut:

Tabel 14.2 Data Pertumbuhan Ekonomi, Upah Minimum dan Tingkat Kemiskinan

KABUPATEN	TAHUN	Y	X1	X2
Kab Lebak	2011	9,2	1,13	1.007.500
Kab Lebak	2012	8,63	1,05	1.047.800
Kab Lebak	2013	9,5	0,98	1.187.500
Kab Lebak	2014	9,17	0,91	1.490.000
Kab Lebak	2015	9,97	0,83	1.728.000
Kab Lebak	2016	8,71	0,76	1.965.000
Kab Lebak	2017	8,64	0,68	2.127.112
Kab Lebak	2018	8,41	0,6	2.312.384

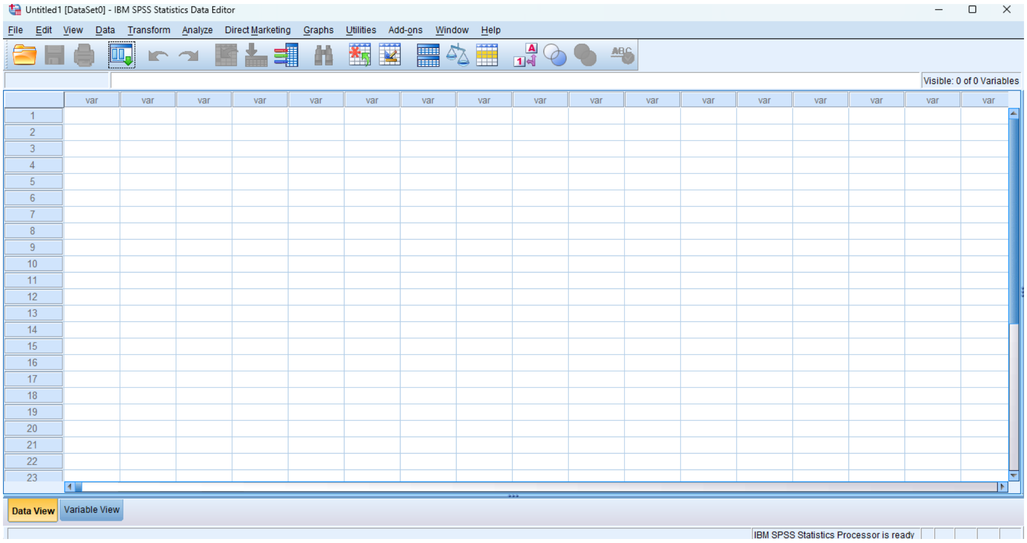
KABUPATEN	TAHUN	Y	X1	X2
Kab Lebak	2019	8,3	0,52	2.498.068
Kab Pandeglang	2011	9,8	0,84	1.015.000
Kab Pandeglang	2012	9,28	0,77	1.050.000
Kab Pandeglang	2013	10,25	0,86	1.182.000
Kab Pandeglang	2014	9,5	0,46	1.418.000
Kab Pandeglang	2015	10,43	0,55	1.737.000
Kab Pandeglang	2016	9,67	0,47	1.999.981
Kab Pandeglang	2017	9,74	0,39	2.164.979
Kab Pandeglang	2018	9,61	0,32	2.363.549
Kab Pandeglang	2019	9,42	0,24	2.542.539
Kab Serang	2011	5,63	1,06	1.189.600
Kab Serang	2012	5,28	0,98	1.320.500
Kab Serang	2013	5,02	0,92	2.080.000
Kab Serang	2014	4,87	0,84	2.340.000
Kab Serang	2015	5,09	0,77	2.700.000
Kab Serang	2016	4,58	0,69	3.010.500
Kab Serang	2017	4,63	0,61	3.258.866
Kab Serang	2018	4,3	0,53	3.542.714
Kab Serang	2019	4,08	0,46	3.827.193
Kab Tangerang	2011	6,42	3,54	1.285.000
Kab Tangerang	2012	5,71	3,47	1.527.000
Kab Tangerang	2013	5,78	3,34	2.200.000
Kab Tangerang	2014	5,26	3,39	2.442.000
Kab Tangerang	2015	5,71	3,24	2.710.000
Kab Tangerang	2016	5,29	3,17	3.021.650
Kab Tangerang	2017	5,39	3,08	3.270.936
Kab Tangerang	2018	5,18	3,01	3.555.835
Kab Tangerang	2019	5,14	2,93	3.841.368
Kota Cilegon	2011	3,98	1,99	1.224.000
Kota Cilegon	2012	3,82	1,9	1.347.000
Kota Cilegon	2013	3,99	1,82	2.200.000
Kota Cilegon	2014	3,81	1,76	2.443.000
Kota Cilegon	2015	4,1	1,68	2.760.590
Kota Cilegon	2016	3,57	1,6	3.078.058
Kota Cilegon	2017	3,52	1,53	3.331.997
Kota Cilegon	2018	3,25	1,46	3.622.215
Kota Cilegon	2019	3,03	1,37	3.913.078
Kota Serang	2011	6,25	2,2	1.156.000
Kota Serang	2012	5,7	2,14	1.231.000

KABUPATEN	TAHUN	Y	X1	X2
Kota Serang	2013	5,92	2,06	1.798.446
Kota Serang	2014	5,7	1,99	2.166.000
Kota Serang	2015	6,28	1,92	2.375.000
Kota Serang	2016	5,58	1,83	2.648.125
Kota Serang	2017	5,57	1,77	2.866.595
Kota Serang	2018	5,36	1,68	3.116.276
Kota Serang	2019	5,28	1,59	3.366.512
Kota Tangerang	2011	6,14	2,66	1.290.000
Kota Tangerang	2012	5,56	2,59	1.527.000
Kota Tangerang	2013	5,26	2,51	2.203.000
Kota Tangerang	2014	4,91	2,43	2.444.301
Kota Tangerang	2015	5,04	2,36	2.730.000
Kota Tangerang	2016	4,94	2,28	3.043.950
Kota Tangerang	2017	4,95	2,21	3.295.075
Kota Tangerang	2018	4,76	2,12	3.582.077
Kota Tangerang	2019	4,43	2,04	3.869.717
Kota Tangerang Selatan	2011	1,5	3,67	1.290.000
Kota Tangerang Selatan	2012	1,33	3,59	1.527.000
Kota Tangerang Selatan	2013	1,75	3,51	2.200.000
Kota Tangerang Selatan	2014	1,68	3,44	2.442.000
Kota Tangerang Selatan	2015	1,69	3,36	2.710.000
Kota Tangerang Selatan	2016	1,67	3,28	3.021.650
Kota Tangerang Selatan	2017	1,76	3,21	3.270.936
Kota Tangerang Selatan	2018	1,68	3,13	3.555.835
Kota Tangerang Selatan	2019	1,68	3,04	3.841.368

Sumber: Mudawanah, (2022)

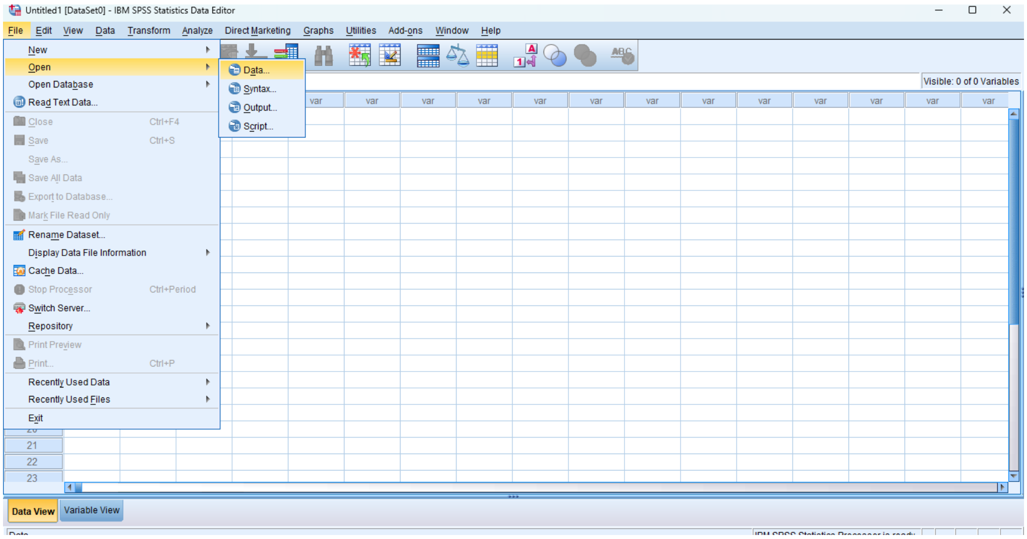
b. Masukkan data ke dalam aplikasi SPSS

1. Buka aplikasi SPSS

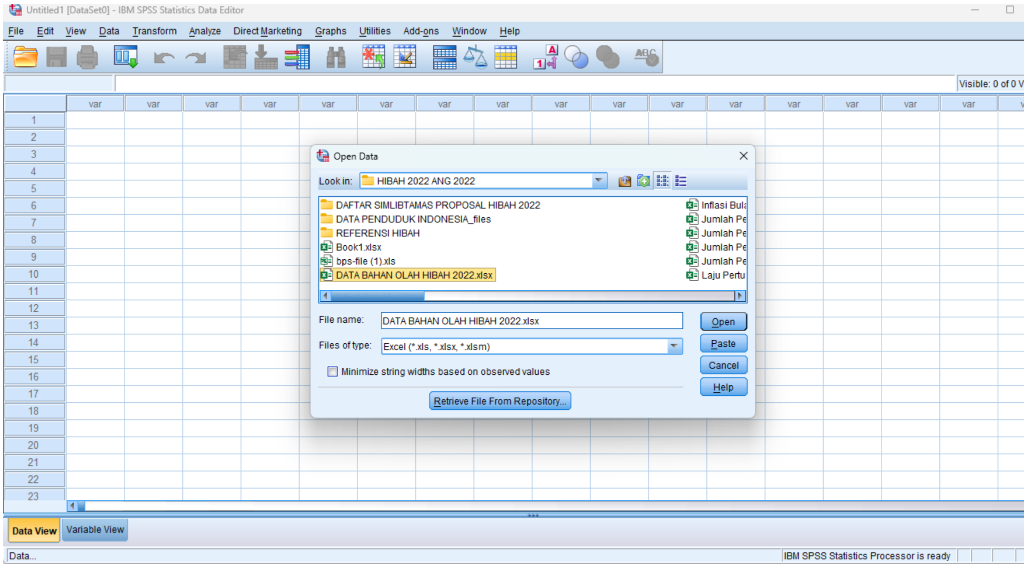


2. *Open file*

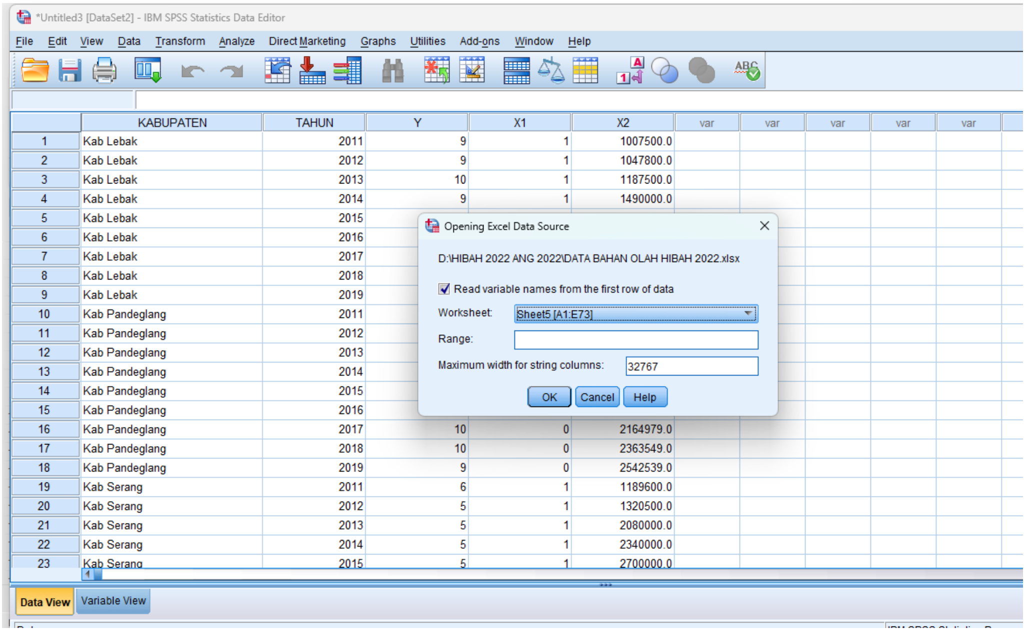
   Klik *file-open-data*



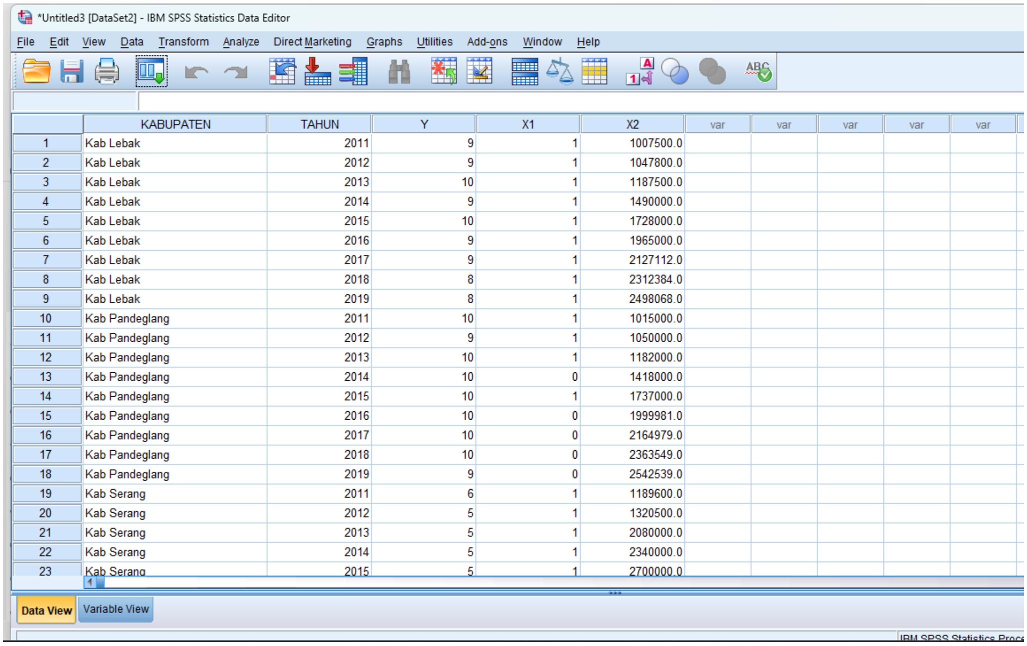
Pilih *file* yang telah kita buat dalam bentuk ms.excel



Pastikan *sheet* pada ms.excel yang akan kita masukkan telah sesuai

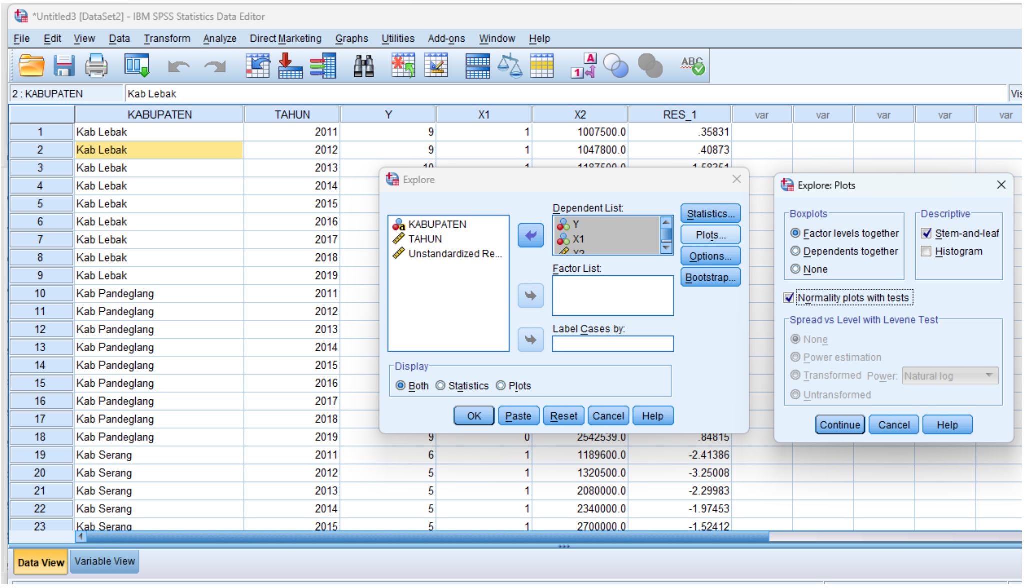


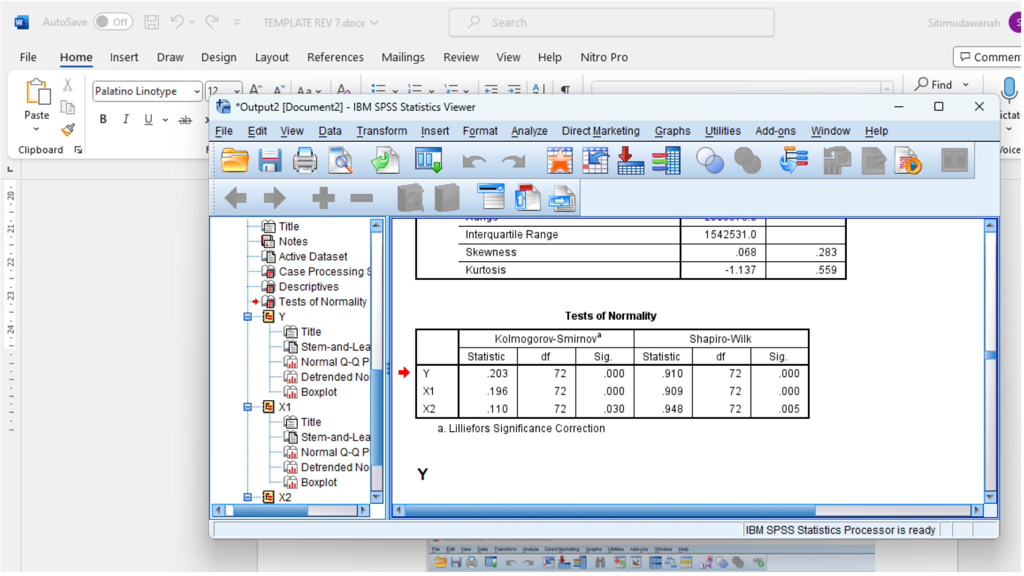
Akan muncul tampilan berikut ini



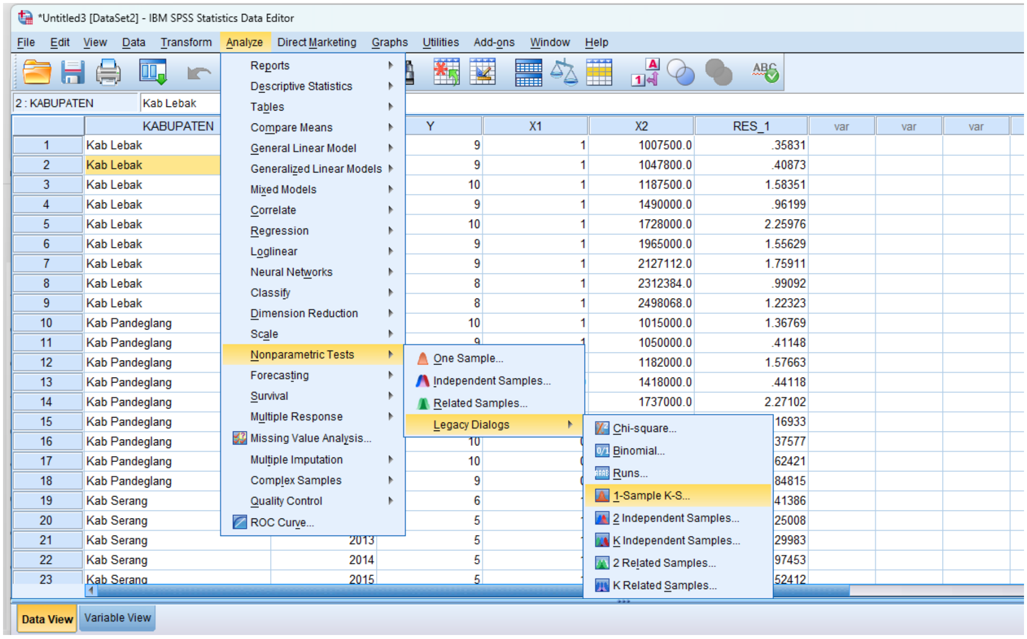
	KABUPATEN	TAHUN	Y	X1	X2
1	Kab Lebak	2011	9	1	1007500.0
2	Kab Lebak	2012	9	1	1047800.0
3	Kab Lebak	2013	10	1	1187500.0
4	Kab Lebak	2014	9	1	1490000.0
5	Kab Lebak	2015	10	1	1728000.0
6	Kab Lebak	2016	9	1	1965000.0
7	Kab Lebak	2017	9	1	2127112.0
8	Kab Lebak	2018	8	1	2312384.0
9	Kab Lebak	2019	8	1	2498068.0
10	Kab Pandeglang	2011	10	1	1015000.0
11	Kab Pandeglang	2012	9	1	1050000.0
12	Kab Pandeglang	2013	10	1	1182000.0
13	Kab Pandeglang	2014	10	0	1418000.0
14	Kab Pandeglang	2015	10	1	1737000.0
15	Kab Pandeglang	2016	10	0	1999981.0
16	Kab Pandeglang	2017	10	0	2164979.0
17	Kab Pandeglang	2018	10	0	2363549.0
18	Kab Pandeglang	2019	9	0	2542539.0
19	Kab Serang	2011	6	1	1189600.0
20	Kab Serang	2012	5	1	1320500.0
21	Kab Serang	2013	5	1	2080000.0
22	Kab Serang	2014	5	1	2340000.0
23	Kab Serang	2015	5	1	2700000.0

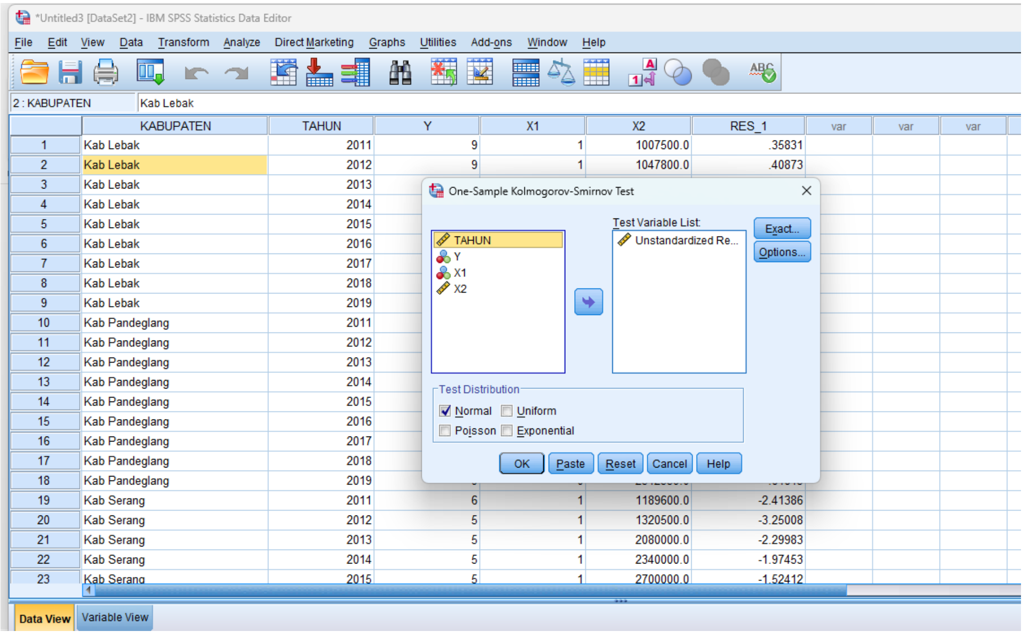
Langkah pertama dapat menguji normalitas data dengan cara klik *analyze-deskriptive statistics-explore* lalu akan muncul kotak dialog pilih *plot-normality plots with tests* (Nuryadi *et al.*, 2017).

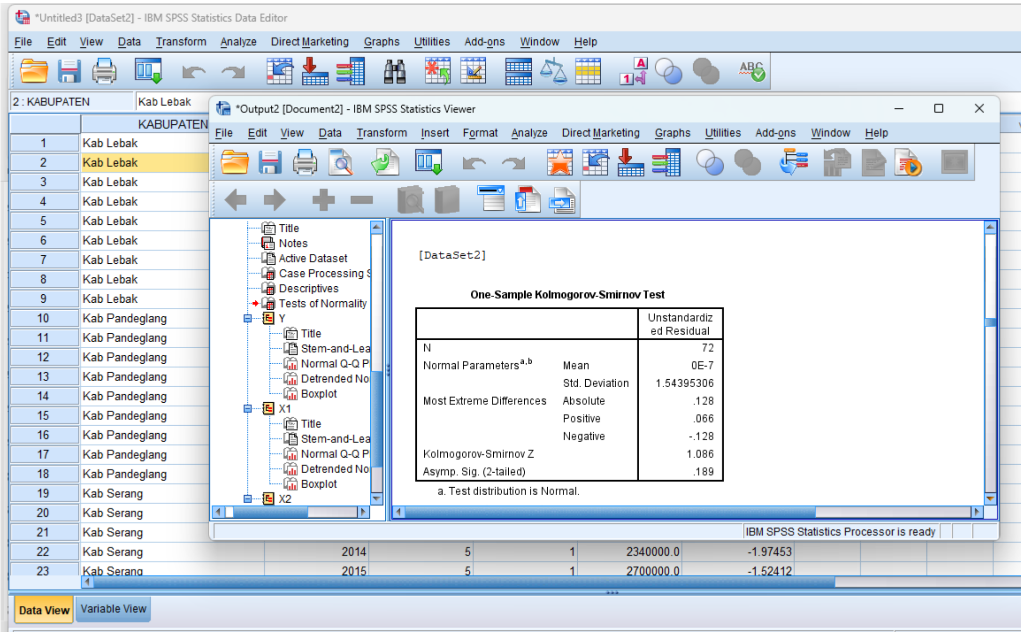




Terlihat hasil uji normalitas bahwa berdasarkan uji *Kolmogorov-smirnov* dan *Shapiro-wilk* masing-masing variabel tidak memiliki distribusi data yang normal. Alternatif dalam uji normalitas menggunakan uji secara Bersama-sama klik menu *analyze-nonparametric tests-legacy dialogs-1-sample k-s* sebagai berikut:

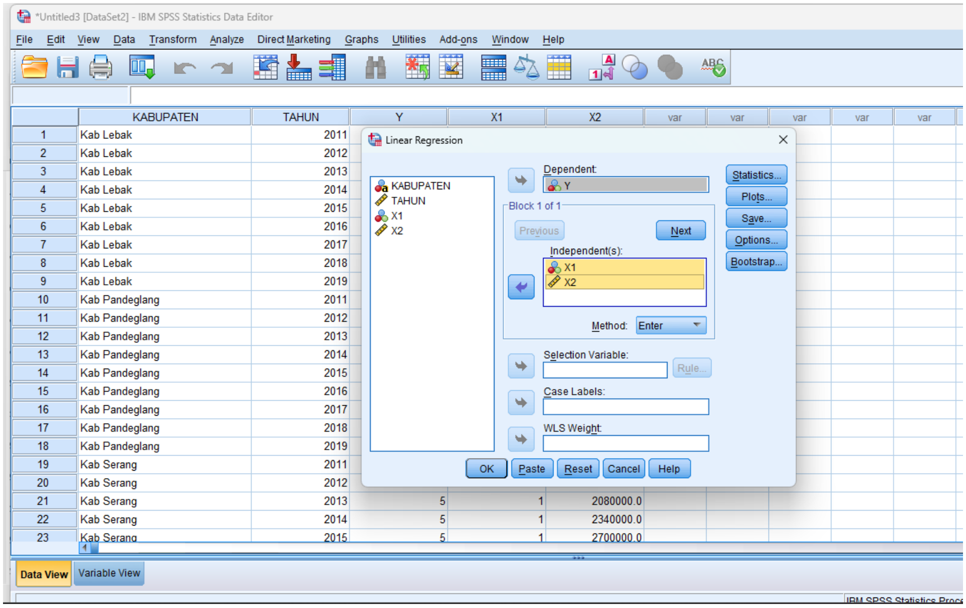




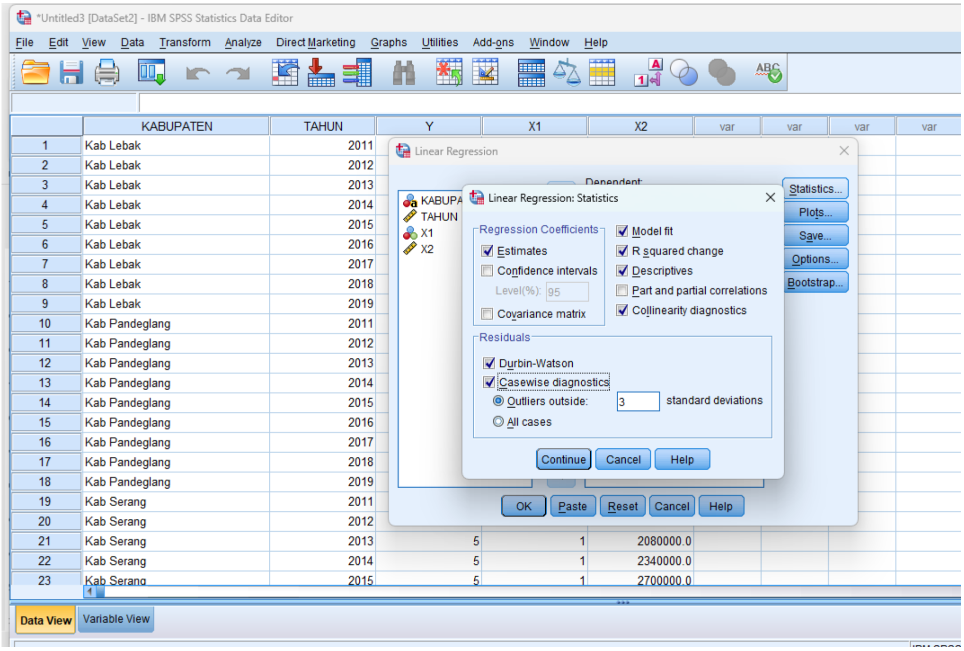


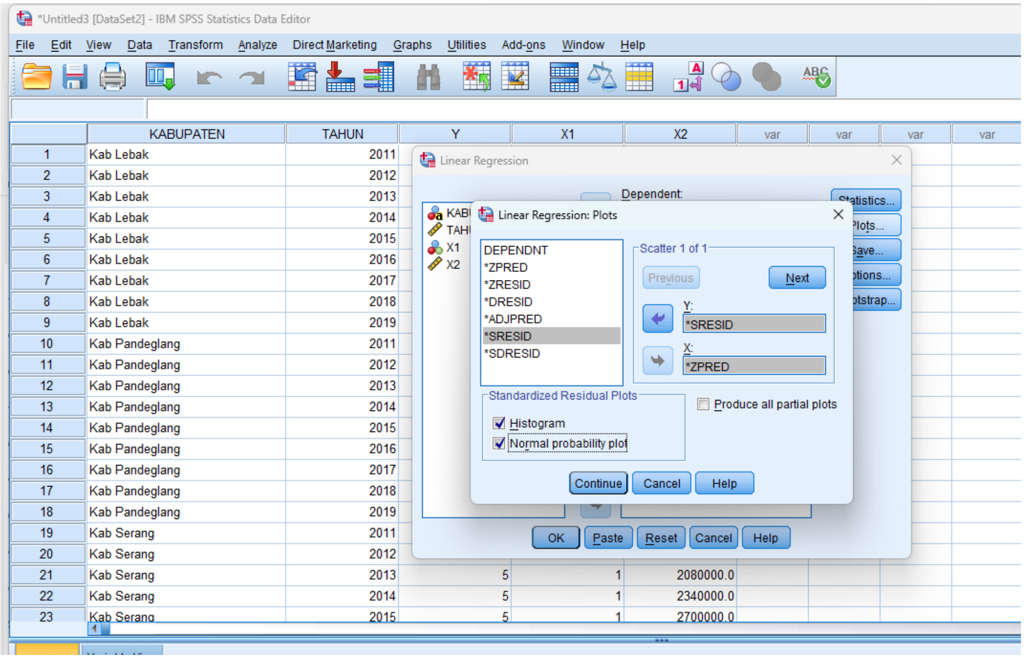
Berdasarkan hasil uji mendapatkan hasil *Asymp. Sig* (2-tailed) adalah sebesar 0,189 artinya nilai lebih dari 0,05 maka data berdistribusi normal, (untuk mendapatkan nilai *unstandardized* dapat menguji *analyze-regression-linier* terlebih dahulu).

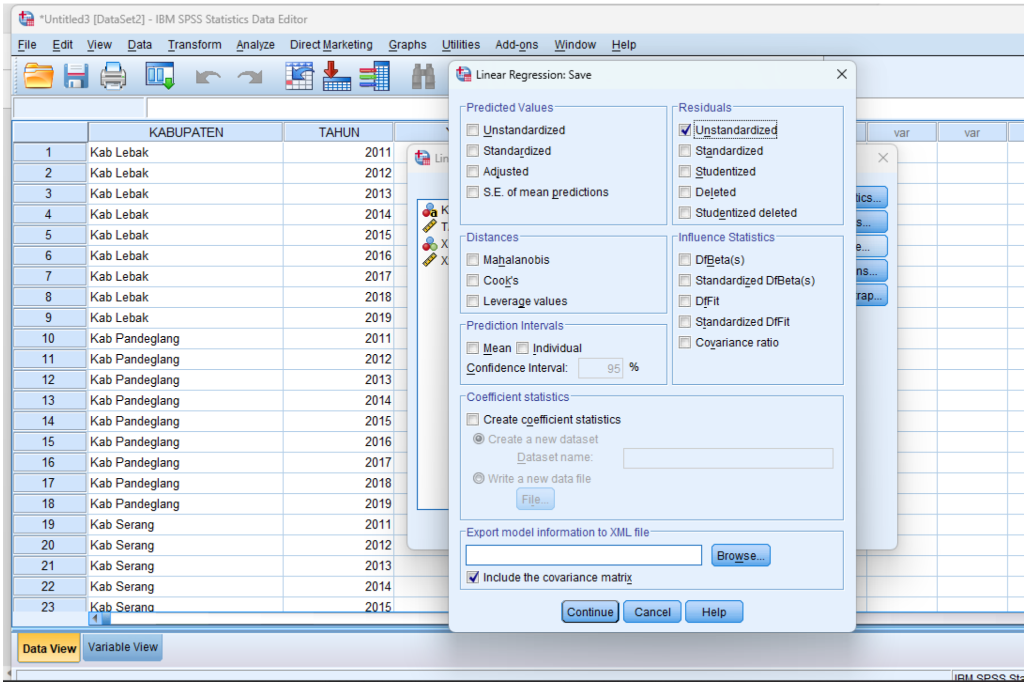
Adapun untuk mengetahui hasil uji t (uji pengaruh secara parsial) dan uji F (pengaruh secara simultan atau bersama-sama) serta terlebih dahulu perlu mengetahui asumsi klasik sebelum uji regresi linier berganda maka dapat dilakukan yaitu pilih menu *analyze-regression-linier*



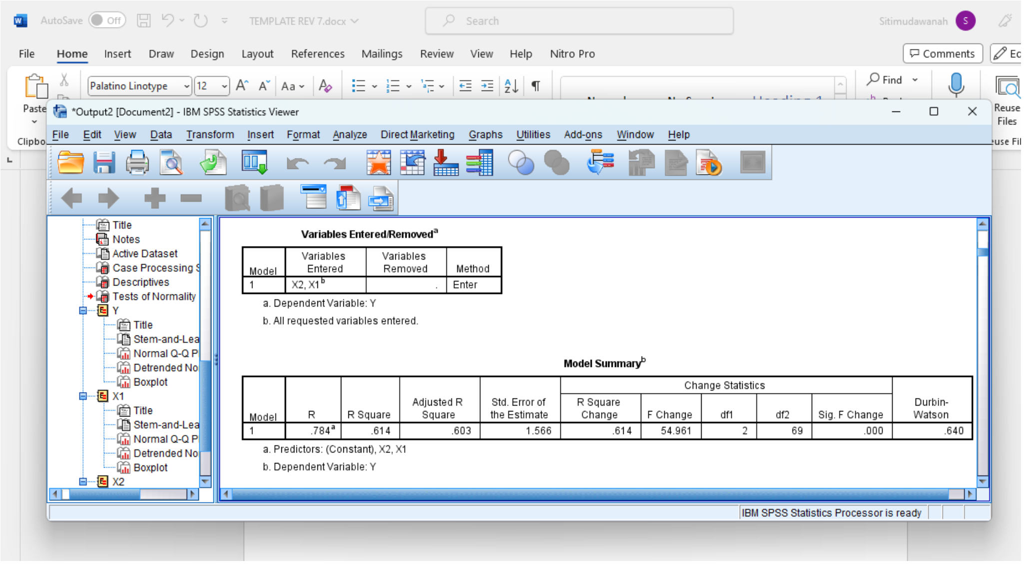
Pada menu *regression liner* maka masukkan variabel Y ke *dependent* dan variabel X1 dan X2 ke *independent*, kemudian lakukan cek ceklis menu *statistics*, *plots* dan *save* klik *continue* seperti di bawah ini:







Lalu klik ok dan akan muncul hasil *output* spss sebagai berikut



Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	X2, X1[b]	.	Enter

a. Dependent Variable: Y

b. All requested variables entered.

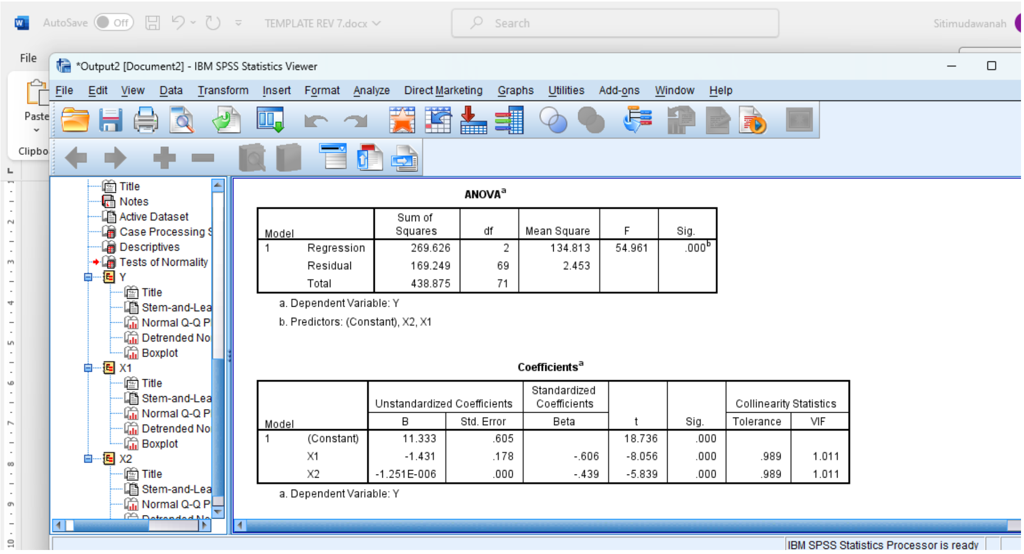
Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.784[a]	.614	.603	1.566	.614	54.961	2	69	.000	.640

a. Predictors: (Constant), X2, X1

b. Dependent Variable: Y

Hasil durbin Watson menggambarkan hasil uji autokorelasi, terlihat sebesar 0,640 tidak terjadi autokorelasi karena nilai diantara -2 sampai dengan +2 (Mudawanah, 2019). Adapun nilai *Adjusted R Square* senilai 0,603 artinya nilai determinan tersebut adalah 60,3% variabel Y dipengaruhi oleh variabel X1 dan X2, sisanya 39,7% dipengaruhi faktor lainnya.



ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	269.626	2	134.813	54.961	.000[b]
	Residual	169.249	69	2.453		
	Total	438.875	71			

a. Dependent Variable: Y

b. Predictors: (Constant), X2, X1

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	11.333	.605		18.736	.000		
	X1	-1.431	.178	-.606	-8.056	.000	.989	1.011
	X2	-1.251E-006	.000	-.439	-5.839	.000	.989	1.011

a. Dependent Variable: Y

Nilai VIF menunjukkan hasil di antara 1 sampai 10 atau di bawah 5 maka memiliki arti bahwa tidak terjadi multikolinearitas (Akinwande, Dikko *and* Samson, 2015). Terlihat nilai signifikan pada masing-masing nilai t pada setiap variabel di bawah 0,05 maka seluruh variabel secara parsial memiliki pengaruh yang signifikan terhadap Y. Juga pada tabel Anova terlihat nilai F yang lebih besar dari F tabel dengan memiliki nilai signifikan di bawah 0,05 maka secara simultan atau Bersama-sama variabel X1 dan X2 berpengaruh terhadap Y.

D. PENUTUP

Demikianlah penerapan metode statistika dalam penelitian utamanya dalam hal ini berfokus pada penelitian kuantitatif dengan diberbentukan alat olah data Aplikasi SPSS (*Statistical Package for the Social Sciences*), untuk pembahasan ini menggunakan SPSS V.20 kemudian disertakan langkah, *output* hasil olah data hingga pada interpretasinya.

Diharapkan setelah membaca pembahasan ini mendapatkan pengetahuan pemahaman mengenai langkah olah data menggunakan SPSS dalam penelitian kuantitatif, mendapatkan pengetahuan pemahaman mengenai *output* hasil olah data menggunakan SPSS dalam penelitian kuantitatif, dan mendapatkan pengetahuan pemahaman mengenai interpretasi hasil olah data menggunakan SPSS dalam penelitian kuantitatif.

Selanjutnya, bagi pembaca kritik saran yang membangun sangat diharapkan oleh penulis untuk pengembangan pada materi ini karena tentu penyusunan materi pembahasan ini masih terdapat beberapa kekurangan, terimakasih.

DAFTAR PUSTAKA

Abdy, M. (2019). Tinjauan Singkat Tentang Regresi Parametrik dan Nonparametrik. *Saintifik* , 58-62.

Akinwande, M. O., Dikko, H. G. and Samson, A. (2015) 'Variance Inflation Factor: As a Condition for the Inclusion of Suppressor Variable(s) in Regression Analysis', *Open Journal of Statistics*, 05(07), pp. 754–767. doi: 10.4236/ojs.2015.57075.

Alwi, W., Muh.Irwan, & Musfirah. (2021). Penerapan Regresi Nonparametrik Spline dalam memodelkan faktor-faktor yang mempengaruhi Indkes Pembangunan Manusia (IPM) Indonesia Tahun 2018. *Jurnal Matemtika dan Statistika serta Aplikasinya* , 70-81.

Al-Ziadi, N.A.J. and Ramadhan, A.M. (2022) 'A New Class of Holomorphic Univalent Functions Defined by Linear Operator', *JOURNAL OF ADVANCES IN MATHEMATICS*, 21, pp. 116–124. Available at: https://doi.org/10.24297/jam.v21i.9258.

Ananda, R., Fadhli, M., 2018. Statistik Pendidikan (Teori dan Praktik Dalam Pendidikan). CV Widya Puspita, Medan

Anggraeni, D. P., Dewi, I. R., & Rio Satriyantara, e. a. (2021). Pelatihan penggunaan Statistik Parametrik Untuk Meningkatkan Motivasi Pegawai Fungional Perencanaan Bappeda KLU dalam Penelitian dan Publikasi Ilmiah. *Alamtana : Jurnal Pengabdian Masyarakat* , 75-80.

Anwar, A. (2009) *Statistika untuk Penelitian Pendidikan dan Aplikasinya dengan SPSS dan Excel*, *IAIT Press*.

Asra, A., Sutomo, S. 2016. Pengantar Statistika I; Panduan Bagi Pengajar dan Mahasiswa. PT Rajagrafindo Persada, Depok.

Asrol., 2019. Modul Kuliah Statistik Ekonomi dan Bisnis. Fakultas Pertanian Universitas Islam Riau, Pekanbaru.

Atmaja, L. S., 2009. *Statistika untuk bisnis dan ekonomi.* Jakarta: Penerbit ANDI.

Azizah, Annisa. *Uji Korelasi Eta, Statistika Non Parametrik*. UIN Sunan Gunung Djati Bandung.

Badan Perencanaan Pembangunan Nasional. (2023). Infografik Investasi Asing di Indonesia, diakses melalui *www.bappenas.go.id*.

Badan Pusat Statistik, (2023). Statistik Sosial dan Kependudukan, diakses melalui *www. bps.go.id*

Bayoud, H. A. (2021) 'Tests of normality: new test and comparative study', *Communications in Statistics: Simulation and Computation*, 50(12), pp. 4442–4463. doi: 10.1080/03610918.2019.1643883.

Bluman, Allan G., 2009. *Elementry Statistics : A Step By Step Approach. 7th ed.* McGraw-Hill, New York.

Boediono, Koster, Wayan, 2014. Teori dan Aplikasi Statistika dan Probabilitas. Penerbit PT Remaja Rosdakarya, Bandung.

Chalmers, D. 2018. Cambridge International AS & A Level Mathematics: Probability & Statistics 1. Cambridge University Press. Cambridge. United Kingdom.

Corder, G.W. and Foreman, D.I. (2014) *Nonparametric Statistics A Step-by-step Approach*. 2nd edn. John Wiley & Sons, Inc.

Dahman, M. (2018) 'Applied Multivariate Statistical Modeling Chapter Nine-Analysis Of Variance (ANOVA) License: CC-By Attribution 4.0 International', in, pp. 2–10. Available at: https://10.0.121.243/osf.io/g8qbw.

Fein, E. C. *et al.* (2021) *Statistics for Research Students*. University of Southern Queensland. Available at: https://usq.pressbooks.pub/statisticsforresearchstudents/.

Fernholz, L. T. (2011) 'Target Estimation: A New Approach to Parametric Estimation', *International Encyclopedia of Statistical Science*, pp. 1583–1585. doi: 10.1007/978-3-642-04898-2_586.

Furqon. 2013. Statistika Terapan Untuk Penelitian. CV Alfabeta. Bandung

Hadi, Sutrisno, 2015. Statistik. Pustaka Pelajar, Yogyakarta.

Hamzah, L, M., Awaluddin, I., Maimunah E., (2016). Pengantar Statistika Ekonomi, CV Anugrah Utama Raharja, Lampung

Hanafiah., Sutedja, A., Ahmaddien, I., (2020). Pengantar Statistika. Penerbit Widina Bhakti Persada Bandung, Bandung.

Harisuddin, M.I., 2021. Kemampuan Pemecahan Masalah Matematis dan Kemandirian Belajar Siswa dengan PJJ Dimasa Covid-19. Teorema : Teori dan Riset Matematika, 6(1), 98-106, Universitas Galuh.

Harisuddin, M.I., Sriyanti, I., Wijaya, H., 2022. *Implementation of Accelerated Learning to Improve Mathematics Communication Ability*. Jurnal Pendidikan MIPA, 23 (4), 1409-1422. Jurusan Matematika FMIPA, Universitas Lampung.

Hartanto, D., Yuliani, S., 2019. Statistik Riset Pendidikan Dilengkapi Analisis SPSS. Cahaya Firdaus, Pekanbaru.

Hasan, I., 2012. *Pokok-pokok Materi Statistika 2*. 7 penyunt. Jakarta: PT Bumu Aksara.

Hasan, M. (2017). *Statistik Inferensi*. Jakarta: Bum iAksara.

Herrhyanto,N., T Gantini ., 2009. Pengantar Statistika Matematis. Yrama Widya, Bandung.

Hidayat, A. (2023, July 14). *Penjelasn Koefisien Kotingensi C dan Cara Hitung Lengkap*. Diambil kembali dari Statistikian: statistikian.com/2012/08/koefisien-kotingensi-c-html

Hidayat, Anwar (2012), *Tutorial Uji Eta dan Contoh Penerapan Rumus Uji Eta*. Statistikian.com

Hooda, R.P., 2013. Statistics for Business and Economics. Vikas Publishing House Pvt Ltd, India.

Huang, K.-W. *et al.* (2019) 'Computer Vision and Metrics Learning for Hypothesis Testing: An Application of Q-Q Plot for Normality Test', *arXiv preprint arXiv:1901.07851*, pp. 1–14. Available at: http://arxiv.org/abs/1901.07851.

Isotalo, J. (2010) *Basics of Statistics, CreateSpace Independent Publishing Platform, 2014*. CreateSpace Independent Publishing Platform, 2014.

Jagostatitsik. (2022, Februari 27). *Contoh Soal Koefisien Korelasi Kontingensi, Penyelesaian Menggunakan SPSS*. Dipetik Juli 25, 2023, dari Statistics Repository: https://statistikapedia.com/artikel/contoh-soal-koefisien-korelasi-kontingensi-penyelesaian-menggunakan-spss/

Jiamwattanapong, K. and Ingadapa, N. (2020) 'Performance of tests for homogeneity of variances for more than two samples 1', *International Journal of Management and Applied Science*, 6(2), pp. 57–62.

Kadir (2010) *Statistika Untuk Penelitian Ilmu-Ilmu Sosial*. Edited by Juredi. Rosematqa Sampurna.

Kenneth, J.B., Janis, E.J., Paul, W.M.Jr., Howard, W.M., Lindsay, A. 2018. Permutation Statistical Methods: Calculation Effciencies. Biostat Biometrics Open Acc J; 4(4): 555640. DOI: 10.19080/BBOAJ.2018.04.555640.

Kim, H.-Y. (2019) ‘Statistical notes for clinical researchers: the independent samples t -test’, Restorative *Dentistry & Endodontics*, 44(3). Available at: https://doi.org/10.5395/rde.2019.44.e26.

Kokoska, S. (2015) *Introduction Statistics, A Problem Solving Approach*. W.H. Freeman & Company.

Krishnainah, P.R. and Sen, P.. (1984) *Handbook of Statistics*. 4th edn. Elsevier Science Publishers.

Kurniasari, W., Kusnandar, D., & Sulistianingsih, E. (2019). Estipasi Parameter Regresi Spline dengan Metode Penalized Spline. *BuletinIlmiah Mat.Stat dan Terapannya (Bimaster)* , 175-184.

Kvam, P.H. and Vidakovic, B. (2007) ‘Nonparametric Statistics with Applications to Science and Engineering’, *Nonparametric Statistics with Applications to Science and Engineering*, pp. 1–429.

Lemeshko, B. Y. and Novikova, A. Y. (2018) ‘Application and Power of Tests for Homogeneity of Variances’, *2018 14th International Scientific-Technical Conference on Actual Problems of Electronic Instrument Engineering, APEIE 2018 - Proceedings*, pp. 146–152. doi: 10.1109/APEIE.2018.8545188.

Mudawanah, S. (2019) ‘Analisis Operating Leverage (DOL), Financial Leverage (DFL), Dan Combination Leverage (DCL) Terhadap Earning Per Share (EPS) Pada Perusahaan LQ45 Di Bursa Efek Indonesia’, *Jurnal Studia Akuntansi dan Bisnis*, 7(3), pp. 187–198.

Mudawanah, S. (2022) ‘Pengaruh Pertumbuhan Ekonomi Dan Upah Minimum Terhadap Tingkat Kemiskinan Di Kabupaten Dan Kota Provinsi Banten’, *Jurnal Studia Akuntansi dan Bisnis*, 10(1), pp. 61–72.

Muhid, A. (2019) *Analisis Statistik Edisi 2*, *Journal of Chemical Information and Modeling*.

Mulyono, S. (2017). *Statistika untuk Ekonomi & Bisnis.* Jakarta: Mitra Wacana Media.

Mundir (2012) *Statistik Pendidikan, Pengantar Analisis Data Untuk Penulisan Skripsi dan Tesis*, STAIN Jember Press.

Mundir, 2012. Statistik Pendidikan Pengantar Analisis Data Untuk Penulisan Skripsi dan Tesis. STAIN Jember Press, Jember.

Nugroho, S. (2008) *Statistika Nonparametrik*. 1st edn. UNIB Press.

Nuryadi *et al.* (2017) *Buku Ajar Dasar-dasar Statistik Penelitian*.

Nuryadi *et al.* (2017) *Dasar-dasar Penelitian*. Yogyakarta: Sibuku Media.

Nuryadi, Astuti, T.D., Utami, S.R., Budiantara, M., 2017. Dasar Dasar Statistik Penelitian. Sibuku Media, Yogyakarta.

Nuryadi., Astuti, T, D., Utami, E, S., Budiantara, M., 2020. Dasar-Dasar Statistik Penelitian. Sibuku Media, Yogyakarta.

Oppong, F. B. and Agbedra, S. Y. (2016) 'Assessing Univariate and Multivariate Normality, A Guide For Non-Statisticians', *Mathematical Theory and Modeling*, 6(2), pp. 26–33.

Pasaribu, B., Aji, R, H, L., Utomo, K, W., Herawati., 2021. Statistika Untuk Ekonomi dan Bisnis. Edu Pustaka, Jakarta.

Peck, R., Olsen, C., Devore, J. 2008. Introduction to Statistics and Data Analysis, Third Edition. Thomson Brooks/Cole. USA.

Purwanto, S. (2016). *Statistika untu kEkonomi dan Keuangan Modern Buku 2 Edisi 3.* Jakarta: Salemba Empat.

Riadi, E. (2014) *Metode statistika parametrik & nonparametrik.* Pustaka Mandiri.

Riadi, Edi (2014) *Metode Statistika Parametrik dan Nonparametrik Untuk Penelitian Ilmu-Ilmu Sosial dan Pendidikan.* Pustaka Mandiri

Riduwan. 2020. Dasar – Dasar Statistika. CV Alfabeta, Bandung.

Rinaldi, A., Novalia, Syazali, M., 2020. Statistika Inferensial untuk Ilmu Sosial dan Pendidikan. IPB Press, Bogor.

Ross, S, M., 2010. Introduction to Probability Model. Tenth Edition. Elsevier, California.

Ruseffendi, E.T., 2005. Dasar-Dasar Penelitian Pendidikan dan Bidang Non Eksakta Lainnya. Tarsito, Bandung.

Salmaa, 2022. Distribusi Probabilitas: Pengertian, Karakteristik, Macam, dan Contohnya. Dalam: s.l.:deepublish.

Siagian, N. (2021). Statistika Dasar: Konseptualisasi dan Aplikasi, CV. Kultura Digital Media, Surakarta.

Simamora, B. (2023, may 29). *Regresi Linier Bergnda*. Diambil kembali dari Bilson Simamora Marketing and Research Center: htpps://www.bilsonsimamora.com

Siregar, S. 2015. Statistika Terapan Untuk Perguruan Tinggi. Kencana. Jakarta

Souza, R. R. de *et al.* (2023) 'Sample size and Shapiro-Wilk test: An analysis for soybean grain yield', *European Journal of Agronomy*, 142(1000), p. 126666. doi: 10.1016/j.eja.2022.126666.

Stolp, C., Dowdy, S. and Wearden, S. (1984) *Statistics for Research*, *Journal of Policy Analysis and Management*.

Subandriyo, B. (2018) 'Statistik Non Parametrik Statistik Non Parametrik n m n', (2), pp. 86–87.

Sudjana, 1989. Metode Statistika. Edisi Ke-5. Tarsito, Bandung.

Sudjana, 2005. Metoda Statistika. Penerbit "Tarsito", Bandung.

Sugiyono, 2013. Statistika untuk Penelitian. Cetakan Ke-22. Alfabeta, Bandung.

Sugiyono. 2016. Statistika Untuk Penelitian. CV Alfabeta, Bandung.

Sukarsa, I.K.G., Kencana, I.P.E.N., 2015. Statistika Dasar. Laboratorium Statistika. Jurusan Matematika FMIPA, Universitas Udayana.

Sulistiyowati, W., Astuti, C, C., 2017. Statistika Dasar Konsep dan Aplikasinya. Umsida Press, Sidoarjo.

Supranto, 2008. Statistik: Teori dan Aplikasi. Penerbit Erlangga, Jakarta.

Supranto, J., 1994. Statistik Teori dan Aplikasi. Edisi ke-5, jilid 2. Erlanggaa, Jakarta.

Syam, R., Sanusi, W., & Adwiyah, R. (2019). Model Regresi Nonparametrik dengan Pendekatan Spline (Studi Kasus: Berat Badan Lahir Rendah di Rumah Sakit Ibu dan Anak Siti Fatimah Makasar). *Journal of Mathematics, Computations and Statistics* , 70-81.

Tabachnick, B. G. and Fidell, L. S. (2013) *Using Multivariat Statistics*, *Pearson*. Pearson. doi: 10.1037/022267.

UCEO. (2016, Mei 16). *Pengertian Korelasi dan Macam-Macam Korelasi*. Diambil kembali dari Universitas Ciputra Entrepreneurship Center: http://ucec.uc.ac.id

Usman, H., & Akbar, P. S. (2020). *Pengantar Statistika.* Jakarta: Bumi Aksara.

Wahyuning, S., 2021. Dasar-Dasar Statistik. Yayasan Prima Agus Teknik, Semarang.

Wallpole, R. E., Myers, R. H., Myers, S. L., Ye, K., 2011. Probability and Statistics for Engineers and Scientists. Pearson Education, London

Walpole, R. E. & Myers, R. H., 2012. *Probability and Statistics for Engineers and Scientists.* 9 penyunt. Boston: Pearson.

Walpole, R.E. *et al.* (2016) *Probability and Statistic for Engineers and Scientists*. 9th edn, *Jurnal Penelitian Pendidikan Guru Sekolah Dasar*. 9th edn. Prentice Hall.

Widiyanto. M. A. 2013. Statistika Terapan. Elex Media Komputindo

Widodo, A., Andawaningtyas. 2017. Pengantar Statistika. UB Press. Malang

Wirawan, N., 2016. Cara Mudah Memahami Statistika Ekonomi dan Bisnis (Statistika Deskriptif). Keraras Emas, Denpasari.

Wulandary, S., & Purnama, D. I. (2020). Perbandingan Regresi Nonparamterik Kerenl NWE Dan B-Spline Pada Pemodelan Rata-rata Lama Sekolah dan Pengeluaran Perkapita di Indonesia. *Jambura Journal of Probability and Statistics* , 89-97.

Yap, B. W. and Sim, C. H. (2011) 'Comparisons of various types of normality tests', *Journal of Statistical Computation and Simulation*, 81(12), pp. 2141–2155. doi: 10.1080/00949655.2010.520163.

Yusi, M. S., & Idris, U. (2019). *Statistika untuk Ekonomi, Bisnis dan Sosial.* Yogyakarta: Penerbit Andi.

Zach. (2020, September 14). *Koefisien Phi : Definisi dan Contoh*. Diambil kembali dari Statology: statology.org/phi-coefficient/

PROFIL PENULIS

Pandriadi, S.E., M.Si.

Ketertarikan penulis terhadap ilmu ekonomi dimulai pada tahun 1991 dimana penulis memilih untuk masuk jurusan A3 (Ilmu-ilmu Sosial) saat bersekolah di SMAN 1 Baturaja, Kabupaten Ogan Komering Ulu. Ketertarikan ini berlanjut dimana penulis pada tahun 1993 melanjutkan pendidikan strata 1 di Fakultas Ekonomi Universitas Sriwijaya Jurusan Ekonomi Pembangunan dan kemudian pada tahun 2002 kembali melanjutkan pendidikan jenjang strata 2 pada Program Studi Ilmu Ekonomi Program Pasca Sarjana Universitas Sriwijaya.

Penulis memiliki kepakaran di bidang ilmu ekonomi, meliputi ekonomi moneter, ekonomi ketenagakerjaan dan ekonomi industri. Selain aktif mengajar pada beberapa mata kuliah, seperti Metode Penelitian, Statistik, Perekonomian Indonesia dan Teori Pasar Modal.

Beberapa buku yang pernah ditulis yaitu:

1. Manajemen Keuangan 1 (Teori, Soal dan Penyelesaian)
2. Transformasi Indonesia Menuju *Cashless Society*
3. Strategi Pemasaran di Era Digital
4. Pengantar Pasar Modal Indonesia
5. Metode Penelitian Kuantitatif dan Kualitatif
6. Pengantar Ekonomi Mikro

Email Penulis: pandriadi_msi@yahoo.com

Vina N. Van Harling, S.Si., M.Pd.

Penulis lahir di Kota Ambon pada tanggal 19 Desember. Mendapatkan gelar pertamanya pada tahun 2004, dan ditahun 2010 mendapat gelar Magister Manajemen Pendidikan di Universitas Kristen Satya Wacana Salatiga. Saat ini ia tercatat sebagai dosen tetap untuk mata kuliah Statistik dan Perancangan Percobaan di Politeknik Saint Paul Sorong. Selain mengajar ia juga aktif dalam kegiatan tridarma lainnya diantaranya ialah penelitian dan pengabdian.

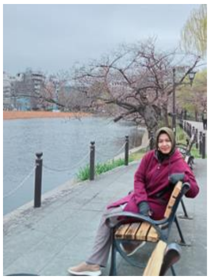
Saat ini ia pun diamanahi menjadi editorial *board* jurnal *Soscied, Electro Luceat* dari Politeknik Saint Paul Sorong, menjadi *reviewer* Jurnal dari Sekolah Tinggi Ilmu Komunikasi dan Sekretari Tarakanita. Beberapa penelitian yang berhasil didanai oleh Ristekdikti dari tahun 2017 berjudul: Analisis Hubungan Antara Motivasi Kerja, Kompetensi Dosen, Kepemimpinan, Lingkungan Kerja Dan Komitmen Profesi Terhadap Kinerja Dosen Politeknik Katolik Saint Paul Sorong (2018), Pengaruh Optimasi Transformator Daya Terhadap Perkembangan Beban Feeder Untuk Meminimalisasi Gangguan Dan Defisit Beban Listrik Di Wilayah Sorong-Papua Barat (2018), Studi Perencanaan Pembangunan PLTMH Di Kampung Sasnek Distrik Sawiat Kabupaten Sorong Selatan Provinsi Papua Barat (2017). Adapun karya buku yang telah ditulisnya sejak tahun 2020, diantaranya berjudul:

1. Pedoman Tata Tulis Usulan Penelitian dan Tugas Akhir
2. Pengaruh Optimasi Transformator Daya Terhadap Perkembangan Beban Feeder Untuk Meminimalisasi Gangguan dan Defisit Beban Listrik
3. Studi Perencanaan Pembangunan PLTMH di Kampung Sasnek Distrik Sawiat Kabupaten Sorong Selatan Provinsi Papua Barat

Dr. Abdul Wahab, S.Si., M.Si.

Penulis lahir di Majene Sulawesi Barat. Penulis menempuh pendidikan dasar pada SD tamat tahun 1991, SLTP tamat tahun 1994, dan SMU tamat tahun 1997 di Majene. Pada tahun yang sama (1997) penulis melanjutkan pendidikan ke Universitas Negeri Makassar dengan mengambil jurusan matematika dan selesai Januari 2003, selanjutnya Program Magister Statistika di Institut Teknologi Sepuluh Nopember Surabaya dan selesai Agustus 2006, kemudian Program Doktor Ilmu Pendidikan dan selesai Nopember 2017. Pada bulan Juli s.d. Nopember tahun 2018 penulis mengikuti Program Magang Dosen di Universitas Gadjah Mada Yogyakarta bidang pengelolaan PT, Pengajaran, Penelitian dan Pengabdian Kepada Masyarakat, serta Kerjasama, yang dilaksanakan oleh Direktorat Jenderal Sumber Daya IPTEK dan PT Kemenristekdikti. Pada Bulan Januari tahun 2020, penulis diterima menjadi dosen tetap pada Jurusan Tarbiyah Fakultas Agama Islam (FAI) Universitas Muslim Indonesia (UMI) Makassar. Email aktif iwahabumi@gmail.com dan abdulwahab79@umi.ac.id.

Sisca Vaulina, S.P., M.P.



Penulis lahir di Tembilahan, Kabupaten Indragiri Hilir Propinsi Riau, pada tanggal 21 Oktober 1983. Lulus Magister Pertanian pada tahun 2011 di Universitas Padjadjaran. Sejak tahun 2013, penulis berhidmat sebagai dosen tetap di Fakultas Pertanian Universitas Islam Riau. Jabatan yang pernah diemban antara lain sebagai Sekretaris Unit Penjaminan Mutu (UPM) Fakultas Pertanian Tahun 2014-2017; sebagai Ketua Program Studi Agribisnis Fakultas Pertanian Tahun 2020-2024. Menjadi anggota organisasi profesi dosen pada Asosiasi Agribisnis Indonesia (AAI), Perhimpunan Ekonomi Pertanian Indonesia (PERHEPI) Komda Pekanbaru sebagai anggota bidang kesekretariatan tahun 2014-2017, sebagai anggota bidang perilaku konsumen dan ekonomi digital tahun 2021-2024. Beberapa penelitian yang telah didanai oleh Ristek Dikti, dengan judul: Efisiensi Usahatani Kelapa Dalam (*Cocos nucifera Linn*) di Kecamatan Gaung Anak Serka Kabupaten Indragiri Hilir Propinsi Riau (2017); Analisis Faktor Yang Mempengaruhi Produksi Kelapa Dalam (*Cocos Nucifera Linn*) Pada Lahan Gambut dan Lahan Mineral Di Kabupaten Indragiri Hilir (2018). Selain itu, melalui Pusat Studi Agribisnis dan Sumberdaya Fakultas Pertanian Universitas Islam Riau kegiatan penelitian didanai oleh PT. Arara Abadi dan PT. Sinar Perawang. Melalui program Merdeka Belajar Kampus Merdeka (MBKM), penulis terdaftar sebagai Dosen Modul Nusantara Program Pertukaran Mahasiswa (2021) yang diselenggarakan oleh Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia. Kemudian, sebagai dosen pendamping Program Kreativitas Mahasiswa PKM 5 Bidang (2020) yang didanai oleh Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia. Karya buku yang pernah ditulis pada tahun 2022 mengenai Sains Teknologi Pertanian Dalam Kedaulatan Pangan.

Sri Sutjiningtyas, S.Pd., M.Kom.

Penulis lahir di kota Madiun pada tanggal 27 Januari 1970, menyelesaikan studi jenjang S1 di Institut Keguruan dan Ilmu Pengetahuan (IKIP) Bandung yang sekarang dikenal dengan nama Universitas Pendidikan Indonesia di tahun 1993 pada bidang Pendidikan Matematika. Kemudian tahun 2004 menyelesaikan jenjang S2 bidang Ilmu Komputer di Universitas Gadjah Mada Yogyakarta, hingga mendapat gelar Magister Ilmu Komputer.

Penulis menjadi staf dosen sejak tahun 1994 dan saat ini tercatat sebagai dosen tetap di program studi Teknik Informatika Universitas Nurtanio Bandung.

Selain mengajar, penulis aktif membimbing tugas akhir mahasiswa, penelitian dan pengabdian kepada masyarakat. Matakuliah yang diajarkan adalah matakuliah terkait matematika yang relevan dengan bidang ilmu komputer atau informatika, yaitu meliputi matematika dasar, matematika diskrit, aljabar linier, logika matematika dan informatika, metode numerik, statistika dan probabilitas.

Dra. Endang Kusdiah Ningsih, M.Si.

Penulis lulus S1 pada Program Studi Ilmu Ekonomi dan Studi Pembangunan (IESP) Fakultas Ekonomi Universitas Sriwijaya Palembang pada tahun 1981, lulus S2 pada program Pengembangan Wilayah dan Desa (PWD) Universitas Andalas tahun 1994. Saat ini tercatat sebagai dosen tetap yayasan pada Fakultas Ekonomi Universitas IBA Palembang, dan menjadi dosen tamu pada perguruan tinggi lain di kota Palembang. Mata kuliah yang diampu adalah Pengantar Ekonomi Mikro, Pengatar Ekonomi Makro, Matematika Ekonomi, Statistika dan Perekonomian Indonesia. Aktif menulis artikel di beberapa jurnal bidang ekonomi terindeks Sinta, dan pernah menerbitkan satu buku yang berjudul "Perkembangan Inflasi di Indonesia.

Bagus Dwi Hari Setyono, S.Pi., M.P.

Penulis lahir di Kota Mataram pada tanggal 3 Agustus 1984. Ia lulus pada tahun 2008 hingga mendapat gelar Magister Perikanan di Universitas Brawijaya. Saat ini ia tercatat sebagai dosen tetap pada Program Studi Budidaya Perairan, Jurusan Perikanan dan Ilmu Kelautan, Fakultas Pertanian, Universitas Mataram. Selain mengajar, ia aktif dalam kegiatan tridharma Perguruan Tinggi lainnya, yaitu penelitian dan pengabdian kepada Masyarakat. Saat ini ia diamanahi sebagai Ketua Program Studi D3 Budidaya Perikanan Universitas Mataram (PDD) di Kabupaten Lombok Utara, serta menjadi *reviewer* di Jurnal *Grouper*, Universitas Islam Lamongan. Adapun karya buku yang telah ditulisnya sejak tahun 2023, diantaranya berjudul: *Akuaponic for Urban Farming*: "Mewujudkan Petani Inovatif 5.0", Kiat Agribisnis Rumput Laut, Pengantar Ekonomi Mikro, dan Imunologi.

Vini Rizqi, S.Pd., M.P.Mat.

Penulis lahir di Kota Bandung pada tanggal 12 September. Merupakan lulusan Universitas Pendidikan Indonesia jurusan Pengajaran Matematika, kemudian melanjutkan studi di Institut Teknologi Bandung jurusan Magister Pengajaran Matematika. Sampai saat ini bekerja sebagai dosen di Universitas Nurtanio Bandung untuk matakuliah Matematika Ekonomi dan Statistika. Aktivitas lainnya adalah sebagai penggiat salah satu organisasi pemikiran Islam di Bandung. Beberapa karya tulis telah diterbitkan oleh beberapa penerbit serta beberapa penelitian yang telah diterbitkan oleh beberapa jurnal, baik dari jurnal nasional hingga jurnal terakreditasi Sinta.

Muhammad Iqbal Harisuddin, S.T., M.Pd.

Penulis lahir di Subang, 19 Agustus 1981. Menyelesaikan Program S1 Teknik Kimia di ITENAS Bandung tahun 2005 dan program S2 Magister Pendidikan Matematika di Universitas Pasundan tahun 2014. Merupakan penulis aktif dan sudah menerbitkan buku; **Asyiknya Belajar Matematika dengan Geogebra** diterbitkan Deepublish tahun 2019; **Secuil Esensi Berpikir Kreatif dan Motivasi Belajar Siswa** diterbitkan Pantera Publishing tahun 2019; **Berpikir Kreatif**

Motivasi Kemandirian Belajar Siswa diterbitkan Deepublish tahun 2023. Hasil penelitian penulis yang dipublish di Jurnal diantaranya; Kemampuan Pemecahan Masalah Matematis dan Kemandirian Belajar Siswa dengan PJJ Dimasa Covid-19, Jurnal Teorema: teori dan riset matematika tahun 2021; *Implementation of Accelerated Learning to Improve Mathematics Communication Ability*, Jurnal Pendidikan MIPA tahun 2022. Profesi sekarang penulis sebagai Dosen Pendidikan Matematika Universitas Mandiri Subang. Sebagai dosen, penulis mengampu mata kuliah yaitu; Geometri Analitik; Aplikasi Matematika, Sains, Teknologi dan Rekayasa; Statistika Dasar; Program Linear; Persamaan Differensial; Geometri ; dan Metode Numerik.

Syamsidar Gaffar, S.Kel., M.Si.

Penulis lahir di Kota Kendari pada tanggal 15 Mei 1989. Ia Lulus sarjana dari Universitas Hasanuddin dan magister dari Institut Pertanian Bogor dari Program Studi Ilmu Kelautan. Sejak 2019, ia tercatat sebagai dosen di Universitas Borneo Tarakan, yang sebelumnya dari 2016 merupakan dosen tetap di Universitas Haluoleo. Selain mengajar, ia aktif dalam kegiatan tridharma lainnya, seperti penelitian dan pengabdian. Beberapa penelitian yang berhasil didanai oleh Ristekdikti, berjudul: Konservasi Sumberdaya Biota Perairan Lokal melalui DNA *Barcoding* Spons (Porifera) Padang Lamun pada Zona Intertidal Perairan Pulau Derawan dan Pulau Panjang, Kalimantan Timur (2022), Eksplorasi Potensi Bakteri Simbion Spons dari Perairan Pulau Derawan sebagai Antibakteri *Multidrug Resistent* (2022-2023), dan DNA *Barcoding* Ikan Pari Air Tawar (*Elasmobranchii*) dari Sungai Sesayap, Malinau, Kalimantan Utara (2023). Selain itu, penelitian yang didanai oleh LPDP melalui skim Riset Keilmuan, berjudul DNA *Barcoding* dan Analisis Filogenetik Kerang Kapah sebagai Upaya Konservasi Genetik Biota Laut Lokal di Pesisir Pulau Tarakan, Kalimantan Utara (2022). Adapun karya buku referensi yang telah ditulisnya sejak tahun 2023, berjudul:

1. Dinamika Perairan Laut Dangkal,
2. Mikrobiologi Lingkungan, dan
3. Air Bersih Gratis.

Trisna Yuniarti, M.T.

Penulis lahir di Kota Palembang pada tanggal 07 Juni 1983. Penulis menyelesaikan pendidikan Sarjana Teknik Industri dan Magister Teknik di Universitas Indonesia. Berprofesi sebagai PNS sejak tahun 2010 di Kementerian Perindustrian R.I. Saat ini, penulis bertugas sebagai dosen tetap pada Program Studi Manajemen Logistik Industri Elektronika, Politeknik APP Jakarta. Mengampu mata kuliah Statistik Industri, Manajemen Inventori, Dasar Industri 4.0 dan K3. Keahlian dan bidang yang diminatinya adalah *Quality Engineering*, *Service Management and Engineering*, *Applied Statistics*, *Big Data/Data Mining*. Adapun karya buku yang pernah ditulis berjudul Dasar Industri 4.0. dan Sekilas Manajemen Logistik.

Anisa Rahmawati, S.Si., M.Sc.

Penulis lahir pada tanggal 7 November 1994 di Kabupaten Klaten, Jawa Tengah. Ia menempuh pendidikan sarjana dan pasca sarjana di Prodi Matematika, Universitas Gadjah Mada. Ia menyelesaikan studi sarjana pada tahun 2017 dan menyelesaikan studi pasca sarjana pada tahun 2019. Saat ini, ia menjadi dosen tetap di salah satu perguruan tinggi milik Kementrian Perindustrian yaitu Politeknik APP Jakarta. Selain mengajar ia mulai aktif dalam kegiatan tridarma lainnya, salah satunya adalah menulis buku yang terkait dengan bidang keilmuannya. Buku Statistika Dasar ini merupakan buku pertama yang ia tulis. Besar harapan untuk kedepannya ia dapat menulis buku lainnya.

Firdhani Faujiyah, S.T., M.T.

Penulis lahir di Kota Bandung pada Bulan November 1990. Ia menyelesaikan pendidikan Sarjana Teknik Industri di Institut Teknologi Nasional Bandung, dan pendidikan Magister Teknik dan Manajemen Industri di Institut Teknologi Bandung. Berprofesi sebagai dosen semenjak tahun 2014 di Kota Bandung. Saat ini ia tergabung sebagai dosen tetap untuk mata kuliah Statistika Pemasaran pada salah satu unit pendidikan di bawah Kementerian Perindustrian di Kota Jakarta. Selain mengajar, ia aktif dalam kegiatan tridarma lainnya yaitu penelitian dan pengabdian masyarakat dengan minat keahlian di bidang perancangan sistem kerja dan ergonomi, serta perancangan manufaktur. Sebagai akademisi saat ini ia juga diamanahi sebagai *reviewer* Jurnal Politeknik TEDC untuk bidang keilmuan keteknikan. Untuk mendukung pengajaran, penelitian dan pengabdian yang digiatinya, ia tersertifikasi dalam bidang keahlian *data science* dan juga gambar teknik.

Siti Mudawanah, S.E., M.Akt.

Penulis lahir di Kabupaten Lebak Banten, Putri ketujuh dari delapan bersaudara dari Bapak M. Soleh dan Ibu Eni. Menyelesaikan Pendidikan dari SD s.d S1 di Kota Rangkasbitung Kabupaten Lebak Banten. Melanjutkan Pendidikan S2 Akuntansi di Universitas Budi Luhur. Kegiatan sehari-hari mengajar di program S1 Universitas La Tansa Mashiro. Saat ini menjadi Dosen Tetap Yayasan sekaligus menjabat sebagai staf Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM) Universitas La Tansa Mashiro. Menjadi tim editor di jurnal *The Asia Pacific Journal Of Management Studies* dan *editorial board* di *Basic and Applied Management Research Journal (BAMRJ)*. Penelitian yang berhasil di danai Kemenristekdikti Tahun 2021 Penelitian Dosen Pemula dengan judul implikasi Biaya Penyelenggaraan Pendidikan dan Latar Belakang Sosial Ekonomi Terhadap Minat Calon Mahasiswa Program Studi Akuntansi STIE La Tansa Mashiro dalam Keberlanjutan Perguruan Tinggi. Adapun karya buku tahun 2023 yang telah di tuliskan yaitu pengantar akuntansi Bab 11. Interpretasi Laporan Keuangan.

PROFIL EDITOR

Sary Shandy, S.T., M.T.

Penulis lahir di Kota Ambon pada tanggal 27 September 1987. Ia menyelesaikan Pendidikan Sarjana Teknik di Universitas Khairun dan Magister Teknik di Universitas Hasanuddin. Berprofesi sebagai dosen sejak tahun 2015 di Prodi Teknik Sipil Universitas Muhammadiyah Maluku Utara hingga tahun 2020. Saat ini menjadi dosen tetap di Prodi Teknik Sipil Fakultas Teknik Universitas Khairun Ternate. Selain mengajar, penulis aktif membimbing tugas akhir mahasiswa, penelitian dan pengabdian kepada Masyarakat ia juga merupakan anggota Satuan Pengawasan Intern (SPI) Universitas Khairun. Matakuliah yang diajarkan adalah Statistika, Struktur Baja, Analisa Struktur, dan Perencanaan Struktur Baja.

Dr. Yanneri Elfa Kiswara Rahmantya

Penulis merupakan dosen Program Pascasarjana Universitas Kuningan, yang lahir di Malang pada tahun 1987. Penelitian dan kepakarannya berfokus pada pengembangan kemampuan organisasi dan meningkatkan efektivitas manajemen strategis melalui program *research and development*, editor memiliki latar belakang pendidikan serta pengalaman di bidang keuangan, pernah bekerja sebagai analis keuangan dan konsultan pengembangan usaha kecil sebelum memperoleh gelar doktor. Beliau meraih gelar doktor di bidang manajemen strategis dari Universitas Brawijaya. Penelitian-penelitian yang telah dilakukan merupakan penelitian kuantitatif dengan menggunakan berbagai metode dan analisis statistik.

STATISTIKA DASAR

Statistika adalah ilmu yang membahas (meneliti) dan mengembangkan prinsip, metode, dan prosedur yang harus diikuti atau digunakan dalam pengumpulan data digital, penyusunan atau pengelolaan data digital, penyajian, representasi atau deskripsi data numerik, analisis data numerik, membuat kesimpulan, memperkirakan, serta menyiapkan prediksi ilmiah/matematis berdasarkan kumpulan data numerik tersebut.

Statistika memegang peranan penting dalam kehidupan manusia dan perkembangan ilmu pengetahuan. Oleh karena itu, tidak heran jika saat ini statistika telah berkembang dalam berbagai bidang keilmuan, misalnya *tatistic* ekonomi, *tatistic* bisnis, *tatistic* pendidikan, *tatistic* sosial, *tatistic* kedokteran, dan lain-lain. Demikian pula dalam dunia pendidikan, statistika digunakan sebagai "alat" untuk memberikan gambaran suatu fenomena, sebagai ujian pengambilan keputusan.

Penerbitan buku ini dimaksudkan untuk membekali mahasiswa dengan pengetahuan statistika yang dipandang perlu dan relevan bagi para peneliti, pendidik, dan penyelenggara pendidikan.

ISBN 978-623-459-752-3

9 786234 597523

Penerbit widina
www.penerbitwidina.com