Pustaka Galeri Mandiri

STATISTIKA PARAMETRIK DAN NON PARAMETRIK

(dilengkapi dengan Aplikasi SPSS dan Excel)

Syafiadi Rizki Abdila, S.T., Ph.D

Prof. Dr. Ir. Drs. Syafwandi, M.Sc.

Dr. Yulasmi, S.E., M.M

Dr. Robby Dharma, S.E., M.M

Dr. Syafri Arlis, S.Kom, M.Kom

STATISTIKA PARAMETRIK DAN NON PARAMETRIK

(dilengkapi dengan Aplikasi SPSS dan Excel)

Syafiadi Rizki Abdila, S.T., Ph.D
Prof. Dr. Ir. Drs. Syafwandi, M.Sc.
Dr. Yulasmi, S.E., M.M
Dr. Robby Dharma, S.E., M.M
Dr. Syafri Arlis, S.Kom, M.Kom

Penerbit

Pustaka
Galeri Mandiri

STATISTIKA PARAMETRIK DAN NON PARAMETRIK

(dilengkapi dengan Aplikasi SPSS dan Excel)

Penulis : Syafiadi Rizki Abdila, S.T., Ph.D
Prof. Dr. Ir. Drs. Syafwandi, M.Sc.
Dr. Yulasmi, S.E., M.M
Dr. Robby Dharma, S.E., M.M
Dr. Syafri Arlis, S.Kom, M.Kom
Design Cover : by rawpixel.com on Freepik
Layout Isi : Tim Kreatif Penerbit

ISBN : 978-623-8164-31-8

Diterbitkan oleh :
Pustaka Galeri Mandiri
Perum Batu Kasek E11, Jl. Batu Kasek, Pagambiran Ampalu Nan XX
Lubuk Begalung, Padang. SUMBAR. 25226
e-mail : pgm@pustakagalerimandiri.co.id
homepage : pustakagalerimandiri.co.id
fansfage FB : Pustaka Galeri Mandiri, Instagram : @pustakagaleri
Youtube : pustaka galeri mandiri
Jurnal Ilmiah : http://jurnal.pustakagalerimandiri.co.id

PRAKATA

Puji syukur ke hadirat Allah SWT, atas terselesaikannya buku ajar Statistika yang merupakan rangkuman modul dari berbagai referensi.Diharapkan, bukuini mudah dipelajari untuk memahami ilmu statistik dan dapat digunakan sebagai alat bantu menganalisis secara statistic bagi para mahasiswa pada berbagai program studi yang membutuhkannya.

Materi yang dikandung Buku Ajar Statistik ini terdiri atas statistik deskriptif, teori probabilitas, analisis korelasi,dan uji hipotesis parametrik serta uji hipotesis non parametrik; mengingat materi tersebut banyak digunakan untuk analisis statistic pada penyusunan tugas akhir mahasiswa.

Penulis menyampaikan terimakasih kepada pihak manapun yang telah memberi saran positif untuk lebih menyempurnakan buku ini. Namun demikian, penulis menyadari bahwa masih banyak kekurangannya.Untuk itu, penulis sangat terbuka menerima kritik dan saran untuk meminimalisir kekurangan-kekurangan tersebut. Banyak terimakasih.

Harapan penulis,semoga buku ini bermanfaat bagi pembacanya.

Medio Juli 2022.

DAFTAR ISI

BAB 1

PENDAHULUAN

Statistika adalah sebuah cabang ilmu yang merupakan derivasi dari ilmu matematika dengan melibatkan berbagai kesepakatan khususnya pada statistika inferensial. Formula-formula umum yang digunakan di dalamnya tetap berdasar kepada kaidah matematika; namun pada tahapan keputusan hasil analisis, pertimbangan kesepakatan itu diberlakukan.

Kemampuan Akhir yang Diharapkan::
- Mahasiswa mampu memahami perkembangan ilmu statistika, - Mampu memahami tujuan ilmu statistika.

1.1 Pendahuluan

Statistik, pada awal abad ke-18 adalah sebuah kata yang berasal dari kata *state*. Kata itu digunakan sampai sekarang dengan arti sebuah rangkaian data numeric yang lebih berkaitan dengan pajak, populasi penduduk, kesejahteraan masyarakat, produksi, panen, ekspor, impor danbanyak hallainyang menjadi perhatian danperlu diketahui pemerintah; untuk tujuan pengendalian dan penyusunan kebijakan publik. Statistika merupakan sebuah ilmu yang diturunkan dari matematika, dengan memasukkan berbagai kesepakatan.

Definisi stastistika saat sekarang telah berkembang menjadi sebuah ilmu untuk mengkoleksi dan mengorganisasi data, analisis untuk pembuatan keputusan, permodelan data, interpretasi dan mempresentasi data[1]. Statistik dibedakan dengan statistika. Yang dimaksud dengan serangkai data numerik (statistik) itu telah meluas, seperti statistik pasar saham, harga jual dolar, penjualan, tingkat suku bunga, statistik sepak bola, statistik pendidikan, statistik kesehatan, statistic buruh, rata-rata upah buruh per jam, statistik produksi, statistik lalulintas, statistik pertanian, dan lain-lain; namun bukan data angka abstrak seperti data pada tabel akar kuadrat.

Pada perguruan tinggi, statistika digunakan untuk menunjang penelitian-penelitian ilmiah yang berbasis kuantitatif, seperti : pengukuran nilai pusat (*tendency central value*) dan penyebaran data (*dispersion*), pengukuran probabilitas, uji pola hubungan antar variabel, uji hipotesis, analisis regresi, dan bahkan sekarang ini telah berkembang analisis multivariat. Pada dasarnya ilmu statistika dapat dikatagorikan menjadi dua katagori utama, yaitu : (i) statistika deskriptif, dan (ii) statistika inferensial. Statistika deskriptif berkaitan dengan pengukuran kecenderungan nilai pusat dan penyebarannya, termasuk di dalamnya pembahasan tentang jenis rata-rata dan penyebarannya (varians) dan grafik. Statistika inferensial berkaitan dengan uji hipotesis.

Secara umum, manfaat mempelajari statistikaadalah paling tidak untuk memahami data dengan lebih baik, cara menangani masalah data, dan menjadi yakin untuk tidak tertipu oleh pernyataan tidak benar yang secara ilmiah tidak dapat dipertanggung jawabkan. Pemahaman statistik membuat

[1]Doane dan kawan-kawan,2007:3

perusahaan memiliki kemampuan berkompetisi misal dalam pemasaran, memiliki informasi mengenai kekuatan dan kelemahan internal perusahaan. Beberapa alas an tentang perlunya mempelajari statistika, antara lain karena bermanfaat untuk: (1) komunikasi; (2) keterampilan komputer; (3) manajemen informasi; (4) pengembangan karir; (5) pengendalian dan peningkatan kualitas; (6) pemasaran; (7) pembelian; (8) perawatan kesehatan; (9) teknik peramalan;dan (10) *auditing*.

1.2 Pokok Bahasan Dalam Statistika

Sebagai ilmu, statistika adalah sebuah bidang studi yang terbagi atas empat bagian utama pembahasan, yaitu: (1) statistika deskriptif; (2) probabilitas; (3) analisis keputusan; dan (4) statistika inferensial.

(1) Statistik Deskriptif.

Statistik Deskriptif merupakan bagian ilmu statistika yang menjelaskan karakteristik serangkaian data, misal nilai kecenderungan pemusatan data (*tendency central value*), dan penyebarannya (*variability*). Pola data ini kemudian dapat dibentuk menjadi tabel distribusi frekuensi dan grafik, berikut kemencengan dan keruncingan kurve distribusi frekuensi.

(2) Probabilitas.

Sebuah probabilitas merupakan ukuran frekuensi relatif terjadinya sebuah peristiwa pada tingkat keyakinan tertentu atau dalam waktu panjang. Peluang munculnya data itu kemudian disebut sebagai probabilitas, misal probabilitas atau peluang seseorang memperoleh gambar kepala pada pelemparan koin mata uang logam, peluang perusahaan meraih penjualan lebih 50.000 unit/tahun, peluang lahan pertanian menghasilkan jagung minimal 5.000 kg/musim.

(3) Statitika untuk analisis keputusan.

Statistik yang digunakan untuk analisis pembuatan keputusan, misal: keputusan pengusaha dalam menginvestasikan dana sebesar $ 10,000.00 pada agribisnis tertentu jika peluang untuk memperoleh hasil (*return*) minimal $ 2,500.00 pada waktu mendatang adalah 75,00%; dan jika peluangnya lebih kecil dari itu, maka pengusaha itu memutuskan untuk tidak menginvestasikan dananya. Materi statistik untuk analisis pembuatan keputusan, banyak digunakan untuk mendukung matakuliah Teori Pembuatan Keputusan (TPK).

(4) Statistika Inferensial.

Statistika inferensial membahas permasalahan dengan sejauh mana sampel random yang diamati merefleksikan populasinya. Sampel random menjadi penting, karena dapat menentukan peluang kebenaran kesimpulan inferensial mengenai polulasi. Merupakan statistik yang digunakan untuk menguji hipotesis. Penggunaan statistik inferensial berkaitan dengan populasi yang diamati, sampel dan teknik pengambilan sampel, dan alat uji hipotesis. Contoh: sebuah perusahaan pengalengan ikan ingin mengetahui tingkat kerusakan produknya dengan mengamati sampel produk yang dipilih secara random. Jika perusahaan menemukan bahwa dari sampel yang diamati itu sebesar 5,00% ternyata rusak; maka mereka tidak serta percaya bahwa tingkat kerusakan populasi hasil produksinya itu benar-benar 5,00%. Perusahaan harus lebih lanjut menganalisis hipotesis untuk membuktikan bahwa peluang kerusakan populasi hasil produksinya antara 3,00% sampai 8,00% itu sebesar 95,00%..Ini merupakan bidang bahasan pada statistika inferensial. Secara statistik, metode uji hipotesis terbagi atas dua, yaitu : (i) uji hipotesis parametrik dan (ii) uji hipotesis non parametrik; di mana perbedaan kedua metode uji terletak pada penggunaan parameter data atau tidak.

1.3. Rangkuman

Definisi stastistika saat sekarang telah berkembang menjadi sebuah ilmu untuk mengkoleksi dan mengorganisasi data, analisis untuk pembuatan keputusan, permodelan data, interpretasi dan mempresentasi data. Penekanan kandungan materi (*emphases*) statistika sangat bervariasi tergantung bidang ilmu lainnya yang didukungnya, artinya pada bidang ilmu fisika penekanan materi pembahasan sedikit berbeda dengan pada bidang ilmu biologi. Namun demikian, ada suatu materi pokok yang umum, seperti nilai kecenderungan pemusatan data (*tendency.central value*) dan dispersi. Pada bidang ilmu manajemen, penekanan lebih dutamakan untuk analsis pembuatan keputusan dan statsitika inferensial; sedang pada bidang ilmu pertanian/peternakan lebih diutamakan rancangan percobaan (*experimental design*).

1.4 Diskusi

Statistik inferensial dapat berbentuk uji hipotesis parametrik dan non parametrik, keduanya menghasilkan kesimpulan hasil uji yang sama. Pilihan metode bisa bersifat subyektif peneliti. Metode uji non parametrik lebih disukai karena prosedurnya lebih sederhana dibanding metode uji parametrik, dan relatif tidak membutuhkan ukuran sampel yang besar. Saat sekarang ini, telah bermunculan pengembang baru untuk uji non parametrik dan diyakini akan masih berkembang lagi. Pertanyaannya, apakah statsitik inferensial parametrik akan ditinggal para analis? Tetap saja dibutuhkan, karena metode parametrik tetap menjadi dasar pengembangan metode non parametrik.

1.5 Referensi

Agresti, Alan, ***Catagorical Data Analysis***, 2002, John Wiley and Son Coy., Tokyo.

Bowen, Earl, K., and Martin K. Starr, 2002, ***Basic Statistics for Business and Economics*,** McGraw Hill Book Company, Tokyo.

Doane, David, P., and Lori E. Seward, 2007, ***Applied Statistics in Business and Economics***, McGraw Hill/Irwin Series, New York.

Hoel, P., ***Introduction to Mathematical Statistics***, 2005, 7th Ed. John Wiley & Sons, Inc., New York.

1.6 Latihan Soal

a. Apakah perbedaan kata statistik dengan statistika ?
b. Jelaskan perkembangan statistika sejak terminologi itu muncul sampai saat sekarang.
c. Apa tujuan mempelajari statistika sebagai ilmu ?
d. Sebutkan empat pembahasan utama dalam ilmu statistika.
e. Berikan contoh kasus yang dibahas dalam statistika inferensial.

BAB 2

DATA DAN SAMPEL

Data tunggal atau jamak, berbentuk angka numerik atau non numerik. Dalam statistika, data merupakan input dasar yang akan dianalisis. Sampel merupakan bagian kecil dari populasi, dan diharapkan dapat merepresentasi populasi yang bersangkutan.

Kemampuan Akhir yang Diharapkan:
- Mampu memahami klasifikasi data berdasar tiga aspek (karakteristik, sumber perolehan dan waku perolehan), - Mampu memahami dan membedakan macam-macam skala pengukuran (nominal, ordinal, interval dan ratio). - Mampu memilih metode pengambilan dan penentuan ukuran sampel.

2.1 Pendahuluan

Data berasal dari kata Latin, *datum*.Data statistik adalah ukuran, informasi, nilai, atau karakter yang dapat menjelaskan suatu obyek yang diamati. Sedang variabel adalah suatu obyek yang nilainya berubah-ubah sesuai dengan perbedaan waktu, tempat, situasi dan kondisi; sehingga variabel itu oleh beberapa ahli juga disebut sebagai “peubah”. Contoh, harga pakan ternak;

di mana harga itu akan berbeda atau berubah sesuai dengan perbedaan waktu, tempat dan kondisi. Nilai variabel adalah data kuantitatif yang terukur. Di lain sisi, ada data kualitatif, di mana nilainya tidak terukur secara numerik; contoh: merk penjualan komputer, seperti McIntosh, IBM, Compaq, Apple, Toshiba, dan Lenovo.

Dalam penelitian ilmiah, data dibangun dari eksperimen yang dicatat dengan sistimatis. Dalam bisnis, biasanya berasal dari transaksi akuntansi atau proses manajemen (contoh: persediaan, penjualan, daftar gaji).

2.2 Syarat Data

Untuk memperoleh data yang benar, data harus *reliabel* (dapat diandalkan), harus memenuhi tiga syarat, yaitu: (i) obyektif; (ii) representatif; dan (iii) akurat.

Obyektif, maksudnya data harus sesuai dengan keadaan yang sebenarnya. Contoh, untuk mengukur tinggi beberapa orang, harus menggunakan alat ukur yang sama.

Representatif, maksudnya data harus mewakili obyek yang yang diteliti, seperti: total produksi tebu pada suatu area, harus dihitung pada seluruh plot dalam area, bukan hanya pada lahan yang subur saja.

Akurat, maksudnya data harus teliti, dengan tingkat kesalahan minimum. Contoh, untuk mengukur ketebalan logam, harus menggunakan alat dengan ketelitian skala ukur yang detil.

Namun dengan teknologi yang tinggi, ketiga syarat data ini hampir pasti telah terpenuhi. Komputer ukur yang detil telah banyak digunakan pada berbagai bidang.

2.3. Klasifikasi Data

Klasifikasi data dapat dirinci berdasar tiga aspek, yaitu: (i) menurut karakteristik data; (ii) menurut sumber perolehannya; dan (iii) menurut waktu pengumpulannya.

Menurut Karakteristik Data

Menurut sifat dan karakter, data dapat diklasifikasi menjadi empat peringkat data, yaitu: nominal, ordinal, interval dan ratio.

(a) Data Nominal.

Data nominal adalah data dengan peringkat yang terendah, yang hanya berfungsi untuk membedakan saja, dan belum bisa diperingkatkan. Contoh: pria = 1, wanita = 0. Nilai ini hanya membedakan, sebab tidak berarti bahwa 1 lebih besar daripada 0. Data nominal tidak dapat dirata-ratakan, dijumlahkan, dikurangkan atau diperkalikan/dibagikan.

(b) Data Ordinal.

Data ordinal adalah data dengan peringkat yang lebih tinggi dibanding data nominal, sebab selain bisa membedakan (karakter nominal) telah ada fungsi peringkat dan jarak, walaupun peringkat yang berurutan itu tidak harus memiliki selisih jarak yang sama. Contoh:

Nilai Ujian Statistik	Peringkat	Selisih/Jarak
A = 80	2	1 □ 2 = 5
B = 85	1	2 □ 3 = 20
C = 60	4	3 □ 4 = 10
D = 70	3	4 □ 5 = 10
E = 50	5	

Dengan demikian, tidak dapat dinyatakan bahwa mahasiswa dengan peringkat 1 lebih pandai 5 poin daripada mahasiswa dengan peringkat 2. Jarak antar peringkat berurutan tidak sama. Data ordinal tidak dapat dirata-ratakan, dijumlahkan, dikurangkan atau diperkalikan/dibagikan.

(c) Data Interval.

Data interval adalah data yang memiliki fungsi nominal dan ordinal, dan jarak antar peringkat memiliki nilai yang sama, walau nilai peringkat itu tidak bisa menjelaskan berapa kali lebih tinggi/rendah. Contoh: Skala Likert dengan rentang skor 1 – 5.

Sangat Puas	= 5
Puas	= 4
Ragu-ragu	= 3
Tidak Puas	= 2
Sangat Tidak Puas	= 1

Jarak antar peringkat = 1 poin. Artinya, selisih peringkat orang sangat puas dengan orang yang puas = 1 poin, sama besar dengan selisih orang yang sangat tidak puas dengan orang yang tidak puas. Namun demikian, tidak berarti bahwa orang yang puas (=4) itu dua kali tingkat kepuasannya dibanding orang yang tidak puas (=2). Data interval tidak dapat dirata-ratakan, dijumlahkan, dikurangkan atau diperkalikan/ dibagikan; namun setidaknya telah bisa dihitung frekuensi kemunculannya melalui distribusi nilainya untuk mengetahui modusnya. Pada data interval, tidak memungkinkan untuk memiliki nilai = 0.

(d) Data Ratio.

Data ratio adalah data yang memiliki fungsi nominal, ordinal, dan interval. Jarak antar peringkat memiliki nilai yang sama, dan nilai peringkat itu bisa menjelaskan berapa kali tinggi/rendah antar nilai. Pada data ratio memungkinkan untuk memiliki nilai = 0.

Contoh:

Petani dengan penghasilan = Rp 4 juta/bulan, lebih tinggi dua kali dibanding penghasilan nelayan yang hanya Rp 2 juta/bulan. Luas lahan persawahan 3 ha = dua kali lebih luas daripada lahan persawahan dengan luas 1,5 ha.

Menurut Sumber Perolehan

Menurut sumber perolehannya, data diklasifikasikan sebagai: (i) data primer; dan (ii) data sekunder.

(a) Data Primer

Data primer adalah data yang diperoleh secara langsung dari obyeknya, kemudian diolah sendiri. Contoh: seseorang yang ingin mengetahui rata-rata harga hasil tangkapan ikan laut di pasar pada suatu hari, harus pergi ke pasar ikan, mendata harga secara langsung kepada penjual ikan, kemudian menghitung rata-rata harga.

(b) Data Sekunder

Data sekunder adalah data yang diperoleh secara tidak langsung dari obyeknya, tetapi diperoleh dari pihak lain yang memiliki data. Contoh: seseorang yang ingin mengetahui pertumbuhan penduduk di sebuah kabupaten pada tahun tertentu, harus pergi ke Dinas Kependudukan, dan menerima data yang sudah terolah.

Menurut Waktu Pengumpulan

Berdasar waktu pengumpulan, data diklasifikasikan menjadi: (i) data rangkai waktu (*time series*); dan (ii) data seksi silang (*cross section*).

(a) Data *time series*.

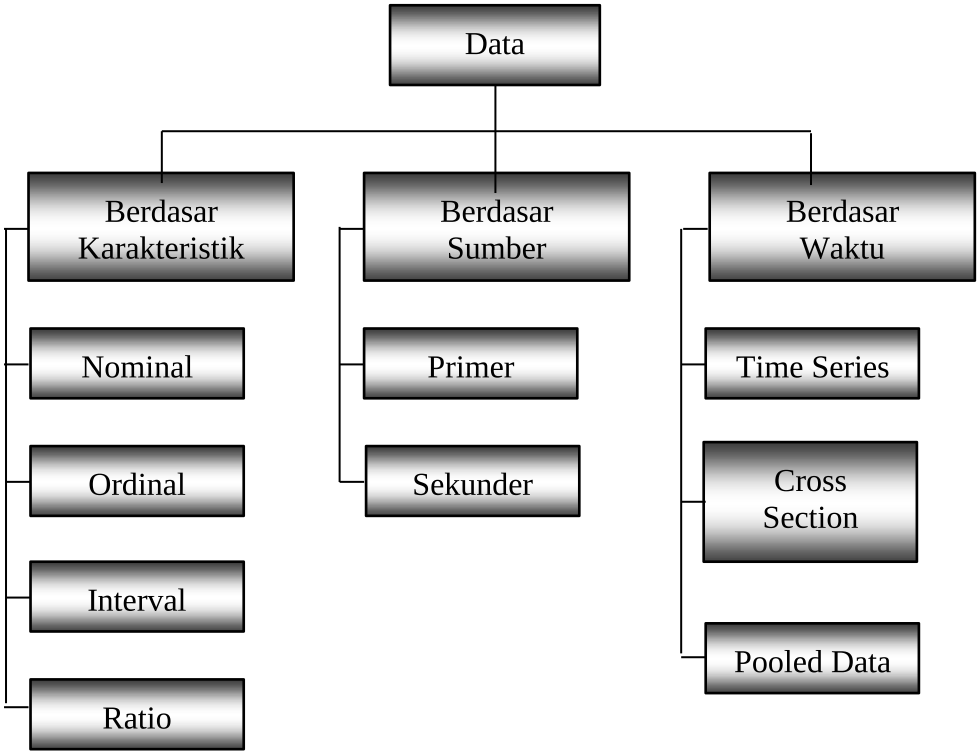
Data *time series* adalah data yang menjelaskan sebuah kegiatan/ kejadian tertentu dari waktu ke waktu. Contoh: data populasi ternak domba per tahun pada daerah tertentu dari tahun 2010 – 2016. Tujuannya adalah untuk mengamati pertumbuhan populasi ternak domba. Dengan kata lain, jika pengamatan sampel merepresentasi perbedaan antar waktu (tahun, bulan, minggu atau hari), maka datanya merupakan data rangkai waktu, contoh: GDP, tingkat konsumsi, indeks harga. Dari data jenis ini, seseorang bisa membuat pola data atau garis trend-nya

(b) Data *cross section*.

Data *cross section* adalah data yang menjelaskan beberapa obyek pada tahun tertentu. Contoh: data populasi ternak kambing, domba, ayam dan kelinci pada suatu daerah dalam tahun 2017. Data penjualan saham dari beberapa perusahaan emiten pada suatu hari tertentu.

Untuk kepentingan analisis tertentu, seorang peneliti seringkali menggabungkan kedua jenis data ini menjadi data gabungan (*pooled data*); artinya selain ada rangkai waktunya, juga ada seksi silangnya. Penggabungan data semacam ini, harus ada alasan yang logis untuk dilakukan, biasanya karena banyaknya pengamatan yang terlampau sedikit. Contoh: populasi ternak domba pada beberapa daerah dalam beberapa tahun berurutan.

Secara grafis, rincian klasifikasi data tersebut di atas dapat digambarkan sebagai berikut:



Gambar 2.1 Klasifikasi Data

YYBeberapa ahli statistik lain, memandang klasifikasi data dari sisi yang lebih sederhana; yaitu: (i) data atribut; dan (ii) data numerik[2].

(a) Data atribut.

(1) Data label/verbal disebut juga sebagai data katagorial, nominal atau data kualitatif; di mana

[2]Bowen and Starr;2002:33

nilainya dijelaskan dengan kata-kata. Contoh: sistim transmisi mobil = manual, otomatis.

(2) Data kode menggunakan angka untuk menjelaskan katagori, contoh: jenis film, 1 = *action*, 2 = klasik, 3 = komedi, 4 = horor. Angka-angka itu bukan numerik, sehingga tidak dapat diartikan sebagai peringkat, namun kadang- kadang ada kode yang juga menjelaskan peringkat, contoh:1 = SD, 2 = SMP, 3 = SMA, 4 = Sarjana,5 = Magister,dan 6 = Doktor.

(3) Data binari menggunakan hanya dua nilai yang mengindikasikan adanya karakteristik (= 1) dan tidak adanya karakteristik (= 0). Contoh:1 = pria, 2 = wanita.

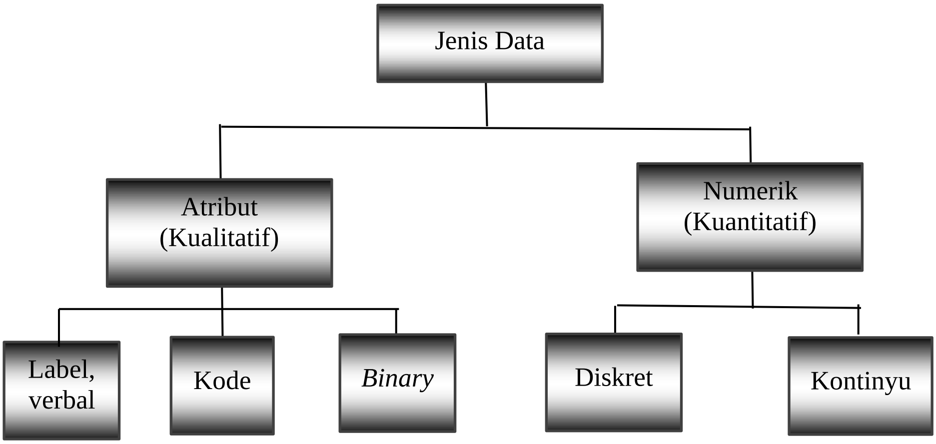
(b) Data numerik.

Data numerik atau data kuantitatif berasal dari pengukuran terhadap sesuatu, atau berasal dari operasi matematik, contoh: banyaknya klaim asuransi pada bulan Januari = 114, penjualan pada bulan Maret = Rp 25 juta. Sebagian besar data ini merupakan ukuran fisik, seperti: berat, panjang, waktu dan kecepatan. Data numerik terklasifikasi berdasar kemungkinan besaran nilai sebagai data deskrit dan data kontinyu.

(1) Data deskrit, nilainya integer (bukan pecahan), contoh: banyaknya mahasiswa yang hadir pada perkuliahan tertentu = 38 orang, banyaknya karyawan yang tidak masuk kerja pada hari tertentu = 2 orang.

(2) Data kontinyu, nilainya adalah sembarang angka pada suatu interval tertentu, jadi bias angka integer atau pecahan, contoh: berat beras = 100,00 kg atau 45,20 kg; tinggi badan = 172,0 cm atau 169,23 cm.

Secara umum, jenis data secara umum dapat digambarkan pada gambar berikut:



Gambar 2.2 Jenis Data

2.4. Sampel dan Sensus

Sebuah sampel melibatkan pengamatan terhadap beberapa item yang dipilih dari sebuah populasi, sedang sensus melibatkan pengamatan seluruh item dari populasi. Data sensus kadang-kadang bias menyesatkan karena masalah ketelitian. Contoh: sensus penduduk Amerika pada tahun1990, kehilangan data sekitar delapan juta penduduk, sedang sensus pada tahun 2000, kelebihan data sekitar 1,3 juta penduduk. Berikut ini rujukan pilihan pengumpulan data dengan sampel atau melalui sensus:

Tabel 2.1.Pilihan Sampel atau Sensus.

Sampel	Sensus
Populasi infinit (tidak berbatas). Jika populasi tidak berbatas,maka sensus tidak mungkin dilakukan, contoh: jumlah produksi baut pada sebuah lini produksi, jumlah pasien yang bisa diperiksa oleh seorang dokter.	**Populasi kecil.** Jika populasinya kecil, maka bisa dilakukan sensus.
Uji kerusakan. Untuk mengukur tingkat kerusakan hasil produksi. Jika hasil produksi banyak sekali, maka gunakan sampel untuk mempekirakan tingkat kerusakan barang.	**Populasi besar.** Jika kebutuhan banyaknya sampel mendekati populasinya, sebaiknya lakukan sensus.
Hasil tepat waktu. Jika ketersediaan data dibutuhkan dalam waktu singkat atau waktu tertentu.Hasil olahan data yang berasal dari sampel disebut sebagai "statistik".	**Tersedianya basis data.** Jika data lengkap tersedia pada suatu sumber. Contoh: jumlah penduduk Indonesia berdasar pendidikan, pendidikan, jenis kelamin,dan lainnya. Data hasil olahan sensus disebut *true value* atau parameter.
Ketelitian. Jika penyebaran sumber data yang terbatas, lebih baik menggunakan sampel. Ini juga berkaitan dengan biaya pengumpulan data.	**Kebutuhan legalitas.** Bank butuh perhitungan jumlah kas pada seluruh *teller* pada suatu hari.

Biaya. Walaupun sensus itu mungkin dilakukan, namun mengingat anggaran biaya yang terbatas, maka dapat dilakukan pengumpulan data melalui sampel.	**Biaya** Biaya cukup besar mengingat pengumpulan data membutuhkan tenaga yang banyak dan waktu yang relatif panjang. Contoh: sensus penduduk biasanya dilakukan dalam 10 tahun sekali.
Informasi sensitif. Jika informasi yang dibutuhkan merupakan hal sensitif, sebaiknya gunakan sampel. Contoh untuk mengamati pelecehan pada buruh wanita di suatu industri.	

Sumber:Doane,2007:31.

2.5 Metode Penyampelan

Dalam memilih sampel, ada dua katagori cara yaitu: (1) *probability sampling*, dan (2) *non probability sampling*. Metode pemilihan sampel tersebut dijelaskan pada tabel di bawah ini:

Tabel 2.2 Metode Penyampelan (*Sampling method*).

Probability	*Non Probability*
Simple random sample. Menggunakan angka random (acak) untuk memilih item dari sebuah daftar data.	***Judgment sample*.** Membutuhkan keahlian khusus untuk memilih item, misal: karyawan yang diwawancara.

Systematic sample. Memilih itemke-k dari sebuah daftar data atau urutan.Contoh: memilih konsumen ke-5 pada setiap restoran yang diamati.	***Convenience sample.*** Memilih sampel yang tersedia pada suatu waktu.Contoh: pada saat makan siang di restoran.
Stratified sample. Memilih secara acak pada strata terdefinisi. Contoh: berdasar umur, jenis kelamin, dan tempat tinggal.	
Cluster sample. Mirip *stratified sampling*, namun berdasar area geografis. Contoh: berdasar kode ZIP, kecamatan atau kota.	

Sumber : Bowen and Starr; 2002:34.

Simple Random Sample.

Ukuran populasi = N, dan ukuran sampel = n. Setiap item dari N memiliki peluang yang sama untuk dipilih sebagai sampel. Pilih-n integer, di mana n berada dalam rentang 1 sampai dengan N. Yang dimaksud dengan 'integer' adalah angka bulat, tidak pecahan.

<u>Contoh 2.1</u>: *Simple Random Sampel*

Tabel 2.3 Daftar Nama Karyawan.

No.	Nama	No.	Nama	No.	Nama
1	Achmad	17	Gazali	33	Nano
2	Adam	18	Gregy	34	Ningrum
3	Ade	19	Gatot	35	Nunung
4	Arief	20	Hadi	36	Pungki
5	Bambang	21	Haris	37	Purnomo
6	Bobi	22	Imam	38	Rafli
7	Citra	23	Inul	39	Rahman
8	Diana	24	Jadikan	40	Rudy
9	Dita	25	Jirin	41	Saleh
10	Dodi	26	Kardiman	42	Samirun
11	Edi	27	Karimullah	43	Totok
12	Erman	28	Kirman	44	Tukirin
13	Erwin	29	Liliek	45	Tulus
14	Fatah	30	Mardi	46	Wahyudi
15	Farida	31	Mustika	47	Wardi
16	Faruk	32	Mimiek	48	Wayan

Jika hanya 1 sampel yang dibutuhkan, maka dari ke-48 karyawan, dipilih secara acak seorang karyawan sebagai sampel, misal no. 30: Mardi. Atau dengan perintah Excel: ”=RANDBETWEEN(1,48)”.

<u>Contoh 2-2</u>:*Simpel Random Sample.*

Jika dibutuhkan lebih banyak sampel, dapat menggunakan dasar pemilihan sampel melalui Tabel Random Digit (biasanya pada buku-buku Statistik tabel ini disediakan). Atau dengan aplikasi Excel, angka random bisa dihasilkan melalui perintah misal:”=RANDBETWEEN(1;70000)”. Di mana 1 adalah batas bawah, dan 70000 = batas atas. Makin besar batas atasnya, makin besar digit angka random yang dihasilkan.

Tabel 2.4 Angka Random

628 44	618 35	248 47	089 81	429 27	492 37	076 65	222 00	667 72	071 49
313 08	201 95	483 70	616 32	507 56	300 75	413 37	372 66	249 08	472 76
161 30	581 38	128 43	296 12	643 57	477 63	357 79	502 47	481 13	246 38
259 51	638 27	314 18	615 04	113 10	376 53	019 33	305 79	144 68	415 45
626 52	604 05	261 55	401 43	498 48	019 67	073 22	087 75	533 07	658 74
072 43	517 13	187 09	045 85	225 13	236 04	206 46	671 07	323 75	543 23
547 48	284 10	020 07	319 88	032 77	275 08	253 91	450 45	310 09	346 82
203 28	273 06	613 54	021 17	524 87	402 60	363 67	033 41	027 17	152 62
245 88	691 95	570 61	123 22	013 09	269 43	029 79	682 16	395 65	657 30
269 52	303 11	380 83	251 81	294 86	640 90	050 65	699 80	533 59	294 18

Sumber : Bowen and Starr;2002:26.

Pada sebuah perusahaan yang mempekerjakan 500 orang karyawan, setiap karyawan telah diberi nomor urut 001 s/d.500.

Jika seandainya dibutuhkan 30 orang karyawan sebagai sampel untuk dievaluasi etos kerjanya, maka: pilih tiga digit pertama dari angka random pada Tabel 2.4 (mengingat 500 merupakan angka tiga digit) yang di bawah 500.

Tabel 2.5 Karyawan Terpilih Sebagai Sampel.

248	089	429	492	076
222	071	313	291	483
300	413	372	249	472
161	128	296	477	357
481	246	259	314	113
376	019	073	087	072

(2). *Systematic Sample*.

Memilih setiap sampel berdasar observasi ke-k, di mana k ditentukan secara sembarang.

Tabel 2.5 di atas, dapat diterjemahkan ke dalam daftar berikut:

			X		
	X				X
			X		
	X				X
			X		
	X				X
			X		
	X				X

Misal setiap k = 4, maka karyawan nomor = 4, 8, 12, 16 dan seterusnya, dipilih sebagai sampel.

(3) *Random Stratified Sample*.

Kadang-kadang peneliti dapat meningkatkan efisiensi sampel dengan menggunakan informasi awal populasi. Jika populasi dapat dibagi menjadi sub kelompok yang *relative homogeny* (ini disebut sebagai strata), maka pada setiap strata tersebut kemudian dipilih sampel secara acak. Atau pilih sampel secara acak dari populasi, kemudian strata diestimasi dengan menggunakan kombinasi bobot yang sesuai.

$n = n_1 + n_2 + n_3 + ... n_L$

Bobot setiap strata ke-j adalah $w_j = n_j/n$.

Contoh: jika banyaknya sampel = 200 orang karyawan, dan proporsi (sebagai pembobotan) diketahui, pria = 55,00% dan wanita = 45%; maka sampel pria = 110 orang dan sampel wanita = 90 orang.

Indeks Harga Konsumen adalah juga contoh sampel terstrata dari 90.000 item dan 364 katagori; yang dipilih dari sekitar 20.000 toko pengecer pada 85 area geografi.

***Random Cluster Sample*.**

Jenis sampel ini sebenarnya juga sampel terstrata, namun stratanya berdasar regional geografis. Contoh, pada sebuah kota, dibagi lagi menjadi sub daerah, seperti: blok tempat pemukiman, sekolah pada kecamatan. Pada *cluster sample* satu tahap, sampel mengandung seluruh elemen dari sub daerah (*cluster*) yang dipilih secara acak sebanyak k-sub daerah. Pada dua tahap, pertama pilih secara acak sebanyak k-sub daerah, kemudian pilih sebuah elemen pada setiap sub daerah.

Karena elemen dalam sebuah *cluster* kedekatan jarak, waktu perjalanan dan biaya pewawancara diupayakan minimum, maka cara penyampelan ini sangat bermanfaat jika:

(a) Kerangka populasi dan karakteristik strata tidak tersedia,

(b) Untuk memperoleh sampel terstrata terlalu mahal,

(c) Biaya perolehan data meningkat tajam sesuai dengan peningkatan jarak, (d) Kehilangan sedikit reliabilitas masih bisa ditoleransi.

Sesungguhnya cara ini cepat dan murah, dan seringkali hasilnya akurat, karena keluarga-keluarga yang berdekatan tempat tinggalnya cenderung memiliki penghasilan yang mirip, kemiripan etnis,dan latar belakang pendidikan. *Cluster sample* sangat sesuai untuk mengetahui pendapat politik, survey harga BBM, vaksinasi.

***Judgment Sample*.**

Metode ini merupakan *non probability sampling*, dan membutuhkan keahlian khusus peneliti untuk menentukan item yang dapat merepresentasi populasi. Contoh: untuk memilih sampel perusahaan pada industri peralatan medis berkaitan dengan biaya R&D yang dikeluarkan, peneliti membutuhkan diskusi dengan ahli perindustrian untuk memilih perusahaan yang dijadikan sampel. *Controlled quota judgment sampling* juga merupakan *judgment sampling*. Contoh: jika diketahui proporsi penduduk berdasar tingkat penghasilan:rendah = 35,00%, menengah = 45,00% dan tinggi = 20,00%. Pada sampel sebanyak 1.000 keluarga, sampel itu terdiri atas 350 keluarga berpenghasilan rendah, 450 keluarga berpenghasilan menengah dan 200 keluarga berpenghasilan tinggi. Sampel dipilih tidak secara random, pencacah menggunakan pertimbangan pribadi dalam memilih sampel.

(6). *Convenience Sample*.

Cara ini merupakan cara tercepat. Idenya adalah “meraih apa yang bisa terjangkau tangan”. Seorang guru besar akuntansi yang ingin mengetahui berapa banyak mahasiswa MBA akan memilih matakuliah Akuntansi Internasional, bisa mensurvey pada kelas yang sekarang diajarnya.

2.6 Ukuran sampel

Banyak pendapat yang menyatakan bahwa jika populasinya besar, maka dibutuhkan sampel yang juga besar untuk memperoleh tingkat ketelitian estimasi yang lebih tinggi. Pendapat ini sebenarnya tidak tepat. Ketelitian estimasi parameter data sebenarnya tergantung pada ukuran sampel bukan populasinya. Walaupun tidak ada ketentuan yang jelas tentang berapa ukuran sampel yang sesuai, tetap masih ada cara menentukan ukuran sampel optimal.

Untuk estimasi Rata-rata populasi.

Jika peneliti ingin mengestimasi rata-rata sebuah populasi,maka formula ukuran sampel optimal adalah:

$$n=\left(\frac{Z\,\sigma}{E}\right)^2 \qquad ....(2.1)$$

Di mana, n = ukuran sampel, E = kesalahan estimasi yang ditoleransi, z = adalah nilai area di bawah kurve normal, ini merupakan representasi tingkat keyakinan (*confidence interval*), umumnya = 5,00%. σ populasi didekati dengan s (standard penyimpangan sampel).

Ada empat cara untuk mengestimasi σ (standard penyimpangan populasi):

(1) Melakukan uji coba dengan sampel awal (*preliminary sample*),untuk menghitung standard penyimpangan sampel (s).

(2) Berasumsi bahwa populasi berdistribusi uniform[3]. Estimasi standard penyimpangan populasi: $\sigma = [(b–a)^2/12]$, di mana b = batas atas data dan a = batas bawah data.

(3) Berasumsi bahwa populasi berdistribusi normal.

[3]Balakrishnan,2003:56

Estimasi standard penyimpangan populasi: σ = (b–a)/4.

(4) Berasumsi bahwa populasi berdistribusi Poisson[4]. Estimasi standard penyimpangan populasi: σ = √λ, di mana λ adalah rata-rata kedatangan per jam dalam teori antrian.

<u>Contoh2.3</u>: Ukuran sampel untuk estimasi rata-rata populasi.

Seorang manajer produksi ingin mengetahui berapa rata-rata berat barang yang dikirim oleh seorang suplier, dengan menggunakan tingkat keyakinan 95,00% dan tingkat kesalahan estimasi yang dapat ditoleransi $\pm$ 1ons. Pengamatan terhadap data *preliminary* (12 unit barang yang dikirim) menunjukkan bahwa standard penyimpangan sampel (s) = 3,60 ons.Pada α = 95,00%, z = 1,96. E = 1, ukuran sampel dapat dihitung: $n = [(1{,}96)(3{,}60)/1]^2$ = 49,79 dibulatkan = 50 unit. Rata-rata berat populasi diestimasi dengan menggunakan sampel sebanyak 50 unit.

(2) Untuk estimasi proporsi populasi.

Jika peneliti ingin mengestimasi kebutuhan sampel sebagai proporsi sebuah populasi, maka formula ukuran sampel optimal adalah:

$$\boxed{n = \left(\frac{z\,\sigma}{E}\right)^2 \pi\,(1 - \pi)} \qquad \text{....(2.2)}$$

π = proporsi populasi, ini tidak diketahui. Untuk pendekatan dapat menggunakan *preliminary* sampel, atau dengan diasumsikan bahwa π = 0,50. E adalah tingkat kesalahan

[4]----------------, 2003:57

yang bias ditoleransi. z = nilai z pada kurve Z pada tingkat keyakinan (α) tertentu.

Contoh 2.4: Ukuran sampel dengan menggunakan estimasi proporsi populasi.

Seorang pimpinan bank kas pembantu yang berlokasi di sebuah universitas ingin mengetahui berapa proporsi penarikan uang oleh mahasiswa melebihi Rp 100.000,00 melalui ATM. Pada tingkat keyakinan = 95,00% dan tingkat kesalahan estimasi yang bisa ditoleransi $\pm$ 2,00%; jumlah kebutuhan sampel : n = $(1,96/0,02)^2$ (0,50)(0,50) = 2.401 penarikan melalui ATM. Jika pimpinan bank dari data tahun lalu mengetahui bahwa proporsi penarikan ATM melebihi Rp 100.000,00 = 27,00%; maka kebutuhan sampel yang harus diamati minimum: n = $(1,96/0,02)^2$ (0,27)(0,73) = 1892,94 dibulatkan 1.893 penarikan ATM.

Jika anggaran untuk pengumpulan data terbatas, maka peneliti bisa menurunkan tingkat keyakinannya, misal menjadi 90,00%, dan kesalahan estimasi yang bisa ditoleransi = 4,00%. Dengan demikian, kebutuhan sampel untuk diteliti, dengan asumsi bahwa π = 0,50 : n = $(1,645/0,04)^2$ (0,50)(0,50) = 422,81 dibulatkan menjadi 423 penarikan ATM.

2.7 Rangkuman

Data statistik adalah ukuran, informasi, nilai, atau karakter yang dapat menjelaskan suatu obyek yang diamati. Untuk memperoleh data yang benar, data harus *reliabel* (dapat diandalkan), harus memenuhi tiga syarat, yaitu: (i) obyektif; (ii) representatif; dan (iii) akurat. Berdasar karakteristiknya, data terklasifikasi sebagai: (a) data nominal ; (b) data ordinal; (c) data interval; dan data ratio.

Berdasar sumber perolehan, data terklasifikasi sebagai : (a) data primer; dan (b) data sekunder. Berdasar waktu pengumpulan, klasifikasi data adalah: (a) data rangkai waktu (*time series*); (b) data silang (*cross section*); dan (c) data gabungan (*pooled*).

Metode penyampelan (*sampling method*) terbagi atas: (a) metode probabilistik); dan (b) metode non probabilistik. Rincian metode probabilistik: *simple random sample*, *systematic sample*, *stratefied sample*, *cluster sample*. Rincian metode non probabilistik: *judgment sample* dan *convenience sample*.

Penentuan ukuran sampel tergantung tujuan mengumpulkan sampel, untuk mengestimasi rata-rata populasi, formula yang digunakan:

$$n = \left(\frac{z\,\sigma}{E} \right)^2 \qquad \text{....(2.1)}$$

Untuk mengestimasi proporsi populasi, formula yang digunakan:

$$n = \left(\frac{z\,\sigma}{E} \right)^2 \square\,(1 - \square) \qquad \text{.....(2.2)}$$

2.8 Diskusi

Banyak peneliti yang tidak tepat dalam mengolah data. Contoh, data yang bersifat ordinal, bahkan data nominal, terkadang diperlakukan sebagai data interval atau ratio. Mereka menjumlahkan, merata-rata dan menghitung standard penyimpangannya; kemudian menginterpretasikan nilai-nilai tersebut sebagai data yang bersifat interval atau ratio. Akibatnya, muncul skor baru yang diperoleh dari penjumlahan atau perata-rataan, yang kemudian dinarasikan sebagai data interval atau ratio. Situasi ini sering terjadi pada perhitungan skor variabel laten; di mana skor variabel laten merupakan jumlah atau rata-rata skor indikator. Pada analisis deskriptif frekuensi distribusi, selayaknya untuk data ordinal, nominal dan interval selayaknya gunakan nilai modus, median, kuartil, desil, atau persentil.

2.9 Referensi

Balakrishnan, N., and V. B. Nevzorov, 2003, ***A Primer on Statistical Distributions***, John Wiley and Son Inc., New York.

Bowen, Earl, K., and Martin K. Starr, 2002, ***Basic Statistics for Business and Economics,*** McGraw Hill Book Company, Tokyo.

Doane, David, P., and Lori E. Seward, 2007, ***Applied Statistics in Business and Economics***, McGraw Hill/Irwin Series, New York.

Huber, Peter, J., 2003, ***Robust Statistics***, John Wiley and Son Coy., Tokyo.

Kendall, M. G., and Stuart, A., 2005, ***The Advanced Theory of Statistics***, Vol. 1 6th Edition, McMillan Publishing Company, NewYork, 2005.

Supranto, J, 2008, Statistik, **Teori dan Aplikasi**, Penerbit Erlangga, Jakarta.

2.10 Latihan Soal

Jelaskan klasifikasi data berdasar tiga aspek: (i) menurut karakteristik data; (ii) menurut sumber data; dan (iii) menurut waktu perolehannya. Berikan masing-masing contohnya.

Klasifikasikan karakter dan sifat dari data berikut ini (nominal, ordinal, interval dan ratio):

(i) Pendidikan SMA = 1, Sarjana Muda = 2, Sarjana = 3.

(ii) Islam = 1, Kristen = 2, Budha = 3, Lainnya = 4.

(iii) Nilai ujian : $\geq$ 80 = A, 70 = B, 60 = c dan $\leq$ 50 = D.

(iv) Madura = 0, Jawa = 1, dan Sunda = 2.

(v) Populasi ternak sapi di daerah A = 1.200 ekor, di daerah B = 800 ekor.

(vi) Masa kerja pegawai: 0, 1, 2 tahun.

Manakah yang termasuk data primer dari informasi berikut:

(i) Data populasi penduduk di beberapa kabupaten.

(ii) Jumlah anggota dalam setiap keluarga pada sebuah RT.

(iii) Jumlah kendaraan bermotor yang parkir di lahan parkir tertentu pada pukul 08.00 – 12.00.

(iv) Jumlah hasil produksi kopi afdeling A, dan B pada tahun 2015.

Manakah yang termasuk data gabungan (*pooled data*) dari informasi berikut:

(i) Jumlah hasil produksi jagung, padi dan sorgum di daerah A, B, C dan D pada tahun 2015 – 2017.

(ii) Hasil penjualan ternak ayam pedaging di daerah A pada tahun 2016 – 2017.

(iii) Jumlah hasil produksi susu sapi di tiga kecamatan Kabupaten A pada tahun 2015.

Seorang direktur produksi pada sebuah perusahaan daerah perkebunan (PDP) ingin meneliti tentang berapa paket kemasan sampel bibit yang harus diamati, jika ia mengetahui dari data historis bahwa tingkat kerusakan bibit diterima dari supplier = 10,00%. Pada tingkat keyakinan = 95,00% dan toleransi kesalahan = $\pm$ 5,00%, berapa paket kemasan yang harus dipilih sebagai sampel ?

BAB 3

PENYAJIAN VISUAL DATA

Penyajian data dapat berbentuk tabel atau grafis. Bentuk grafis lebih mudah diinterpretasikan dibanding bentuk tabel. Kesulitan menyusun bentuk grafis sekarang ini telah dapat diminimalisir dengan berbagai aplikasi program komputer.

Kemampuan Akhir yang Diharapkan:
- Mahasiswa mampu menyajikan data secara visual dalam berbagai bentuk tabel dan grafis. - Mampu menyusun distribusi frekuensi data dengan metode Sturgest.

3.1. Pendahuluan

Ahli statistik pada berbagai institusi harus mengorganisasi, meng-eksplorasi danmembuat ringkasan data secara singkat namun jelas. Caranya bisa visual (dalam bentuk grafik) atau numerik (dalam bentuk tabel). Langkah-langkah untuk menyajikan data adalah:

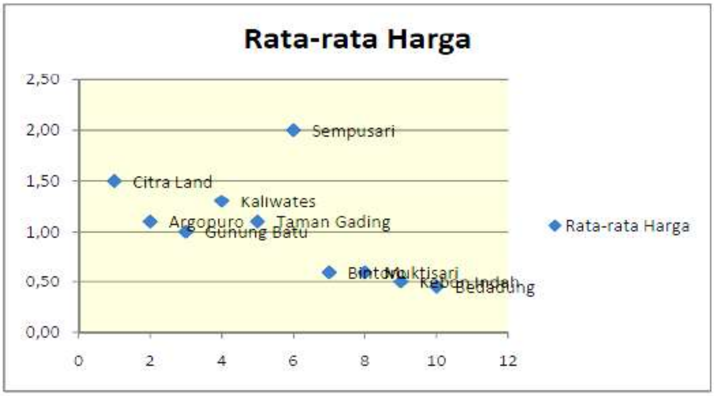
(1) Melakukan *sorting*. *Sorting* adalah mengurutkan data secara *ascending*, yaitu dari nilai terkecil ke nilai lebih besar.
(2) Membuat grafik atau tabel seperti yang dibutuhkan.
Pilihan grafik adalah:
(a) plot titik atau *dot plots*,
(b) histogram,
(c) grafik garisatau *line chart*,
(d) grafik kolom atau *bar chart*,
(e) diagram pencar atau *scatterplot*,
(f) peta dan *pictogram*.

3.2. Penyusunan grafik

(1). *Dot plots*.

Dot plots adalah grafik paling sederhana. Langkah-langkah menyusun *dot plots* adalah:
(a) membuat skala sesuai data,
(b) tandai sumber datar dan buat labelnya,
(c) plot setiap data sebagai titik di atas skala pada lokasi yang sesuai. Walaupun data sedikit, *dot plots* ini menghasilkan gambaran data dengan mudah.

Contoh 3.1: *Dot Plots* Rata-rata Harga Rumah.



Gambar 3.1 *Dot Plots* Rata-rata Harga Rumah

(2). Grafik Garis.

Grafik garis sederhana biasanya digunakan untuk data rangkai waktu, membuat garis *trend* atau untuk membandingkan data antar waktu. Jika dua variabel yang ditampilkan, di mana skala ukuran bisa berbeda, maka sumbu sebelah kiri untuk ukuran variabel ke-1, dan sumbu sebelah kanan untuk ukuran variabel-2.

Contoh 3.2: Grafik Garis Keuntungan Bersih/Tahun.

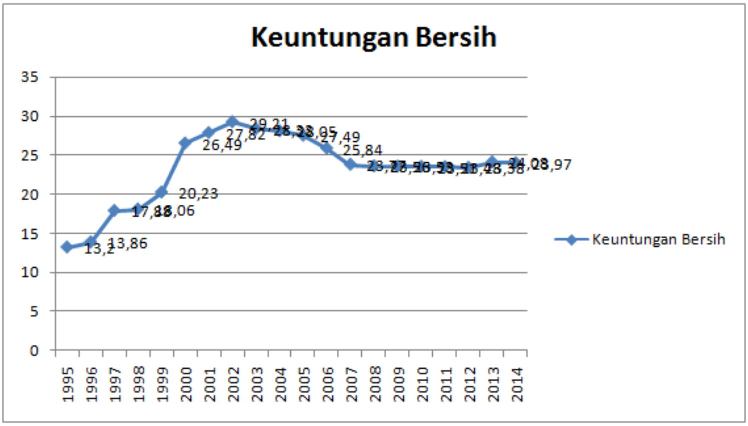
Laporan keuangan yang telah diaudit dari perusahaan A sejak tahun 1995 sampai dengan tahun 2014 adalah:

Tabel 3.1 Perkembangan Keuntungan Bersih/Tahun Perusahaan A.

Tahun	Keuntungan Bersih ($juta)	Tahun	Keuntungan Bersih ($juta)
1995	13,20	2005	27,49
1996	13,86	2006	25,84
1997	17,88	2007	23,77
1998	18,06	2008	23,56
1999	20,23	2009	23,53
2000	26,49	2010	23,51
2001	27,82	2011	23,48
2002	29,21	2012	23,38
2003	28,33	2013	24,08
2004	28,05	2014	23,97

Sumber : Bagian Keuangan Perusahaan A, 2015.

Secara visual, tabel ini tidak segera memperlihatkan pola data, maka dengan grafik garis sederhana, pola data bisa segera tampak.



Gambar 3.2 Grafik Garis Keuntungan Berih/Tahun Perusahaan A

Sebuah grafik garis menjadi efektif, jika :

(a) Grafik garis digunakan untuk data rangkai waktu dan bukan data silang.

(b) Angka numerik menjadi sumbeu tegak (Y), angka waktu menjadi sumbu datar (X).

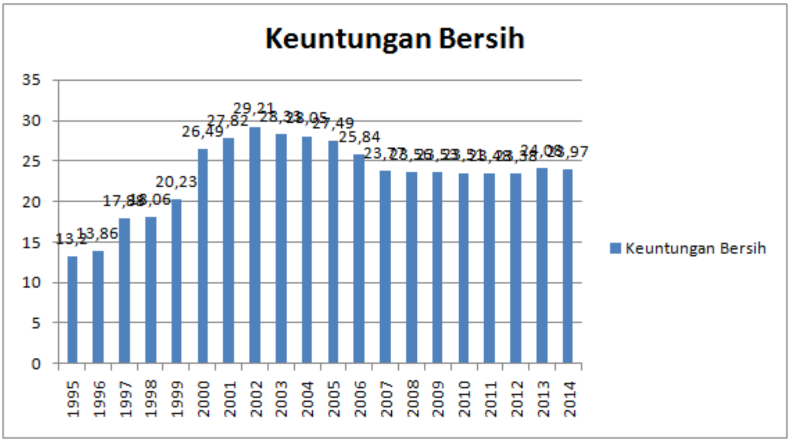
(c) Untuk menghindari ketidak jelasan data, bisa ditambahkan label data pada setiap titik pengamatan, dan garis *grid*.

(3). Grafik Kolom.

Grafik kolom menggambarkan atribut data pada sumbu datar, nilainya pada sumbu tegak.

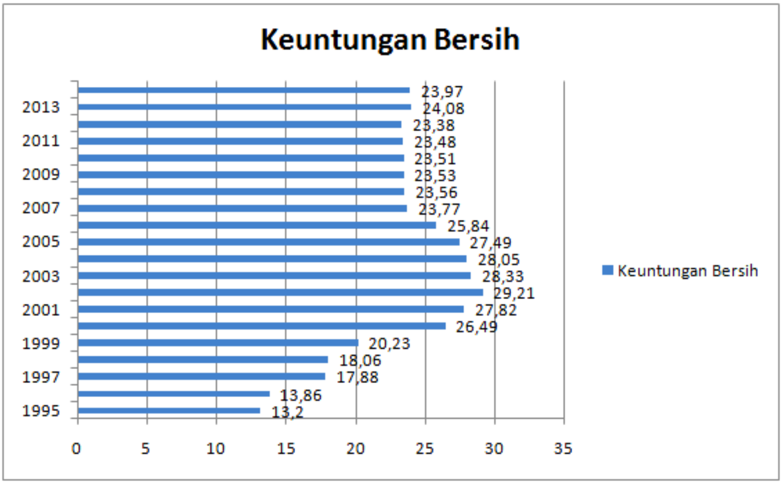
<u>Contoh 3.3</u>: Grafik Kolom Keuntungan Bersih/Tahun.

Kembali ke Contoh 3.2, grafik kolom dapat dibuat sebagai berikut:



Gambar 3.3 Grafik Kolom Keuntungan Bersih/Tahun Perusahaan A

Jika sumbu datar terlampau panjang, grafik kolom dapat dibentuk horisontal, hasil untuk contoh data di atas adalah:



Gambar 3.4 Grafik Kolom Keuntungan Berih/Tahun Perusahaan A.

Gambar 3.4 di atas menjelaskan pola perubahan data antar tahun yang sama dengan pola pada Gambar 3.3.

<u>Contoh 3.4</u>: Grafik Kolom Berjenjang.

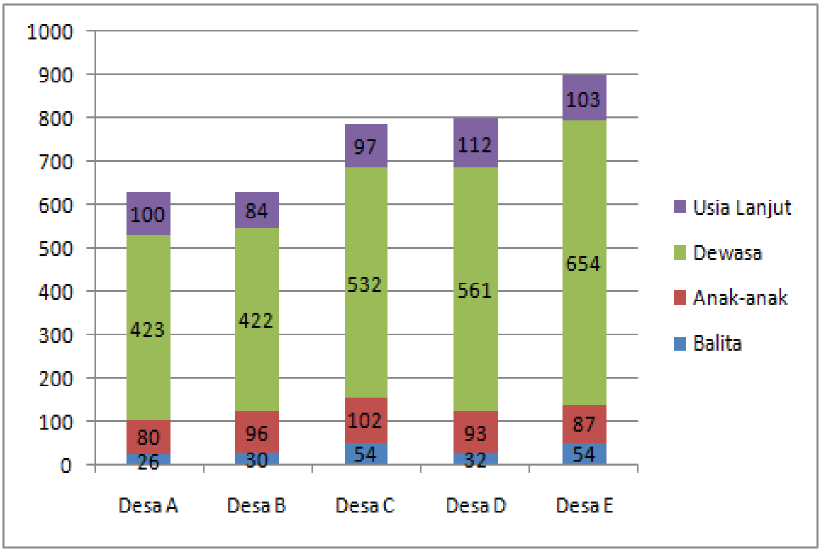
Jika pada setiap obyek yang diamati terdapat beberapa unit analisis, maka grafik kolom dibentuk menjadi grafik kolom yang berjenjang. Contoh: pada lima desa A, B, C, D dan E, dilaporkan jumlah orang lanjut usia, dewasa, anak-anak dan balita seperti yang dirangkum dalam tabel berikut:

Tabel 3.2. Jumlah Penduduk Desa Berdasar Usia, Tahun 2015.

Kelompok Usia	Desa A	Desa B	Desa C	Desa D	Desa E
Lanjut Usia	100	84	97	112	103
Dewasa	423	422	532	561	654
Anak-anak	80	96	102	93	87
Balita	26	30	54	32	54

Sumber : Dinas Kependudukan Kota X, 2016.

Grafik kolom berjejang dari data di atas adalah:



Gambar 3.5 Grafik Kolom Berjenjang Berdasar Struktur Usia.

Sebuah grafik kolom menjadi efektif, jika:

(a) Data numerik diperlihatkan sebagai sumbu Y, dan katagori sebagai sumbu X.

(b) Ketinggian kolom sesuai dengan proporsi data.

(c) Berilah label pada setiap kolom.

(4). Diagram Pencar.

Diagram pencar merupakan grafik titik koordinat dari dua variabel, di mana sumbu-sumbunya adalah nilai kedua variabel yang diperkirakan memiliki hubungan.

<u>**Contoh 3.5**</u>: Diagram Pencar (*Scatter Plot*).

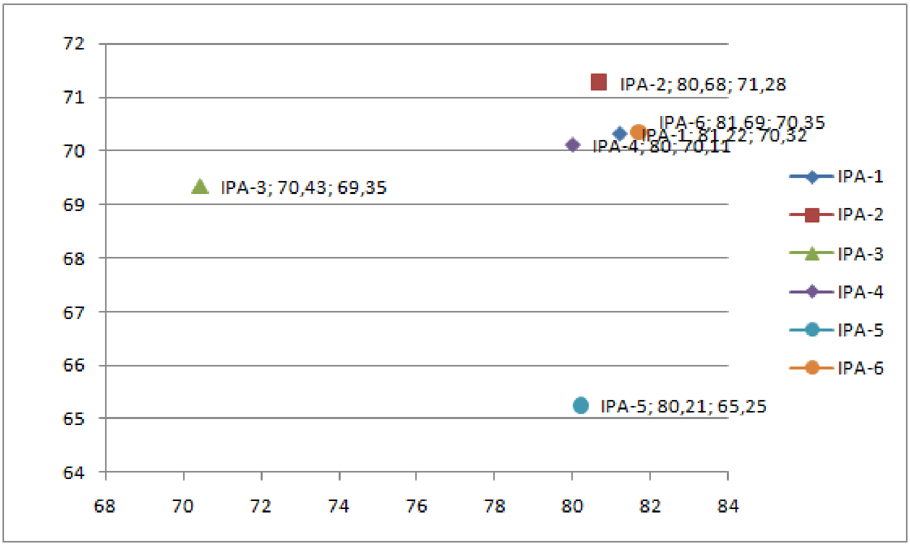
Rata-rata nilai Kimia dan rata-rata nilai Matematik siswa kelas 2-IPA padaenam kelas di sebuah SLTA pada Semester-2 Tahun Ajaran 2016/2017 adalah:

Tabel 3.3 Rata-rata Nilai Kimia dan Matematik Siswa IPA pada Semester 2 Tahun Ajaran 2016/2017.

Kelas	IPA-1	IPA-2	IPA-3	IPA-4	IPA-5	IPA-6
Kimia	70,32	71,28	69,35	70,11	65,45	70,35
Matematik	81,22	80,68	70,43	80,00	80.21	81,69

Sumber : Bagian Akademik SLTA, 2018.

Dari Tabel 3.3 di atas, dapat dibuat diagram pencar sebagai berikut:



ambar 3.6 Grafik Diagram Pencar Nilai Kimia dan Matematik.

Sumbu tegak (*vertical axis*) adalah rata-rata nilai Kimia, dan sumbu datar (*horizontal axis*) adalah rata-rata nilai matematik.

(5). Grafik Pie

Grafik Pie masih sering digunakan pada perusahaan, walaupun tingkat akurasinya kurang baik; karena umumnya tidak memberikan angka absolut, tetapi dalam bentuk persentase.

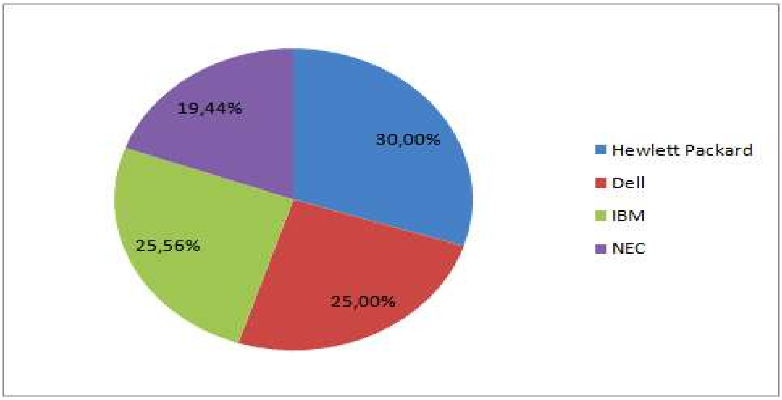
Contoh 3.6: *Pie Chart*.

Tabel 3.4 Penjualan Berbagai Merk PC di Kota Jember Tahun 2015.

Tahun	Hewlett Packard	Dell	IBM	NEC
2015	30,00%	25,00%	25,56%	19,44%

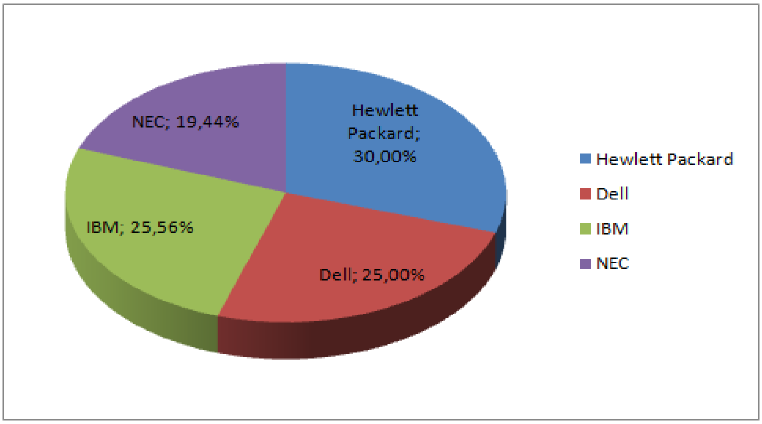
Sumber : Dinas Perindustrian dan Perdagangan Kabupaten Jember, 2016.

Dari tabel ini, dapat dibuat *pie chart* sebagai berikut:



Gambar 3.7 *Pie Chart* Penjualan Berbagai Merk Komputer Tahun 2015.

Kadang-kadang untuk mendapatkan tampilan yang lebih menarik, grafik pie digambarkan dalam bentuk tiga dimensi seperti:



Gambar 3.8 *Pie Chart* Tiga Dimensi Penjualan Berbagai Merk Komputer Tahun 2015.

3.3. Penyusunan Tabel.

Tabel merupakan bentuk penyajian data yang paling sederhana, melalui pengaturan baris dan kolomnya, sehingga arti data dengan mudah dapat dipahami[5].

Judul Tabel: Ditulis ditengah-tengah bagian teratas dengan huruf awal kapital pada setiap kata; singkat dan jelas dicantumkan meliputi: apa, macam atau klasifikasi, serta bilangan dan satuan atau unit data yang digunakan.

Tiap baris hendaknya menjelaskan sebuah pernyataan lengkap, dan jangan dilakukan pemisahan bagian kata. Judul kolom, ditulis dengan singkat dan jelas, bisa dalam beberapa baris.

[5]Setyawan, 2009:8

Sel Tabel: Adalah tempat data dituliskan

Catatan: Merupakan catatan-catatan yang diperlukan, dan juga sumber yang menjelaskan dari mana data itu dikutip.

Contoh 3.7 : Pengeluaran Biaya Operasional Sekolah.

Tabel 3.5 Pengeluaran Biaya Operasional Sekolah dan Perguruan Tinggi Di Kabupaten Jember pada Tahun 2000 – 2005 (dalam jutaan Rupiah).

Tahun	Seluruh Sekolah dan PT	SekolahDasar+SLTP+SLTA			Akademi+Universitas		
		Total	Negeri	Swasta	Total	Negeri	Swasta
2000	422	155	79	76	267	132	135
2001	429	150	80	70	279	143	136
2002	453	160	85	75	293	154	139
2003	470	171	88	83	299	159	140
2004	489	177	90	87	312	162	150
2005	508	183	92	91	325	165	160

Sumber: Dewan Pendidikan Kabupaten Jember, 2006.

Kolom ‘Seluruh Sekolah’ merupakan jumlah pengeluaran total SD. SLTP, SLTA + total pengeluaran Akademi, Universitas. Kolom “Total” merupakan pengeluaran masing-masing sekolah negeri + swasta, atau masing-masing akademik negeri + swasta.

Tabel yang efektif dibuat dengan :

(1) Buat tabel yang sederhana, konsisten dengan kegunaannya. Tabel merupakan ringkasan, di mana detilnya dapat dibuat dalam sebuah lampiran tersendiri.

(2) Atur data dengan dasar kolom, karena hasil penelitian membuktikan bahwa dasar kolom lebih mudah dipahami oleh pembacanya.

(3) Untuk keperluan presentasi, sebaiknya gunakan angka-angka bulat

(4) Judul kolom dan baris dibuat sederhana dan jelas.

(5) Dalam sebuah kolom buat digit decimal angka yang konsisten dan seragam.

3.4 Penyusunan Distribusi Frekuensi.

Sebuah distribusi frekuensi adalah sekumpulan data yang dikelompokkan dalam beberapa kelas, di mana pada setiap kelas ada batas bawah dan batas atasnya. Tujuan penyusunan distribusi frekuensi adalah meringkas data sehingga lebih mudah untuk diinterpretasikan.

Penyusunan distribusi data bisa subyektif, namun ada sebuah acuan yang banyak diikuti oleh para analis, yaitu metode Sturgest. Langkah-langkah dalam menyusun distribusi frekuensi adalah[6]:

(1) Menentukan banyaknya kelas (*class*).
Formula : dengan metode Sturgest.

$$\mathbf{k = 1 + 3{,}222 \log (n),}$$

di mana n = banyaknya data, k = banyaknya kelas.
Contoh, jika n = 100
k = 1 + 3,222 log(100) = 1 + 6,444 = 7,644 atau dibulatkan 7 kelas.

(2) Menentukan interval kelas (*class interval*, C),

Formula:

$$\boxed{C = \frac{X_U \pm X_L}{k}} \qquad \text{. . . . . (3.1)}$$

X_L = data terkecil, X_U = data terbesar

[6]Balakrishnan, 2003.

Syafiadi Rizki Abdila, S.T., Ph.D, Prof. Dr. Ir. Drs. Syafwandi, M.Sc
Dr. Yulasmi, S.E., M.M, Dr. Robby Dharma, S.E., M.M
Dr. Syafri Arlis, S.Kom, M.Kom

(3) Batas kelas (*class boundary*).
Batas kelas atas (*upper boundary*) adalah nilai tertinggi pada sebuah kelas, dan batas kelas bawah (*lower boundary*) adalah nilai terendah pada sebuah kelas. Batas semu adalah batas kelas yang mengandung nilai batas di antara kelas yang satu dengan kelas berikutnya.

(4) Titik tengah adalah nilai yang terdapat ditengah-tengah sebuah kelas, yang dihitung sebagai batas kelas bawah + batas kelas atas dibagi 2.

(5) Rentang data (*data range*) adalah selisih nilai tertinggi dengan nilai terendah dari seluruh data.

<u>Contoh 3.8</u> : Penyusunan Distribusi Frekuensi.

Menyusun distribusi frekuensi untuk data modal usaha dari 100 perusahaan:

75	86	66	86	50	78	66	79	68	60
80	83	87	79	80	77	81	92	57	52
58	82	73	95	66	60	84	80	79	63
80	88	58	84	96	87	72	65	79	80
86	68	76	41	80	40	63	90	83	94
76	66	74	76	68	82	59	75	35	34
65	63	85	87	79	77	76	74	76	78
75	60	96	74	73	87	52	98	88	64
76	69	60	74	72	76	57	64	67	58
7	80	7	5	7	8	7	4	7	5

2	2	6	3	2	8	5	5	6

n = banyaknya pengamatan = 100.

Jumlah kelas = k = 1+3,222 log(100) = 1+3,222(2) = 7,644,dibulatkan = 7 kelas. Nilai terendah (X_L) = 34 dan nilai tertinggi (X_U) = 98, dari nilai-nilai ini dapat ditentukan kelas interval atau rentang kelas, C = (98-34)/7 = 9,14, dibulatkan = 9.

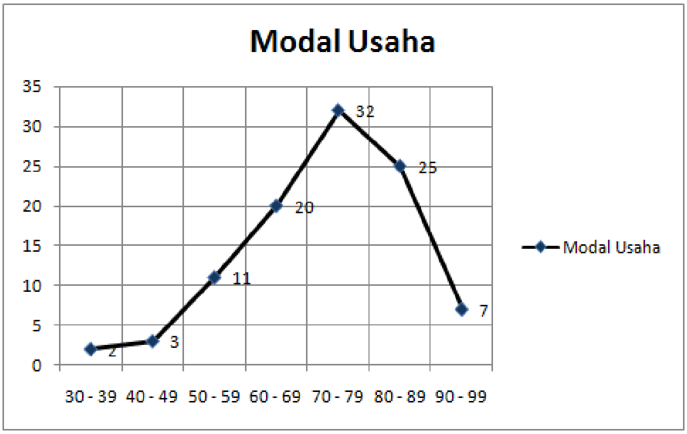
Tabel distribusi frekuensi dapat disusun sebagai berikut:

Tabel3.6.Distribusi Frekuensi Data.

Batas Kelas	Nilai Tengah	Frekuensi
30-39	34,5	2
40-49	44,5	3
50-59	54,5	11
60-69	64,5	20
70-79	74,5	32
80-89	84,5	25
90-99	94,5	7

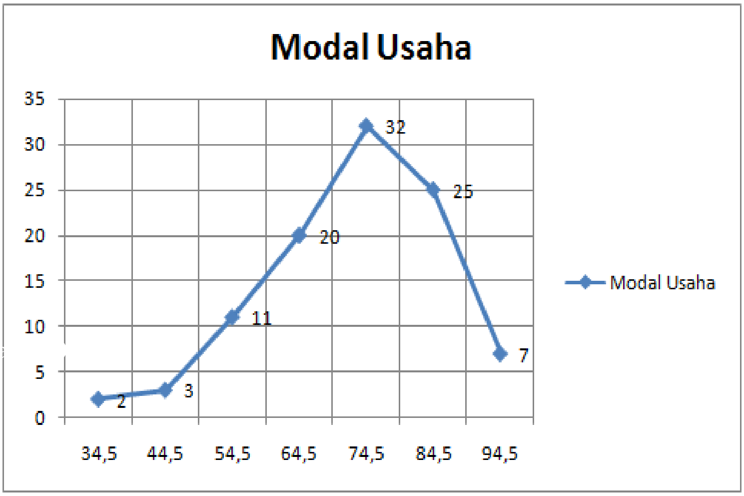
Ada tujuh kelas (k = 7), dan rentang kelas (C) pada setiap kelas = 9. Frekuensi diperoleh dengan menghitung frekuensi secara manual dari data.

Selanjutnya, dari tabel di atas, dapat dibuat grafik garis sebagai berikut:



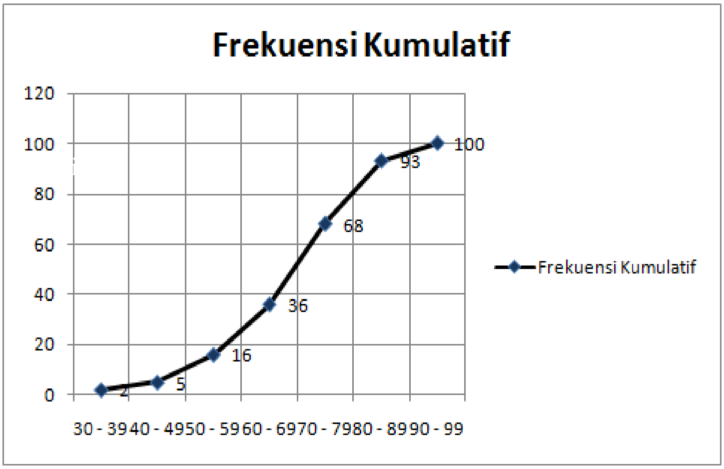
Gambar 3.9 Grafik Distribusi Frekuensi Kepemilikan Modal Usaha.

Dapat pula dibuat frekuensi dengan memanfaatkan kelas tengah (*class mark*) pada setiap kelas. Grafik dengan memanfaatkan nilai tengah disebut "polygon".



Gambar 3.10 *Polygon* Kepemilikan Modal Usaha

Bentuk histogram lain adalah distibusi frekuensi relatif kumulatif, yaitu dengan membuat grafik kumulatif frekuensinya, hasilnya sebagai berikut:



Gambar 3.11 Frekuensi Kumulatif Perusahaan Berdasar Modal Usaha.

Aplikasi Komputer (Excel) untuk membuat distribusi frekuensi, dapat dilakukan melalui langkah-langkah seperti contoh di bawah ini:

12	2	1	
	7	6	
15	2	1	
	1	8	
20	1	1	Min = 12
	8	7	
22	1	2	Max = 32
	9	3	
14	1	2	n = 20
	8	8	
14	2	1	
	2	3	
15	3		
	3		

Dengan metode Sturgest:

Jumlah kelas = 1+3,333 log(n) = 1 + 3,333 (1,301) = 5,336 atau 5 kelas,

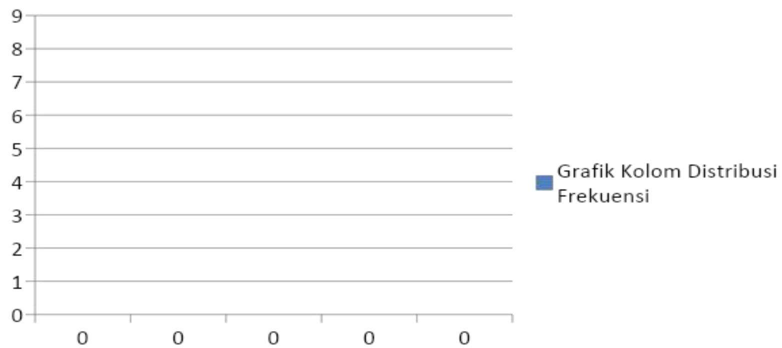
Classinterval = (33- 12)/5 = 4,2 dibulatkan menjadi 4

Selanjutnya buat kelas sebagai berikut:

dimulai dari angka puluhan yang lebih kecil terdekat dengan angka minimum, yaitu =10.Kelas dapat dibuat sebagai berikut:

10–14	$f_t = 4$
15–19	$f_t = 8$
20–24	$f_t = 5$
25–29	$f_t = 2$
30–34	$f_t = 1$

Batas bawah pada kelompok pertama ditentukan dengan memilih angka puluhan terdekat dengan angka minimum. Angka minimum = 12, maka batas bawah pada kelompok pertama = 10. Grafik kolom dapat dibuat sebagai berikut:



Gambar 3.12 Grafik Kolom Distribusi Frekuensi

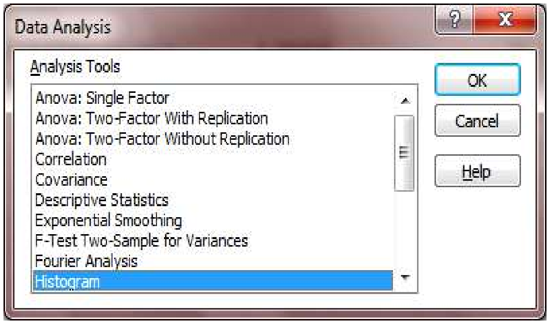
(a). Dengan aplikasi Excel, ketik data pada lembar kerja:

	A	B	C	D	E	F	G
1	12						
2	15	Bin					
3	20	14					
4	22	19					
5	14	24					
6	14	29					
7	15	34					
8	27						
9	21						
10	18						
11	19						
12	18						
13	22						
14	33						
15	16						
16	18						
17	17						
18	23						
19	28						
20	13						

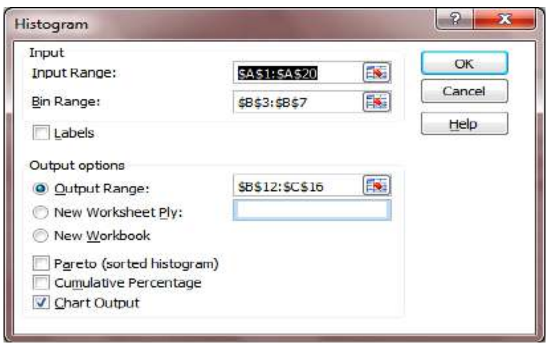
Data diketikkan pada sel A1 s/d A20. Ketik 'Bin' pada kolom B2. Angka bin merupakan batas atas kelas, ketikkan pada sel B3 s/d B7 (mengingat ada 5 kelas).

(b). Selanjutnya pilih menu 'Data'.

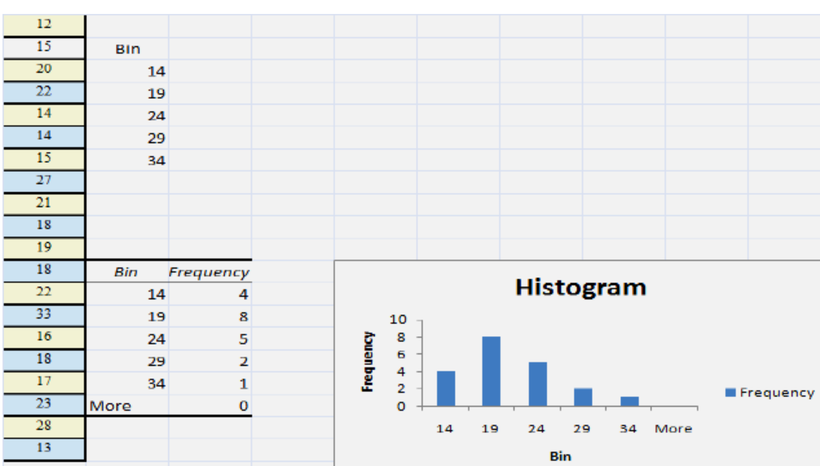
(c). Pilih sub menu 'Data Analysis'. Muncul tampilan sebagai berikut:



Pilih 'Histogram'. Tekan OK.



Isikan input range: $A$1:$A$20. Isikan bin range: $B$3:$B$7. Isikan output range: $B$12:$C$16 (untuk lima kelas).Klik Ok, hasilnya adalah:



Gambar 3.13 Grafik Kolom Distribusi Frekuensi Modal Usaha.

<u>Contoh-3.9</u>:Penyusunan Distribusi Frekuensi dengan Excel.

Data penghasilan per bulan (dalam sepuluh ribuan rupiah):

255,5	285,3	287,5	215,2	310,	313,	248,9	341,8
0	2	2	3	52	98	2	1
333,1	291,7	290,4	254,1	334,	228,	233,9	373,2
2	2	1	9	27	42	1	5
358,2	297,0	299,4	216,7	289,	257,	267,9	274,0
1	1	2	6	62	92	4	5
282,6	308,6	235,7	251,4	259,	258,	242,1	279,4
2	2	2	3	12	14	1	4
303,6	275,2	294,5	293,6	276,	278,	250,2	274,0
0	1	5	5	31	22	7	8
272,0	268,0	366,5	354,8	365,	256,	261,1	289,7
1	3	4	3	42	42	2	3
259,7	260,1	280,2	292,4	219,	295,	236,2	296,2
2	3	7	4	03	43	2	1
298,6	309,3	282,7	281,3	305,	311,	282,9	283,9
5	7	9	4	02	74	0	1
315,7	325,6	286,4	288,0	318,	325,	345,7	351,6
2	2	3	6	12	41	9	2
314,7	306,4	312,0					
8	3	4					

Dengan aplikasi Excel:

	A	B	C	D	E	F	G	H
1	255,5	285,32	287,52	215,23	310,52	313,98	248,92	341,81
2	333,12	291,72	290,41	254,19	334,27	228,42	233,91	373,25
3	358,21	297,01	299,42	216,76	289,62	257,92	267,94	274,05
4	282,62	308,62	235,72	251,43	259,12	258,14	242,11	279,44
5	303,6	275,21	294,55	293,65	276,31	278,22	250,27	274,08
6	272,01	268,03	366,54	354,83	365,42	256,42	261,12	289,73
7	259,72	260,13	280,27	292,44	219,03	295,43	236,22	296,21
8	298,65	309,37	282,79	281,34	305,02	311,74	282,9	283,91

9	315,72	325,62	286,43	288,06	318,12	325,41	345,79	351,62
10	314,78	306,43	312,04					

- Hitung banyaknya kelas: k = 1+3,222 log(n), hasil Excel =7,25014, dibulatkan menjadi, k = 8 kelas, n = 75. Ini dilakukan dengan perintah Excel:"=COUNT(A1:H10)", Excel menghitung banyaknya data yang ada pada selA1 s/dH10, sel-sel kosong tidak dihitung.
- Hitung interval kelas: C={max– min}/k

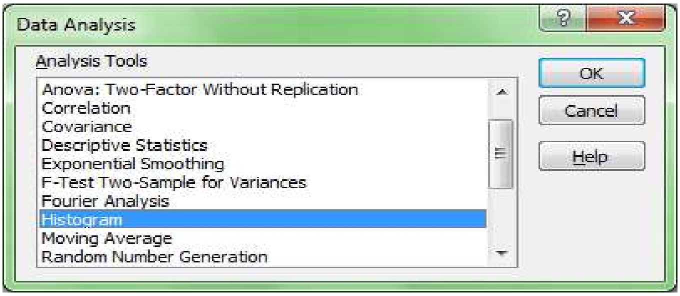
  Nilai maksimum dan minimum dari data di atas, diperoleh dengan perintah Excel "=MAX(A1:H10)", dan "=MIN(A1:H10)". Hasilnya, nilai maksimum = 373,25 dan nilai minimum = 215,23.

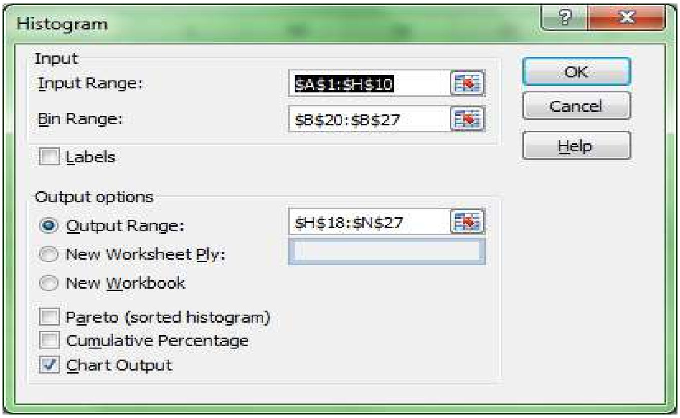
  Hasil Excel memperlihatkan bahwa C=19,7525 dan dibulatkan = 20. Batas bawah untuk kelompok pertama adalah angka bulat terdekat, dengan nilai minimum yaitu : 215.
- Dengan demikian, dapat dibuat kelompok data sebanyak 8 kelas, batas bawah = 215, dan C= 20, hasilnya adalah sebagai berikut:

Kelas	Bin
215 - 234,99	234,99
235 - 254,99	254,99
255 - 274,99	274,99
275 - 294,99	294,99
295 - 314,99	314,99
315 - 334,99	334,99
335 - 354,99	354,99
355 - 374,99	374,99

Batas atas adalah batas bawah + C, contoh: untuk kelompok pertama *range* kelas = 215 –235; namun agar tidak terjadi tumpang tindih (*overlap*) data, buat koreksi pada batas atas kelas dengan dikurangi 0,01; sehingga *range* kelas berubah menjadi 215 – 234,99, dan seterusnya perlakukan hal yang sama untuk setiap kelompok berikutnya. Bin adalah batas kelas atas.

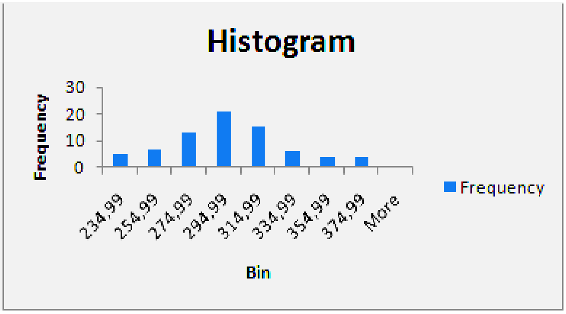
Selanjutnya pilih ‘Data’ pada lembar kerja Excel:





Tekan OK, hasilnya adalah:

Bin	*Frequency*
234,99	5
254,99	7
274,99	13
294,99	21
314,99	15
334,99	6
354,99	4
374,99	4
More	0



Gambar 3.14 Grafik Kolom Distribusi Frekuensi Penghasilan.

3.5 Rangkuman

Penyajian data dapat berbentuk visual (grafik) atau numerik (tabel). Langkah-langkah penyusunan adalah : (a) sorting data, dan (b) membuat grafik atau tabel. Jenis grafik adalah: (a) *dot plot*; (b) histrogram; (c) *line chart*; (d) *bar chart*; (e) *scatter plot*; (f) *pie chart*; dan (g) *pictogram*. Tabel dapat berbentuk tabel dengan data tidak terkelompok, dan data terkelompok (frekuensi distribusi). Metode penyusunan distribusi frekuensi yang paling banyak digunakan adalah metode yang dikembangkan Strugest.

Formula yang digunakan pada metode Sturgest untuk menyusun distribusi frekuensi adalah:

(a) Penentuan banyaknya kelas: **k = 1 + 3,222 log (n)**

(b) Penentuan interval kelas:

$$\mathbf{C = \frac{X_U \pm X_L}{k}} \quad . . . . . (3.1)$$

X_L = data terkecil, X_U = data terbesar

(c) *Class boundary* adalah : *upper boundary* (batas atas kelas) dan *lower boundary* (batas bawah kelas).

(d) Rentang data (*data range*) adalah selisih nilai tertinggi dan terendah dari seluruh data.

3.6 Diskusi

Penyajian data secara visual dalam bentuk grafik memiliki keunggulan: cepat untuk dipahami pola datanya, namun kurang detil jika dibanding penyajian data dalam bentuk tabel. Dengan demikian, dapat dilakukan penyajian data dalam bentuk grafik dan dilengkapi pula dalam bentuk tabel.

Sekarang ini, penyusunan grafik dan distribusi frekuensi dapat dipermudah dengan aplikasi program Excel atau program lainnya.

3.7 Referensi

Balakrishnan, N., and V. B. Nevzorov, 2003, ***A Primer on Statistical Distributions***,John Wiley and Son Inc., New York

Doane, David, P., and Lori E. Seward, 2007, ***Applied Statistics in Business andEconomics***, McGraw Hill/Irwin Series, New York

Kirkpatrick, E., G., 2009, ***Introductory Statistics and Probability for Engineer,Science and Technology*, **Prentice-Hall Inc., Englewood Cliffs, N. J.

Setyawan, HB, 2009. **Statistika (Deskriptif dan Diferensial),** Perpustakaan Nasional:Katalog dalm Terbitan ISBN 978-602-98107-0-7, Yayasan Dharma Nusantara, Jember

3.8 Latihan Soal

Data hasil panen kopi Robusta dan Arabica (dalam satuan ton) di perkebunan kopi rakyat di Kecamatan Panti berdasar blok kebun pada tahun 2016 adalah sebagai berikut:

Blok Kebun	Robusta	Arabica
132	11,23	7,32
56	8,99	6,12

112	7,00	8,09
76	12,87	3,45
34	10,56	6,88
57	8,45	7,45

(a) Buatlah *dot plots* untuk hasil produksi kopi Robusta Arabica pada keenam blok kebun.
(b) Buatlah grafik garis untuk hasil produksi kopi Robusta pada keenam blok kebun.
(c) Buatlah grafik kolom berjenjang untuk data tersebut.

Data penjualan beras (dalam satuan kuintal) dari 20 toko pengecer di Kecamatan Sukowono dan Sumberjambe pada triwulan 1 tahun 2017 adalah:

Toko Pengecer	Bulan-1	Bulan-2	Bulan-3
A	15	14	16
B	16	13	14
C	8	10	9
D	9	11	10
E	4	5	6
F	7	8	7
G	12	14	13
H	14	12	11
I	10	9	12
J	12	10	11

(a) Berapa nilai minimum dan maksimum?
(b) Berapa kelas yang harus dibuat?
(c) Berapa kelas interval?
(d) Buatlah distribusi frekuensi untuk data tersebut.

Penggunaan *pie chart* dianggap memiliki kelemahan, apakah itu?

Nilai pendapatan 20 petani di desa Sumber Agung pada bulan Juni 2017, adalah:

No.	Pendapatan (Rp)	No.	Pendapatan (Rp)	No.	Pendapatan (Rp)	No.	Pendapatan (Rp)
1	3.000.000	6	2.500.000	11	4.600.000	16	5.000.000
2	3.250.000	7	4.600.000	12	6.000.000	17	3.450.000
3	2.800.000	8	3.000.000	13	3.250.000	18	2.500.000
4	3.450.000	9	6.000.000	14	2.800.000	19	2.500.000
5	6.000.000	10	3.000.000	15	3.250.000	20	3.450.000

a. Buatlah distribusi frekuensi jumlah petani berdasar pendapatan.
b. Buatlah grafik garis untuk distribusi frekuensi tersebut.
c. Buatlah *polygon.*

BAB 4

TENDENCY CENTRAL VALUE DAN VARIABILITAS

Nilai kecenderungan utama dari data adalah berbagai jenis rata-rata dan penyerbarannya (dispersi atau variabilitas). Pengukuran nilai kecenderungan data ini banyak digunakan untuk parameter dalam pengujian hipotesis.

Kemampuan Akhir yang Diharapkan::
- Mahasiswa mampu memahami beberapa jenis nilai pemusatan data (rata-rata, modus dan median), dan variabilitas (dispersi atau penyimpangan) data. - Mampu memahami tingkat kemiringan dan keruncingan kurve yang menggambarkan pola data.

4.1. Pendahuluan

Nilai kecenderungan pemusatan (*central tendency value*) dan penyebarannya (*variability* atau *dispersion*) merupakan parameter dasar dari sekumpulan data,yaitu:

a. Rata-rata artimatika,
b. Rata-rata tertimbang,
c. Rata-rata untuk data terkelompok,
d. Median,
e. Modus,
f. Rata-rata geometrik,
g. Rata-rata harmonik,
h. Varians, standard deviasi,
i. Kemencengan (*skewness*), dan keruncingan (*kurtosis*).

4.2. Rata-rata

(1) Rata-rata aritmatika

Rata-rata aritmatika sesuai digunakan untuk menghitung rata-rata data di mana setiap data memiliki bobot yang setara, artinya x_1 sama bobotrnya dengan x_2, sama bobotnya dengan x_3, dan seterusnya.

Formula:

$$\bar{X} = \frac{© X_t}{n} \quad \text{untuk } t = 1, 2, 3, ..., n \qquad . . . . . (4.1)$$

Di mana, ΣX_t = jumlah data, contoh: jika banyaknya data (n) = 5, maka $\Sigma X_i = X_1+X_2+X_3+X_4+X_5$. n = banyaknya data yang diamati. Karakteristik rata-rata dipengaruhi oleh setiap data.

<u>Contoh-4.1</u>: Rata-rata Aritmatik

Tabel 4.1.Data X_1 dan X_2

No.	X_1	X_2
1	8	4
2	9	3

3	7	2
4	9	4
5	6	
©	39	13

$n_1 = 5$ $n_2 = 4$

$\overline{X}_1 = 39/5 = 7{,}80$

$\overline{X}_2 = 13/4 = 3{,}25$

<u>Contoh4.2</u>: Rata-rata Aritmatika.

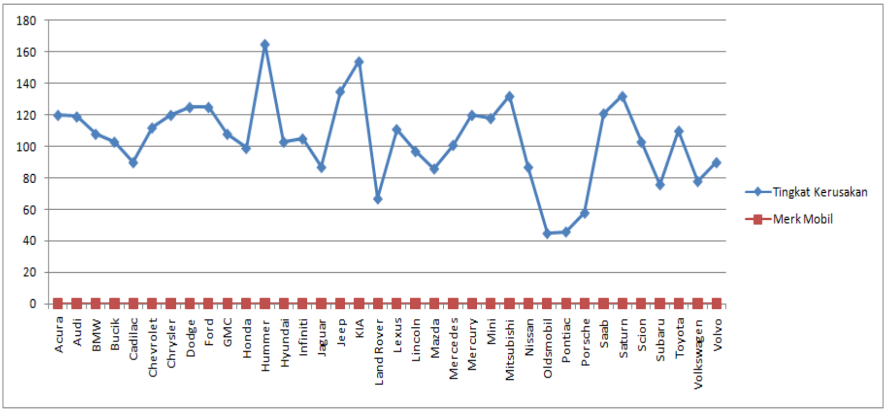
Tabel 4.2 Data Kerusakan Berbagai Merk Mobil, Tahun 2009.

Merk	Kerusakan	Merk	Kerusakan	Merk	Kerusakan
Acura	120	Infinity	105	Oldsmobil	45
Audi	119	Jaguar	87	Pontiac	46
BMW	108	Jeep	135	Porsche	58
Bucik	103	KIA	154	Saab	12 1
Cadilac	90	LandRov er	67	Saturn	13 2
Chevrolet	112	Lexus	111	Scion	10 3
Chrysler	120	Lincoln	97	Subaru	76
Dodge	125	Mazda	86	Toyota	110
Ford	125	Mercedes	101	Volkswagen	78
GMC	108	Mercury	120	Volvo	90
Honda	99	Mini	118		
Hummer	165	Mitsubishi	132		
Hyundai	103	Nissan	87		
Σ	1.497		1.400		859

ΣX_t = 1.497 + 1.400 + 859 = 3.756 unit. n = 36

$\overline{X}$=3.756/36=104,33 unit.

Grafik garis data tersebut dapat digambarkan dengan aplikasi Excel, seperti berikut[7]:



Gambar 4.1 Grafik Garis Tingkat Kerusakan Berbagai Merk Mobil.

(2) Rata-rata tertimbang

Rata-rata tertimbang sesuai digunakan untuk menghitung rata-rata data di mana setiap data memiliki bobot yang tidak setara.

Formula:

$$\boxed{\overline{X} = \frac{\sum X_t W_t}{\sum W_t}} \qquad ....(4.2)$$

Di mana W_t = bobot data-t.

---

[7]Teknik penyusunan grafik diaplikasi dengan Excel.

Contoh-4.3: Rata-rata Tertimbang (*WeightedAverage*)

Tabel 4.3 Data Keuntungan Produk dan Pesanan.

Produk	Profit (X_t)	Order (W_t)	$X_t\,W_t$
Kecil	$ 1	120	120
Medium	$ 3	60	180
Besar	$ 6	20	120
©		200	420

$\bar{X}$= 420/200 = $2,10

Mengkombinasi rata-rata.

Sebuah perusahaan mengoperasikan dua buah pabrik. Pabrik-1 memiliki 50 orang buruh (n_1 =50), Pabrik-2 memiliki 70 orang buruh (n_2 =70). Rata-rata upah buruh pada Pabrik-1 ($\bar{X}_1$)= $3.20/jam dan pada Pabrik-2 ($\bar{X}_2$) = $ 3.50/jam. Rata-rata upah buruh pada kedua pabrik itu merupakan kombinasi rata-rata. Kombinasi rata-rata dihitung dengan formula:

$$Kombinasi\ Rata-rata = \frac{n1X1+n2X2}{n1+n} \quad ....(4.3)$$

$$Kombinasi\ Rata-rata = \frac{(50)(3.20)+(70)(3.50)}{50+70} = \$\ 3.38$$

rata-rata upah pada kedua pabrik

(3) Rata-rata data terkelompok

Data terkelompok adalah data yang telah dikelompokkan dan disarikan pada sebuah tabel distribusi frekuensi[8].

Formula:

[8] Penyusunan distribusi frekuensi data dibahas pada Bab III

$$\bar{X} = \frac{\Sigma f_t X_t}{\Sigma f_t} \qquad ......(4.4)$$

Di mana f_t = frekuensi ke-t, banyaknya data pada kelompok-t

X_t= titik tengah kelompok-t, merupakan nilai batas bawah + nilai batas atas kelompok dibagi 2.

Contoh-4.4: Rata-rata data Terkelompok

Tabel 4.4 Distribusi Frekuensi Data

Kelompok	Ft	Xt	ftXt
4 s/d dibawah5	3	4,5	13,5
5 s/d dibawah6	11	5,5	60,5
6 s/d dibawah7	17	6,5	110,5
7 s/d dibawah8	16	7,5	120,0
8 s/d dibawah9	8	8,5	68,0
9 s/d dibawah10	5	9,5	47,5
Σ	60		420,0

$\bar{X}$ = 420/60 = 7,00

(4) Median

Median adalah posisi tengah data yang telah diurutkan berdasar besaran nilainya (dari kecil ke besar, *ascending*).

Formula:

$$\text{Posisi Tengah} = \frac{n+1}{2} \qquad ......(4.5)$$

Di mana n = banyaknya data yang diamati.

Syafiadi Rizki Abdila, S.T., Ph.D, Prof. Dr. Ir. Drs. Syafwandi, M.Sc
Dr. Yulasmi, S.E., M.M, Dr. Robby Dharma, S.E., M.M
Dr. Syafri Arlis, S.Kom, M.Kom

<u>Contoh-4.5</u>: Median Data Ganjil

Median untuk banyaknya data ganjil.

Tabel 4.5 Data Ganjil

No. Urutan	Data Yang Telah Diurutkan	Posisi Tengah
1	6	
2	9	
3	10	← Median
4	15	
5	20	

Posisi tengah = (5+1)/2 = 3. Median = 10

Median untuk banyaknya data genap.

<u>Contoh-4.6</u>: Median Data Genap

Tabel 4.6 Data Genap

No. Urutan	Data Yang Telah Diurutkan	Posisi Tengah
1	13	
2	14	
3	16	
3,5	…	← Median
4	20	
5	25	
6	30	

Posisi tengah = (6+1)/2 = 3,5 datanya = (16+20)/2 = 18.Median = 18.

Median untuk data terkelompok.

<u>Contoh-4.7</u>: Median Data Terkelompok

Tabel 4.7 Distribusi Frekuensi Data.

Kelompok (Kelas)	fi	Frekuensi Kumulatif (f_c)
3,00–3,49	68	68
3,50–3,99	142	210
4,00–4,49	100	310
4,50–4,99	60	
5,00–5,49	40	
5,50–5,99	20	
6,00–6,49	10	
Σ	440	

Frekuensi kumulatif (f_c) pada sebuah kelompok adalah jumlah frekuensi dari kelompok-kelompok sebelumnya sampai dengan kelompok yang bersangkutan. N = jumlah frekuensi total atau Σ f_i. Median kelompok berada pada kelompok pertama di mana frekuensi kumulatifnya $\geq$ N/2. Untuk N = 440, maka N/2 = 220. Kelompok data (4,00 – 4,49) merupakan kelompok data pertama yang frekuensi kumulaitf > 220; maka kelompok tersebut merupakan median kelompok. Frekuensi pada kelompok (4,00 – 4,49) tersebut = 100. Frekuensi kumulatif pada kelompok sebelumnya = 210, maka selisih untuk mencapai 220 tersebut = 220 – 210 = 10. Median pada kelompok (4,00 – 4,49) kemudian dapat dihitung sebagai:

Median kelompok = Batas bawah + (selisih/frekuensi dalam kelompok median dikalikan interval kelas)

Secara matematis, formulanya dapat dituliskan:

$$\boxed{Md_{kelompok} = L_m + r/f_m \text{ x } c} \quad ......(4.6)$$

Di mana

Md = median

L_m = batas bawah pada kelompok di mana median berada, pada contoh ini $L_m = 4$

r = selisih frekuensi kumulatif pada kelompok sebelumnya kepada N/2, r = 220 – 210 = 10,

f_m= frekuensi pada kelompok di mana median berada, pada contoh ini = 100,

c = interval kelas = beda antara batas bawah dengan batas atas = 0,50

Median kelompok = 4 + (10/100) x 0,50 = 4,05.

<u>Contoh-4.8</u>: Median Data Terkelompok

Tabel 4.8 Distribusi Frekuensi Data.

Kelompok(Kelas)	fi	fc
0,00–4,99	16	16
5,00–9,99	35	51
10,00–14,49	40	91
15,00–19,49	39	130
20,00–24,99	26	156
Σ	156	

N/2=156/2 =78. Frekuensi kumulatif yang pertama melebihi 78 adalah = 91, maka median kelompok berada pada posisi kelompok (10,00– 14,99). Frekuensi pada kelompok tersebut (f_m) = 40. Frekuensi kumulatif pada kelompok sebelumnya = 51, untuk mencapai 78, ada selisih = 78– 51 = 27. Selisih batas bawah dengan batas atas (c) = 5. Selanjutnya menghitung median kelompok: Md =10 + (27/40) x 5,00 = 13,38.

Modus

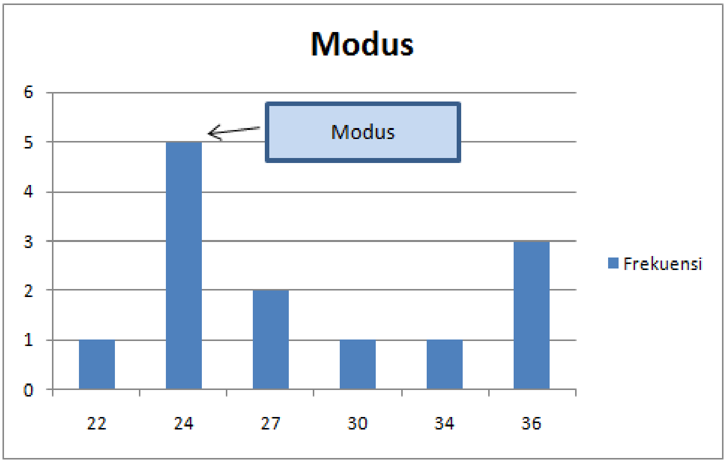
Modus berasal dari kata *mode*, artinya data yang paling sering muncul pada serangkaian data.

Pada sebuah toko pakaian, penjualan terbanyak adalah kemeja dengan ukuran 14. Maka modusnya adalah = 14.

Serangkaian data yang telah diurutkan adalah sebagai berikut:

22, 24, 24, 24, 24, 24, 27, 27, 30, 30, 36, 36, 36

Paling banyak terjadi adalah 24 (sebanyak 5 kali). Modus rangkaian data = 24.

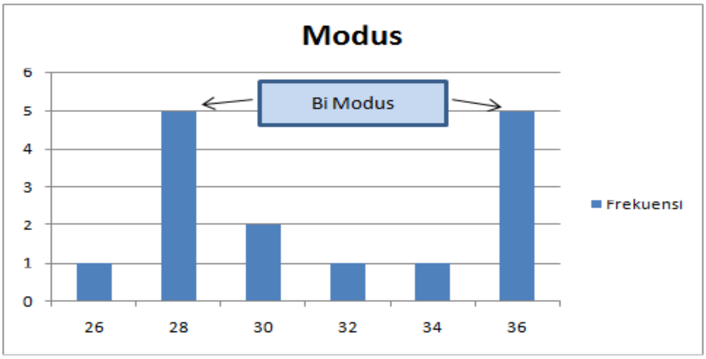


Gambar 4.2 Distribusi Data Dengan Modus Tunggal.

Perhatikan data berikut:

26,28,28,28,28,28,30,30,32,34,36,36,36,36,36

Paling banyak terjadi adalah angka 28 dan 36 (masing-masing terjadi 5 kali). Maka data di atas memiliki dua modus, atau *bi-modus*.



Gambar 4.3 Distribusi Data dengan Bi-Modus.

Modus untuk data terkelompok

Formula untuk menghitung modus:

$$M_o = L_o + \frac{d_1}{(d_1+d_2)} C \qquad \text{......(4.7)}$$

L_o= batas bawah kelompok di mana modus itu berada.

d_1= frekuensi pada kelompok modus dikurangi dengan frekuensi pada kelompok sebelumnya,

d_2= frekuensi pada kelompok modus dikurangi dengan frekuensi pada kelompok sesudahnya,

C= interval kelas.

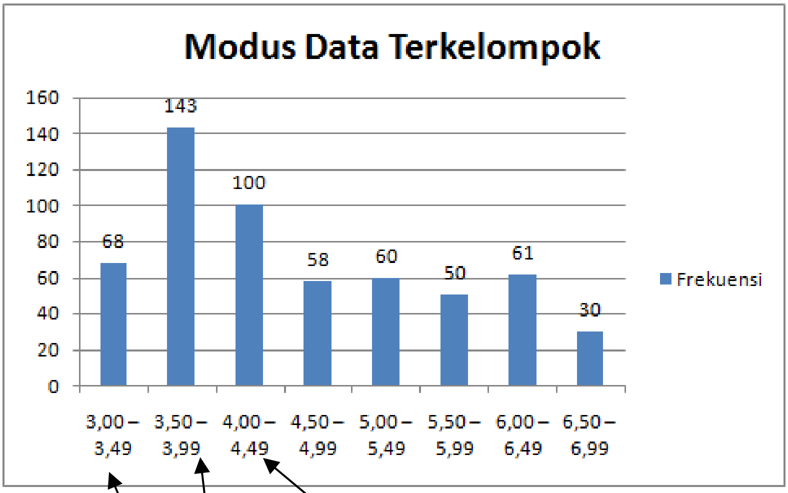
Contoh-4.9: Modus Data Terkelompok

Tabel 4.9 Distribusi Frekuensi Data.

No.	Kelompok	Frekuensi
1	3,00–3,49	68
2	3,50–3,99	143
3	4,00–4,49	100
4	4,50–4,99	58

5	5,00–5,49	60
6	5,50–5,99	50
7	6,00–6,49	61
8	6,50–6,99	30

Untuk mempermudah dalam menentukan kelompok modus, kelompok sebelumnya dan kelompok sesudahnya; digunakan pendekatan grafis kolom dari frekuensi distribusi di atas. Grafik kolomnya adalah:



Kelompok Modus

Kelompok Sebelumnya

Kelompok Sesudahnya

Gambar 4.4 Grafik Kolom Distribusi Frekuensi

L_o= 3,50

d_1 = selisih frekuensi kelompok modus dengan frekuensi kelompok sebelumnya = 143 – 68 = 75.

d_2 = selisih frekuensi kelompok modus dengan frekuensi kelompok sesudahnya = 143 – 100 = 43. C = 0,50

M_o = 3,50 +75/(75 + 43) x 0,50 = 3,82

Contoh-4.10: Modus Data Terkelompok

Tabel 4.10 Distribusi Frekuensi Data.

No.	Kelompok	Frekuensi
1	10,0–14,9	94
2	15,0–19,9	102
3	20,0–24,9	70

←Kelompok Modus

L_o= 15,0 d_1 = 102 – 94 = 8, d_2 = 102 – 70 = 32, C = 5

M_o = 15,0 + (8/40) x 5 = 16,00

(6) Rata-rata geometrik

Rata-rata geometrik adalah ukuran rata-rata untuk data rasio atau persentase pertumbuhan. Ahli lain sering menyatakan sebagai rata-rata ukur. Formula untuk menghitung rata-rata geometrik:

$$G_m = \sqrt[n]{\pi X_t} \qquad \text{.......(4.8)}$$

Di mana, πX_t adalah perkalian data-t, contoh:jika ada 4 data, maka $\pi X_t = X_1.X_2.X_3.X_4$.

Contoh-4.11: Rata-rata Geometrik

Tabel 4.11 Data Ratio.

Tahun	Ratio Pertumbuhan (X_t)
2010	95,60%
2011	112,00%
2012	78,03%
2013	101,80%

Rata-rata ratio pertumbuhan dalam empat tahun:

$G_m = \sqrt[4]{(95,60\%) \times (112,60\%) \times (78,03\%) \times (101,80)} = 96,03\%$

Contoh-4.12: Rata-rata Geometrik

Tabel 4.12 Data Pertumbuhan Pelanggan

Tahun	Jumlah Pelanggan	Pertumbuhan (%)
2010	101.432	-
2011	123.450	21,71%
2012	132.567	7,39%
2013	140.668	5,76%

Rata-rata pertumbuhan pelanggan dalam empat tahun:

$G_m = \sqrt[3]{(21,71\%) \times (7,39\%) \times (5,76\%)} = 9,74\%$

Jika ada data yang bernilai negatif, sehingga hasil perkaliannya menghasilkan angka negatif, dan tidak mungkin diakarkan; maka aplikasi formula G_m mengalami perubahan sebagai berikut:

Contoh-4.13: Rata-rata Geometrik

Tabel4.13 Data Pertumbuhan Penduduk (dalam juta jiwa)

Tahun	Penduduk	Pertumbuhan (%)
2009	100,00	-
2010	101,60	1,60
2011	102,90	1,28
2012	99,00	-3,79
2013	101,30	2,32
	πX_t	-18,03

πX_t = (1,60%) x (1,28%) x (-3,79%) x (2,32%) = -18,03%; angka ini tidak mungkin diakarkan pangkat 4, karena bertanda negatif. Untuk itu, aplikasi G_m dimodifikasi menjadi:

$$G_m = \sqrt[4]{(1{,}60\% + 1) \times (1{,}28\% + 1) \times (-3{,}79\% + 1) + (2{,}32\% + 1)} - 1 = 100{,}32\% - 1 = 0{,}32\%$$

(7) Rata-rata *Harmonic*[9]

Rata-rata *harmonic* adalah rata-rata yang dihitung dari banyaknya data dibagi dengan jumlah *quotient* data.

$$X = \frac{n}{\Sigma(1/X_t)} \quad \text{.......(4.9)}$$

<u>Contoh-4.14</u>: Rata-rata *Harmonic*

Penjualan sarung dalam satu bulan sebagai berikut:
Minggu-1: 20 helai dengan harga Rp 60.000/helai
Minggu-2: 15 helai dengan harga Rp 70.000/helai
Minggu-3: 25 helai dengan harga Rp 75.000/helai
Minggu-4: 15 helai dengan harga Rp 80.000/helai
Rata-rata *harmonic* harga sarung per helai dalam satu bulan:

$$\bar{X} = \frac{4}{1/60000 + 1/70000 + 1/75000 + 1/80000} = Rp\ 70.440{,}25$$

4.3. Standard Deviasi (Penyimpangan)

Penyimpangan merupakan variabilitas atau penyebaran data terhadap rata-ratanya.Makin besar penyimpangannya, makin besar penyebaran data. Penyebaran atau penyimpangan diukur dengan varians dan standard deviasi.

---

[9]Rata-rata *harmonic* sangat jarang digunakan.

Formula untuk menghitung varians adalah:

$$S^2 = \frac{\Sigma(X_i - \bar{X})^2}{n} \quad .....(4.10)$$

Contoh-4.15: **Varians**

Hitung varians dari 6, 4, 8 dan10. n = 4, rata-rata = 28/4 = 7

$\Sigma (X_i - \bar{X})^2 = (6–7)^2 + (4–7)^2 + (8–7)^2 + (10–7)^2 = 20$

$s^2 = 20/4 = 5,0$

Standard deviasi adalah akar kuadrat dari varians.
Formula untuk menghitung standard deviasi sampel adalah:

$$s = \sqrt{\frac{\Sigma(X_i - \bar{X})^2}{n}} \quad ....(4.11)$$

Dengan cepat dapat dihitung: s = √5,0 = 2,24

Standard deviasi juga dapat dihitung dengan formula:

$$s = \sqrt{\frac{©{X_t}^2}{n} - \left(\frac{©X_t}{n}\right)^2} \quad ....(4.12)$$

Tabel 4.14 Lembar Kerja Perhitungan Standard Deviasi.

X_t	Xt2	Rata-rata kuadrat = $(\Sigma X_t^2)/n$ = 216/4 = 54
6	36	Kuadrat rata-rata = $(©X_t/n)^2 = 7^2 = 49$
4	16	Rata-rata = 28/4 = 7
8	64	
10	100	s =$\sqrt{\{\Sigma Xt^2/n - (\Sigma X_t/n)^2\}}$= √(54–49) = 2,24
Σ	216	

Hasil perhitungan standard deviasi dengan formula (4.11) sama dengan hasil formula(4.12).

Standard deviasi untuk data terkelompok

Formula:

$$s = \sqrt{\frac{\copyright(f_t X_t)^2}{\copyright f_t} - \left(\frac{\copyright f_t X_t}{\copyright f_t}\right)^2} \qquad \text{. . . . . (4.13)}$$

<u>Contoh-4.16</u>: Standard deviasi untuk data terkelompok.

Tabel 4.15 Frekuensi Distribusi Data.

Kelompok	Frekuensi (f_t)	Nilai Tengah	$f_t X_t$	X_t^2	$f_t X_t^2$
0 - kurang 10	2	5	10	25	50
10 - kurang 20	6	15	90	225	1350
20 - kurang 30	16	25	400	625	10000
30 - kurang 40	12	35	420	1225	14700
40 - kurang 50	7	45	315	2025	14175
50 - kurang 60	4	55	220	3025	12100
60 - kurang 70	2	65	130	4225	8450
70 - kurang 80	1	75	75	5625	5625
©	50		1660		66450

© $(f_t X_t^2)$/©f_t = 66450/50 = 1329.

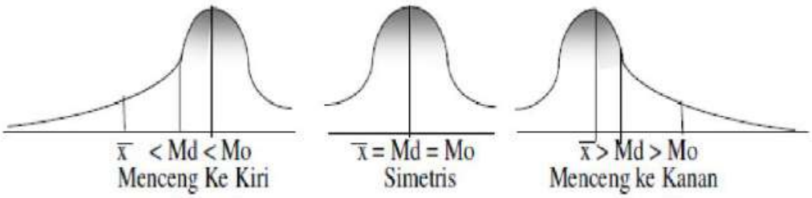
©{($f_t X_t$)/©f_t}^{2} = (1660/50)^2 = 1102,24

Dengan formula (4.13), standard deviasi untuk data terkelompok di atas dapat dihitung:

s = √ (1329 – 1102,24) = 15,06.

4.4. Kemencengan Kurve (*Skewness*)

Bentuk kurve ada tiga yaitu: (1) menceng ke kiri; (2) simetris; dan (3) menceng ke kanan.



Gambar 4.4 Bentuk Kurve Distribusi Frekuensi

Tingkat kemencengan dapat diukur sebagai:

$$Skewness = \frac{X - Mo}{s} \quad .....(4.14)$$

<u>Contoh-4.17</u>: Kemencengan Kurve Distribusi Data

Tabel4.16 Data untuk Menghitung Kemencengan Kurve

7	5	9	4	11
6	4	10	5	12
7	6	12	6	8

$\bar{X}$ = 7,47 Sd = 2,75 Mo = 6 Md = 7

Skewness = (7,47 – 6)/2,75 = 0,5345 → positif, kurve menceng ke kanan (terlihat juga bahwa X > Md > Mo).

Atau dengan formula:

$$Skewness = \frac{3(\bar{x} - Md)}{s} \quad .....(4.15)$$

Skewness (untuk data tidak terkelompok) dapat pula dihitung dengan menggunakan *moment* ketiga (α_3). Formulanya adalah:

$$\alpha 3 = \frac{M_3}{s^3} = \frac{1}{ns^3} \sum_{t=1}^{n} (Xt - \bar{X})^3 \quad ...... (4.16)$$

Jika α_3 positif, maka kurve menceng ke kanan, dan sebaliknya jika α_3 negatif, maka kurve menceng ke kiri. Jika $\alpha_3 = 0$, maka kurve simetris.

Contoh-4.18: *Skewness* Data Tidak Terkelompok

Kembali pada data Tabel 4.16: di mana $\bar{X} = 7,47$. s = 2,75

Tabel 4.17 Data untuk Menghitung Kemencengan Kurve

7	5	9	4	11
6	4	10	5	12
7	6	12	6	8

$\Sigma(X_t - \bar{X})^3 = 127,45$ $\alpha_3 = (1/312)(127,45) = 0,409373$ (kurve menceng

ke kanan).

$n(s)^3 = 15(2,75^3) = 312$

Aplikasi perhitungan di atas dengan Excel dapat dilakukan secara lebih mudah dibanding perhitungan manual.

	A	B	C	D	E
1	7	5	9	4	11
2	6	4	10	5	12
3	7	6	12	6	8

$\bar{X}$ = 7,47 ← =average(A1:E3)

s = 2,75 ← =stdev(A1:E3)

Untuk data terkelompok, sebanyak k kelompok, formula *moment* ke-3 adalah :

$$\alpha 3 = \frac{M_3}{s^3} = \frac{1}{ns^3} \sum_{t=1}^{k} f_t (Xt - \bar{X}) \qquad \text{. . . . (4.17)}$$

Contoh-4.19: *Skewness* Data Terkelompok

Tabel 4.18 Lembar Kerja Untuk Menghitung *Skewness* Data Terkelompok

Kelas	Titik Tengah (Xt)	ft	fc	ftXt	$ftXt^2$	$ft(Xt - X)^3$
110-118	114	6	6	684	77976,00	-68343,75
119-127	123	10	16	1230	151290,00	-24603,75
128-136	132	67	83	8844	1167408,00	-6105,38
137-145	141	24	107	3384	477144,00	2187,00
146-154	150	10	117	1500	225000,00	24603,75
155-163	159	9	126	1431	227529,00	102515,63
164-172	168	4	130	672	112896,00	125023,50
Σ		130		17745	2439243,00	155277,00

X_t = *classmark* atau titik tengah kelas. Banyaknya kelas = k = 7. Kelas interval = c = 8. f_c = frekuensi kumulatif. N = Σf_t = 130.

Rata-rata kelompok:

$$\bar{X} = \frac{\Sigma f_t X_t}{\Sigma f_t} = 17745/130 = 136,500$$

Median kelompok:

$$Mo_{kelompok} = L_m + r/f_m \times c$$

N/2 = 130/2 = 65. Kelas (128-136) merupakan kelompok pertama dengan f_c yang melebihi 65 (yaitu = 83); maka kelompok ini merupakan kelas yang mengandung median, frekuensi pada kelas ini = f_m = 67. Frekuensi kumulatif pada kelas sebelumnya adalah = 16, maka untuk mencapai 65, ada selisih = r = 65 - 16 = 49.

$$Mo_{kelompok} = 128 + 49/67 \times 8 = 133,85$$

Standard deviasi kelompok dengan menggunakan formula (4.13):

$$s = \sqrt{\frac{© (f_t X_t^2)}{© f_t} - \left(\frac{© f_t X_t}{© f_t}\right)^2}$$

$$s = \sqrt{\frac{2439243}{130} - \left(\frac{17745^2}{130}\right)} = 11{,}45$$

$$\alpha_3 = \frac{1}{130(11{,}45)^3} 155277 = 0{,}7952$$

Kurve distribusi frekuensinya menceng ke kanan; makin tinggi α_3, makin menceng kurvenya.

Jika kelas interval (= C) sama, maka *skewness* dapat dihitung sebagai:

$$\langle_3 = \frac{C^3}{s^3} \left\{ \frac{1}{N} \overset{k}{\underset{t=1}{©}} f_t d_t^3 - 3\left(\frac{1}{N} \overset{k}{\underset{t=1}{©}} f_t d_t^2\right)\left(\frac{1}{N} \overset{k}{\underset{t=1}{©}} f_t d_t\right) + 2\left(\frac{1}{N} \overset{k}{\underset{t=1}{©}} f_t d_t\right)^3 \right\} \quad \ldots(4.18)$$

Di mana,

s = standard deviasi data terkelompok,

C = kelas interval, k = banyaknya kelas,

f_t = kelas interval ke-t,

d_t = kode kelas interval ke-t, dan

N = © f_t.

Aplikasi formula (4.18) untuk data pada Tabel-4.18 adalah sebagai berikut:

Tabel-4.19 Lembar Kerja Untuk Menghitung *Skewness* Data Terkelompok.

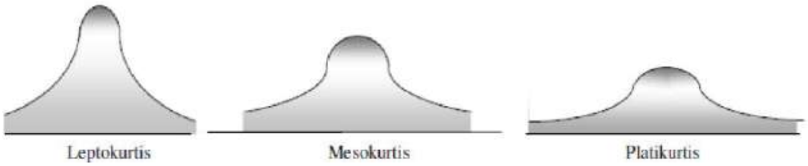
Kelas	Titi k Tengah (x t)	ft	ftXt	dt	ftdt	ftdt2	ftdt3	ftdt4
110-118	114	6	684	-3	-18	54	-162	486
119-127	123	10	1230	-2	-20	40	-80	160
128-136	132	67	8844	-1	-67	67	-67	67
137-145	141	24	3384	0	0	0	0	0
146-154	150	10	1500	1	10	10	10	10
155-163	159	9	1431	2	18	36	72	144
164-172	168	4	672	3	12	36	108	324
Σ		130	17745		-65	243	-119	1191

$_{\text{kelompok}} = 11{,}45.$

$$\alpha_3 = \frac{8^3}{(11{,}45)^3}\left\{\left(\frac{1}{130}\right)(-119) - 3\left(\frac{1}{130}\right)(243)\left(\frac{1}{130}\right)(-65) + 2\left(\frac{1}{130}(-65)\right)^3\right\} = 0{,}5588$$

4.5. Keruncingan Kurve (*Kurtosis*)

Dilihat dari keruncingannya, bentuk kurve distribusi frekuensi terbagi menjadi tiga, yaitu: (1) leptokurtis, (2) mesokurtis dan (3) platikurtis.



Gambar 4.5 Keruncingan Kurve Distribusi Frekuensi

Syafiadi Rizki Abdila, S.T., Ph.D, Prof. Dr. Ir. Drs. Syafwandi, M.Sc
Dr. Yulasmi, S.E., M.M, Dr. Robby Dharma, S.E., M.M
Dr. Syafri Arlis, S.Kom, M.Kom

Formula untuk menghitung kurtosis adalah:

$$\alpha 4 = \frac{\frac{1}{n}\sum_{t=1}^{n}(X_t - X)^4}{s^4}$$ → data tak terkelompok… (4.19)

$$\alpha 4 = \frac{\frac{1}{n}\sum_{t=1}(M_t - \bar{X})^4}{s^4}$$ → data terkelompok. . . . (4.20)

M_t = titik tengah kelas-t pada data terkelompok atau X_t pada data tak terkelompok.

Jika kelas interval pada setiap kelompok itu sama, maka formula (4.20) dapat diganti dengan formula (4.21).

Contoh-4.20: Kurtosis Data Terkelompok

Tabel 4.20 Lembar Kerja Untuk Menghitung Kurtosis Data Terkelompok

Kelas	TitikTe-ngah (xt)	ft	ftxt	dt	ftdt	ftdt2	ftdt3	ftdt4
110-118	114	6	684	-3	-18	54	-162	486
119-127	123	10	1230	-2	-20	40	-80	160
128-136	132	67	8844	-1	-67	67	-67	67
137-145	141	24	3384	0	0	0	0	0
146-154	150	10	1500	1	10	10	10	10
155-163	159	9	1431	2	18	36	72	144
164-172	168	4	672	3	12	36	108	324
Σ		130	17745		-65	243	-119	1191

Jika kelas intervalnya sama, formula di atas berubah menjadi formula (4.21) berikut:

$$\alpha_4 = \frac{C^4}{s^4}\left\{\frac{1}{N}\sum_{t=1}^{k} f_t d_t^4 - 4\left(\frac{1}{N}\sum_{t=1}^{k} f_t d_t^3\right)\left(\frac{1}{N}\sum_{t=1}^{k} f_t d_t\right) + 6\left(\frac{1}{N}\sum_{t=1}^{k} f_t d_t^2\right)\left(\frac{1}{N}\sum_{t=1}^{k} f_t d_t\right)^2 - 3\left(\frac{1}{N}\sum_{t=1}^{k} f_t d_t\right)^4\right\}$$

$$\alpha_4 = \frac{8^4}{11{,}45^4}\left\{\left(\frac{1191}{40}\right) - 4\left(\frac{-119}{40}\right)\left(\frac{-65}{40}\right) + 6\left(\frac{243}{40}\right)\left(\frac{-65^2}{40}\right) - 3\left(\frac{-65^4}{40}\right)\right\} = 2{,}3704$$

Jika $\alpha_4 \geq 3$, leptokurtis, Jika $\alpha_4 = 3$, mesokurtis, Jika $\alpha_4 < 3$, platikurtis.

Distribusi frekuensi data pada Tabel 4.20 tersebut terkatagori platikurtis.

4.6. Rangkuman

*Tendency central value*dan penyebaran (dispersi) adalah nilai kecenderungan terpusat dari sekelompok data, ini merupakan parameter utama dari data tersebut. Parameter tersebut adalah: rata-rata, modus, median, penyimpangan (variabilitas, dispersi), keruncingan pola (*kurtosis*) dan kemencengan/kemiringan pola (*skewness*). Jenis rata-rata adalah: (a) Rata-rata artimatika; (b) Rata-rata tertimbang: (c) Rata-rata untuk data terkelompok; (d) Median; (e) Modus; (f) Rata-rata geometric; dan (g) Rata-rata harmonik. Penyebaran atau dispersi data adalah: Varians dan standard deviasi. Grafik pola kemencengan data (*skewness*),dan keruncingan data (*kurtosis*).

Formula untuk menghitung rata-rata adalah:

(a) Rata-rata artimatika:

$$\bar{X} = \frac{\sum X_t}{n} \qquad \text{untuk } t = 1, 2, 3, ..., n \qquad ..... (4.1)$$

Syafiadi Rizki Abdila, S.T., Ph.D, Prof. Dr. Ir. Drs. Syafwandi, M.Sc
Dr. Yulasmi, S.E., M.M, Dr. Robby Dharma, S.E., M.M
Dr. Syafri Arlis, S.Kom, M.Kom

(b) Rata-rata tertimbang:

$$\bar{X} = \frac{\sum X_t W_t}{\sum W_t} \quad .....(4.2)$$

(c) Kombinasi rata-rata:

$$\text{Kombinasi Rata-rata} = \frac{\overline{n1X1} + \overline{n2X2}}{n1 + n2} \quad .....(4.3)$$

(d) Rata-rata data terkelompok:

$$\bar{X} = \frac{\Sigma f t X t}{\Sigma f_t} \quad .....(4.4)$$

(e) Median:

$$\text{Posisi Tengah} = \frac{n + 1}{2} \quad .....(4.5)$$

Median data terkelompok:

$$Mo_{kelompok} = L_m + r/f_m \text{ x } c \quad ......(4.6)$$

(f) Modus:

Modus data terkelompok:

$$M_o = L_o + \frac{d_1}{(d_1 + d_2)} C \quad .....(4.7)$$

(g) Rata-rata Geometrik:

$$Gm = \sqrt[n]{(\pi Xt)} \quad .....(4.8)$$

(h) Rata-rata Harmonik:

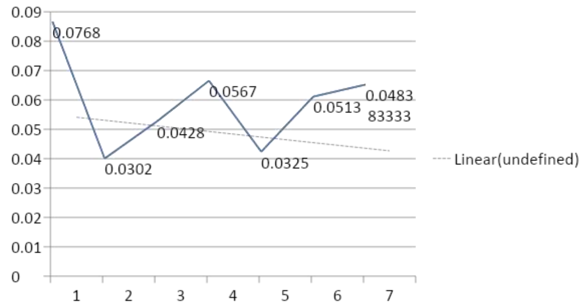
$$\bar{X} = \frac{n}{\Sigma(1/Xt)} \quad .....(4.9)$$

Formula untuk menghitung varians:

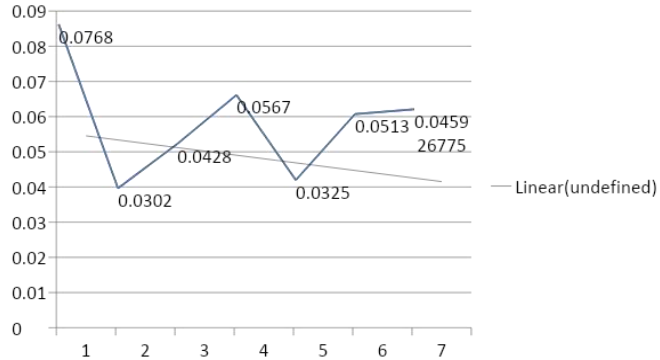
$$s2 = \frac{\sum(Xi - \bar{X})^2}{n} \quad .....(4.10)$$

4.7. Diskusi

Dalam aplikasinya, masih ada beberapa peneliti yang tidak tepat dalam menghitung rata-rata dari sekelompok data. Untuk datadengan nilai persentase (%), rata-ratanya dihitung dengan rata-rata aritmatik, di mana seharusnya menggunakan rata-rata geometrik. Rata-rata pertumbuhan ekonomi yang dinyatakan dalam persentase, seharusnya dihitung dengan rata-rata geometrik; walaupun mungkin hasilnya tidak jauh berbeda; namun hal tersebut tetap tidak benar. Contoh: pertumbuhan ekonomi dalam enam tahun adalah: 7,68%; 3,02%; 4,28%; 5,67%; 3,25% dan 5,13%. Dengan rata-rata geometrik, rata-rata pertumbuhan ekonomi dalam empat tahun empiris = 4,59%; tetapi dengan rata-rata aritmatik = 4,84%.



Rata-rata aritmatik lebih jauh terhadap garis trend-nya, dibanding rata-rata geometrik.



4.8. Referensi

Balakrishnan, N., and V. B. Nevzorov, 2003, ***A Primer on Statistical Distributions***,John Wiley and Son Inc., New York

Doane, David, P., and Lori E. Seward, 2007, ***Applied Statistics in Business andEconomics***, McGraw Hill/Irwin Series, New York

Kirkpatrick, E., G., 2009, ***Introductory Statistics and Probability for Engineer,Science and Technology,*** Prentice-Hall Inc., Englewood Cliffs, N. J.

4.9. Latihan Soal

Hitunglah rata-rata tertimbang harga beras per kwintal:

JenisBeras	Harga (Rp/kw)	Penjualan (kw)
Kepala IR64	1,200,000.00	3,2
Kepala Pandawangi	1,250,000.00	4,0
Broken IR64	1,000,000.00	5,1
Broken Pandawangi	1,150,000,00	6,0
Semi Kepala IR64	1,100,000.00	4,8
Semi Kepala Pandanwangi	1,125,000.00	5,1

Hitunglah rata-rata tinggi badan mahasiswa:

Kelompok	Frekuensi	Kelompok	Frekuensi
150–154	12	165–169	20
155–169	14	170–174	8
160–164	16	175-179	3

Hitunglah rata-rata pertumbuhan laju inflasi dalam 6 tahun yang lalu:

Tahun	Tingkat Inflasi(%)
2003	4,32
2004	5,00
2005	4,45
2006	3,88
2007	4,65
2008	4,32

Hitunglah modus kelompok dari data berikut:

No.	Kelompok	Frekuensi
1	3,00–3,49	42
2	3,50–3,99	76
3	4,00–4,49	89
4	4,50–4,99	35
5	5,00–5,49	38
6	5,50–5,99	41
7	6,00–6,49	45
8	6,50–6,99	23
9	7,00–7,49	14
10	7,50–7,99	32

Hitung median kelompok dari data berikut:

Kelompok(Kelas)	f_i
4,00–4,49	38
4,50–4,99	34
5,00–5,49	109
5,50–5,99	100
6,00–6,49	80
6,50–6,99	43
7,00–7,49	20

Hitung standard deviasi dari distribusi frekuensi:

No.	Kelompok	f_t
1	0–kurang 10	16
2	10–kurang 20	18
3	20–kurang 30	19
4	30–kurang 40	12
5	40–kurang 50	17
6	50–kurang 60	16
7	60–kurang 70	22
8	70–kurang 80	16
9	80–kurang 90	19
10	90–kurang 100	21

Hitung kemencengan kurve distribusi frekuensi data di bawah ini dengan menggunakan formula α3:

Kelas	f_t
120-125	13
126-131	11
132-137	19
138-143	20
144-149	15
150-155	13
156-161	12
162–167	16

Hitung keruncingan kurve distribusi frekuensi data di bawah ini dengan menggunakan formula α4:

No.	Kelas	f_t
1	110-125	13
2	126-141	25
3	142-157	19
4	158-173	22
5	174-189	15
6	190-205	24
7	206-221	28
8	222-237	32

BAB 5

TEORI PROBABILITAS

Probabilitas, secara teori terdiri atas probabilitas klasik dan probabilitas dengan pola distribusi tertentu, seperi: Binomial, Multinomial, Poisson, Hypergeometrik, Geometrik, Normal dan banyak bentuk distribusi lainnya. Probabilitas klasik dipelajari sebagai dasar pemahaman probabilitas secara awal. Namun pada tataran praktek, probabilitas klasik ini jarang digunakan.

Kemampuan Akhir yang Diharapkan:
- Mahasiswa mampu menjelaskan distribusi probabilitas (untuk variabel diskret) dan padatan probabilitas (untuk variabel kontinyu). - Mampu menghitung probabilitas Binomial, Poisson, Hipergeometrik, Geometrik, Multinomial, dan probabilitas Normal.

5.1 Pendahuluan

Distribusi probabilitas adalah sebuah penyebaran probabilitas sebuah variabel random memiliki sebuah nilai pada suatu rentang nilai. Random variabel bisa berbentuk variabel kontinyu atau berbentuk diskret. Variabel diskret adalah variabel dengan nilai yang terbatas (*finite*), contoh : data tingkat absensi buruh yang diukur dalam hari per bulan merupakan variabel diskret, tetapi berat seseorang merupakan variabel kontinyu (*infinite*).

Dalam bentuk persamaan matematik distribusi probabilitas variabel random dapat dituliskan dalam bentuk umum:

$f(xi) = 1/k; 0 < f(xi) < 1$; dan $S f(xi) = 1$

Contoh : $f(x) = 1/6$ untuk $x = 1, 2, 3, 4, 5$ dan 6.

Untuk variabel diskret penyebaran probabilitas disebut sebagai distribusi probabilitas (*probability distribution*), sedang untuk variabel kontinyu disebut dengan padatan probabilitas (*probability density*).

5.2 Distribusi Probabilitas

Bentuk distribusi probabilitas yang banyak diamati oleh peneliti adalah:

a. Distribusi Binomial,
b. Distribusi Hipergeometrik,
c. Distribusi Poisson,
d. Distribusi Geometrik, dan
e. Distribusi Multinomial.

1. Distribusi Binomial

Asumsi yang mendasari distribusi binomial adalah:

(1) Hanya ada dua kemungkinan hasil pada setiap percobaan,

(2) Probabilitas setiap kemungkinan hasil adalah sama pada setiap percobaan,

(3) Ada sejumlah n-percobaan di mana n adalah sebuah konstan,

(4) n-percobaan bersifat independen.

Formula distribusi probabilitas Binomial :

$$b(n;x;p) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{untuk } x = 0, 1, 2, \ldots\ldots n$$

. . . . . (5.1)

Di mana, $\binom{n}{x}$ adalah koefisien binomial atau koefisien Bernoulli, yang dihitung sebagai:

$$\frac{n!}{x!\,(n-x)!}$$

n = banyaknya percobaan,
X = hasil atau *outcome*,
p = probabilitas *outcome*-X.

Teorema yang berlaku pada Distribusi Binomial adalah :
(1) $b(x;n;p) = b(n-x;n;1-p)$
(2) $B(x;n;p) = 1 - B(n-x-1;n;1-p)$
(3) $b(x;n;p) = B(x;n;p) - B(x-1;n;p)$
(4) $b(x;n;p) = B(n-x;n;1-p) - B(n-x-1;n;1-p)$

Formula parameter pada Distribusi Binomial :

Rata-rata : $\boxed{\mu = n.p}$ . . . . . (5.2)

Varians : $\boxed{\sigma^2 = n.p.(1-p)}$ . . . . . (5.3)

<u>Contoh-5.1</u>: Distribusi Binomial-1

Dinyatakan bahwa jika 60,00% dari seluruh instalasi listrik diganti dengan sistim listrik matahari, maka akan ada penurunan biaya operasional paling sedikit sepertiganya. Berapa probabilitas penurunan biaya operasional tersebut dalam :

(a) empat instalasi dari lima instalasi,
(b) paling sedikit empat instalasi dari lima instalasi.
Jawaban :
(a) substitusikan : x = 4, n = 5, dan p = 0,60

$$b(4;5;0.60) = \binom{5}{4} 0,60^4 (1 - 0,60)^{5-4} = 0,259$$

(b) substitusikan : x = 5, n = 5, dan p = 0,60

$$b(5;5;0.60) = \binom{5}{5} 0,60^5 (1-0,60)^{5-5} = 0,078$$

Sehingga, jawabannya adalah :
b(4;5;0,60) + b(5;5;0,60) = 0,259 + 0,078 = 0,337

Jika n merupakan angka yang besar, maka perhitungan probabilitas binomial menjadi lebih rumit, untuk itu dapat dimanfaatkan Tabel Binomial.

Tabel Binomial merupakan nilai-nilai probabilitas kumulatif:

$$B(x;n;p) = \overset{x}{\underset{k=:0}{©}} \; b(x;n;p) \text{ untuk } x = 0, 1, 2, \ldots n$$

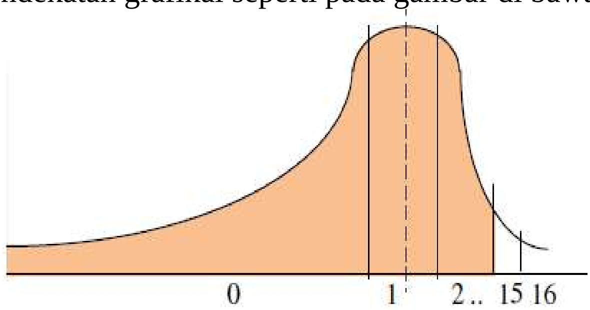
<u>Contoh-5.2</u> : Distribusi Binomial-2

Probabilitas sebuah file tidak terbaca pada saat *loading* = 0,05. Dari 16 files, berapa kemungkinan tidak terbaca pada saat *loading* :

(a) paling banyak dua files?
(b) paling sedikit empat files?
Jawaban :
(a) Paling banyk 2 files
Tabel Binomial menunjukkan bahwa B(2;16;0,05) = 0,9571.

Syafiadi Rizki Abdila, S.T., Ph.D, Prof. Dr. Ir. Drs. Syafwandi, M.Sc
Dr. Yulasmi, S.E., M.M, Dr. Robby Dharma, S.E., M.M
Dr. Syafri Arlis, S.Kom, M.Kom

Untuk memudahkan perhitungan ini, bisa dimanfaatkan pendekatan grafikal seperti pada gambar di bawah ini :



Gambar 5.1 Distribusi Binomial *Loading File*

Atau dihitung secara manual sebagai berikut :

B(2;16;0,05) = b(0;16;0,05) + b(1;16;0,05) + b(2;16;0,05) =

$$b(0;16;0,05) = \binom{16}{0} 0,050\ (1-0,05)^{16} = 0,4401$$

$$b(1;16;0,05) = \binom{16}{0} 0,051\ (1-0,05)^{15} = 0,3706$$

$$b(2;16;0,05) = \binom{16}{0} 0,052\ (1-0,05)^{14} = 0,1463$$

Jumlah = 0,9571

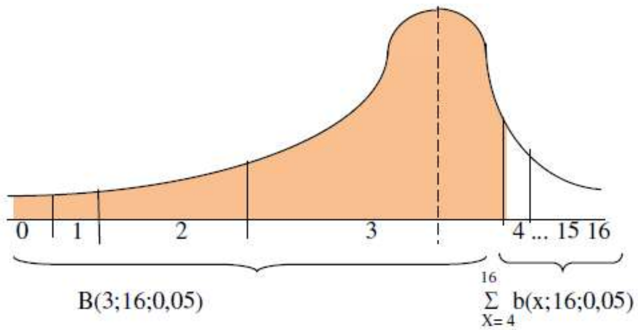
(b) Paling sedikit 4 files

16
© b(X;16;0,05) = 1 – B(3;16;0,05)
X= 4

Pada Tabel Binomial, B(3;16;0,05) = 0,9930, maka :

16
© b(X;16;0,05) = 1 – 0,9930 = 0,0070
X= 4

Pendekatan grafikal dapat mempermudah perhitungan probabilitas di atas sebagai berikut:



Gambar 5.2 Distribusi Binomial*Loading File*

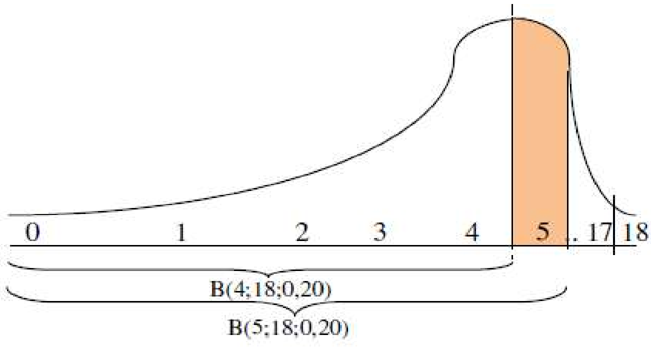
<u>Contoh-5.3</u> : Distribusi Binomial-3

Probabilitas seseorang untuk tidak menyukai sebuah produk baru = 0,20, Berapakah probabilitas 5 orang dari 18 orang yang terpilih akan tidak menyukai produk baru tersebut?

Jawaban :

b(5;18;0,20) = B(5;18;0,20) – B(4;18;0,20). Dari Tabel Binomial, dapat diketahui B(5;18;0,20) = 0,8671 dan B(4;18;0,20) = 0,7164, sehingga: b(5;18;0,20) = 0,8671 – 0,7164 = 0,1507.

Pendekatan grafikal memperlihatkan hal tersebut sebagai berikut:



Gambar 5.3 Distribusi Binomial Preferensi Konsumen

<u>Contoh-5.4</u> : Distribusi Binomial-4

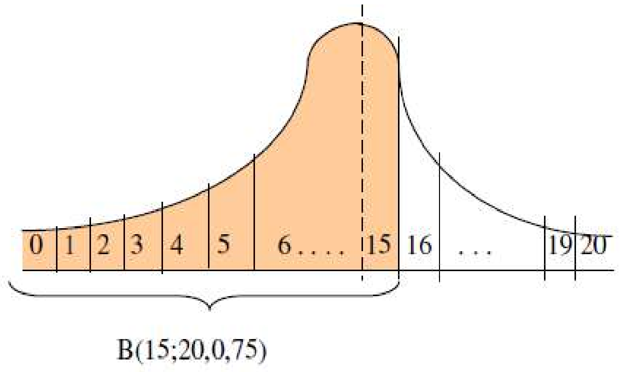
Seseorang manajer pabrik gula (*plant manager*) menyatakan bahwa 75,00% dari seluruh kecelakaan kerja dalam pabrikdapat dihindarkan dengan mengoperasikan sistim keamanan kerja. Berapa probabilitas:

(a) kurang dari 16 dari 20 kecelakaan kerja dalam pabrik dapat dihindarkan dengan pemanfaatan sistim keselamatan kerja?

(b) 12 dari 15 kecelakaan kerja dalam pabrik dapat dihindarkan dengan pemanfaatan sistim keselamatan kerja?

Jawaban :

(a) probabilitas lebih sedikit 16 dari 20 kecelakaan kerja :

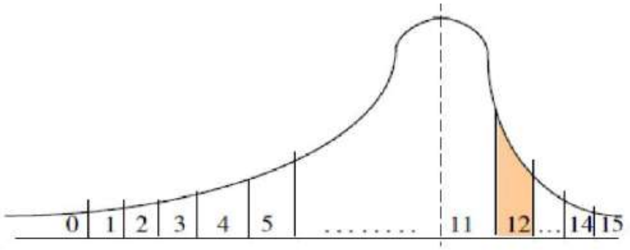


Gambar 5.4 Distribusi Binomial Kecelakaan Kerja

Gunakan teorema distrribusi Binomial: B(X;n;p) = 1 – B(n-X-1;n;1-p)

B(15;20;0,75) = 1 – B(4;20;0,25) = 1 – 0,4148 = 0,5852

(b) probabilitas 12 dari 15 kecelakaan kerja :



Gambar 5.5 Distribusi Binomial Kecelakaan Kerja

b(12;15;0,75) = B(12;15;0,75) – B(11;15;0,75)
Berdasar teorema Binomial, B(12;15;0,75) = 1 – B(2;15;0,25) = 1 – 0,2361 = 0,7639.
B(11;15;0,75) = 1 –B(3;15;0,25). Pada Tabel Binomial, B(3;15;0,25) = 0,4613; maka B(11;15;0,75) = 1 – 0,4613 = 0,5387. Selanjutnya dapat dihitung b(12;15,0,75) = 0,7639 – 0,5387 = 0,2252.

<u>Contoh-5.5</u> : Distribusi Binomial-5

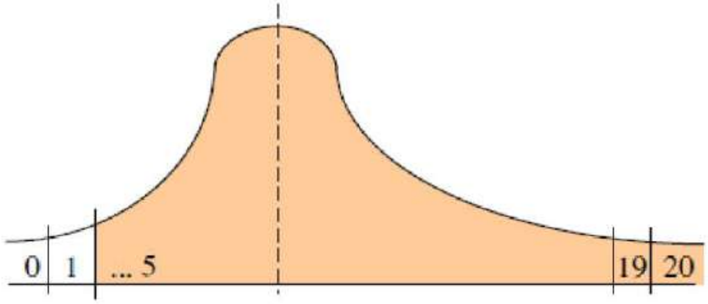
Sebuah perusahaan traktor tangan (*hand tractor*) menyatakan bahwa hanya 10,00% dari mesin-mesin yang mereka hasilkan membutuhkan reparasi dalam jangka waktu garansi (12 bulan). Jika ternyata dari pemeriksaan dilaporkan 5 dari 20 unit *hand tractor* hasil produksinya ternyata membutuhkan reparasi dalam masa garansi, apakah hal ini mendukung pernyataan perusahaan tersebut atau justru melemahkannya?

Jawaban:
P = probabilitas mesin direparasi = 0,10; n = 20 dan x = 5. Probabilitas paling sedikit 5 unit mesin dari 20 unit adalah :

$$\sum_{x=5}^{20} b(5;20;0,10) = 1 - B(4;20;0,10)$$
$$= 1 - 0,9568$$
$$= 0,0432$$

Syafiadi Rizki Abdila, S.T., Ph.D, Prof. Dr. Ir. Drs. Syafwandi, M.Sc
Dr. Yulasmi, S.E., M.M, Dr. Robby Dharma, S.E., M.M
Dr. Syafri Arlis, S.Kom, M.Kom

Karena nilai probabilitas ini sangat kecil, maka pernyataan perusahaan tersebut tidak terbukti benar.



Gambar 5.6 Distribusi Binomial untuk *Hand Tractor*

2. Distribusi Hipergeometrik

Formula probabilitas Distribusi Hypergeometrik :

$$h(x;n;a;N) = \frac{\begin{pmatrix} a \\ X \end{pmatrix}\begin{pmatrix} N\text{-}a \\ n\text{-}X \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}} \quad \text{untuk } x = 0,1,2,\ldots,n \qquad \ldots\ldots (5.4)$$

di mana

N = populasi,

n = ukuran sampel

a = *outcome* pada N

X = *outcome* yang diperkirakan.

X tidak melebihi a, dan (n – X) tidak melebihi (N-a).

Rata-rata Distribusi Hypergeometrik :

$$\mu = n.\left(\frac{a}{N}\right) \qquad \ldots\ldots (5.5)$$

Varians Distribusi Hypergeometrik :

$$\sigma^2 = \frac{n.a.(N-a).(N-n)}{N^2.(N-1)} \qquad \text{. . . . . (5.6)}$$

<u>Contoh-5.6</u> : Distribusi Hipergeometrik-1

Dalam sebuah pengiriman 20 bal tembakau NO Besuki ke gudang di Surabaya, ada 5 unit yang rusak. Jika dipilih secara random 10 bal tembakau NO Besuki, berapakah probabilitas rusak 2 unit ?

Jawaban :

Estimasi tingkat kerusakan (X) = 2; banyaknya sampel yang diamati (n) = 10, N = 20, tingkat kerusakan historis (a) = 5.

$$h(2;10;5;20) = \frac{\binom{5}{2}\binom{15}{8}}{\binom{20}{10}} = \frac{(10)(6{,}435)}{(184756)} = 0{,}348 \text{ atau}$$
34,80%.

<u>Contoh-5.7</u> : Distribusi Hipergeometrik-2

Kembali seperti pada contoh di atas, dengan N = 100 unit dan kerusakan yang ada (a) = 25. Dari sejumlah sampel random (n) = 10, probabilitas 2 unit rusak dihitung dengan Distribusi Hipergeometrik adalah:

$$h(2;10;25;100) = \frac{\binom{25}{2}\binom{75}{8}}{\binom{100}{10}} = 0{,}292$$

Dengan pendekatan Distribusi Binomial, probabilitas 2 unit rusak dari 10 unit sampel random adalah:

$$b(2;10;0{,}25) = \binom{10}{2} \; 0{,}25^2\,(1-0{,}25)^8 = 0{,}282$$

Tampak ada selisih nilai probabilitas sebesar = 0.010, di mana N makin besar (mendekati ∞), selisih tersebut makin kecil. Pendekatan dengan Distribusi Binomial untuk Distribusi Hypergeometric cukup baik jika n < N/10.

3. Pendekatan Poisson terhadap Distribusi Binomial

Jika n dan p kecil, maka Distribusi Binomial dapat didekati dengan Poisson. Formula probabilitas Distribusi Poisson :

$$\boxed{f(X;\lambda) = \frac{\lambda^X e^{-\lambda}}{X!}} \quad \text{untuk } X = 0, 1, 2, \ldots \qquad \ldots\ldots (5.7)$$

di mana:

$\lambda = n.p$

Sebuah *rule of thumb* menyatakan bahwa jika n > 20 dan p < 0.05; dapat dilakukan pendekatan Ditribusi Poisson untuk Distribusi Binomial. Jika n > 100, selama np < 10, maka pendekatan Distribusi Poisson tetap digunakan.

Rata-rata Distribusi Poisson :

$$\boxed{\mu = \lambda} \qquad \ldots\ldots (5.8)$$

Varians Distribusi Poisson :

$$\boxed{\sigma^2 = \lambda} \qquad \ldots\ldots (5.9)$$

<u>Contoh-5.8</u> : Distribusi Binomial dengan Pendekatan POISSON-1

Diketahui bahwa 5,00% dari sejumlah buku yang ada pada sebuah binder mengalami kerusakan. Hitung probabilitas 2 buku dari 100 buku dalam sebuah binder menglami kerusakan, dengan menggunakan :

a. Distribusi Binomial.

b. Distribusi Poisson.

Jawaban :

a. X = 2, n = 100 dan p = 0,05

$$B(2;100;0.05) = \binom{100}{2} \ (0,05)2(0,95)98 = 0,081$$

b. X = 2, ⌊ = n. p = 100 x 0,05 = 5.

$$f(2;5) = \frac{5^2 e^{-5}}{2!} = 0,084$$

Ada selisih nilai hanya sebesar 0.003; selisih yang sangat kecil dan dapat diabaikan, perbedaan yang muncul disebabkan oleh e^{-5}.

<u>Contoh-5.9</u> : Distribusi Binomial Dengan Pendekatan POISSON-2

Sebuah bank menerima rata-rata = 6 cek kosong setiap hari. Berapa probabilitas bank tersebut menerima :

(a) empat cek kosong pada suatu hari,

(b) 10 cek kosong dalam dua hari berturut-turut.

Jawaban :

(a) x = 4 dan ⌊ = 6

$$f(4\,;6) = \frac{6^4\ e^{-6}}{4!} = 0,135$$

(b) $x = 10$ dan $\lambda = 2 \times 6 = 12$

$$f(10\,;12) = \frac{12^{10}\, e^{-12}}{10!} = 0{,}105$$

Atau dengan pendekatan Tabel Distribusi Poisson:

$f(10\,;12) = F(10\,;12) – F(9\,;12) = 0{,}347 – 0{,}242 = 0{,}105.$

4. Distribusi Geometrik

Formula probabilitas Distribusi Geometric :

$$\boxed{g(X\,;p) = p\,(1 – p)^{X-1} \text{ untuk } X = 1, 2, 3, 4, \ldots} \quad \ldots (5.10)$$

Rata-rata Dsitribusi Geometric :

$$\boxed{\mu = \frac{1}{p}} \quad \ldots\ldots (5.11)$$

<u>Contoh-5.10</u> : Distribusi Geometrik-1

Jika probabilitas sebuah ban truk meletus = 0,01; berapa probabilitas ban truk akan meletus untuk pertama kalinya pada saat ia meluncur di jalan keempat kali?

Jawaban :

$X = 4$, $p = 0{,}01$

$g(4\,;0{,}01) = (0{,}01)(0{,}99)^{4-1} = 0{,}010$ atau 1,00%.

<u>Contoh-5.11</u> : Distribusi Geometrik-2

Jika probabilitas sebuah alat pengukur akan memberikan kesalahan ukur = 0,05, berapa probabilitas alat pengukur yang keenam pada saat diuji menunjukkan kesalahan ukur untuk pertama kalinya?

Jawaban :
X = 6, p = 0,05
$g(6\,;0{,}05) = (0{,}05)(0{,}95)^{6-1} = 0{,}039$.

5. Distribusi Multinomial

Formula Distribusi Multinomial:

$$f(x_1, x_2, \ldots, x_k) = \frac{n!}{x_1!\, x_2!\, x_3! \ldots x_k!}\; p_1^{x1} . p_2^{x2} . p_3^{x3} \ldots p_k^{xk} \qquad \text{....(5.12)}$$

Contoh-5.12 : Distribusi Multinomial-1

Probabilitas sebuah jenis bola lampu akan rusak setelah digunakan terus menerus < 40 jam, antara 40 – 80 jam, dan > 80 jam adalah = 0,30, 0,50 dan 0,20. Berapa probabilitas dari 8 buah bola lampu: 2 rusak setelah dipakai < 40 jam, 5 rusak setelah dipakai antara 40 – 80 jam, dan 1 rusak setelah dipakai > 80 jam secara terus menerus?

Jawaban :

n = 8, X_1 = 2, X_2 = 5, X_3 = 1, p_1 = 0,30; p_2 = 0,50 dan p_3 = 0,20

$$f(2,5,1) = \frac{8!}{2!\; 5!\; 1!}\; 0{,}30^2\; 0{,}50^5\; 0{,}20^1 = 0{,}0945 \text{ atau } 9{,}45\%$$

Contoh-5.13 : Distribusi Multinomial-2

Diketahui bahwa probabilitas penggunaan BBM pada mobil-mobil impor rata-rata kurang dari 22 km per galon, antara 22 – 25 km per galon, lebih dari 25 km per galon adalah: 0,40; 0,40 dan 0,20. Dari 12 mobil impor yang diuji, berapakah kemungkinan 4 mobil menggunakan BBM kurang dari 22 km/galon, 6 mobil menggunakan BBM antara 22 – 25 km/galon dan 2 mobil menggunakan BBM di atas 25 km/galon?

Jawaban:

$n = 12$, $X_1 = 4$, $X_2 = 6$, $X_3 = 2$; $p_1 = 0{,}40$, $p_2 = 0{,}40$ dan $p_3 = 0{,}20$.

$$f(4,6,2) = \frac{12!}{4!\ 6!\ 2!}\ 0{,}40^4\ 0{,}40^6\ 0{,}20^2 = 0{,}0581 \text{ atau } 5{,}81\%,$$

<u>Contoh-5.14</u> : Distribusi Multinomial-3

Probabilitas untuk memperoleh gambar kepala 0, 1, atau 2 pada setiap pelemparan dua mata uang adalah : ¼, ½, dan ¼. Berapa probabilitas untuk memperoleh 2 gambar ekor sebanyak 2 kali, 1 gambar kepala dan 1 gambar ekor sebanyak 3 kali, dan 2 gambar kepala sebanyak 1 kali dalam enam kali pelemparan dua mata uang?

Jawaban:

$n = 6$, $X_1 = 2$, $X_2 = 3$, $X_3 = 1$, $p_1 = ¼$, $p_2 = ½$, dan $p_3 = ¼$.

$$f(2,3,1) = \frac{6!}{2!\ 3!\ 1!}\ 0{,}25^2\ 0{,}50^3\ 0{,}25^1 = 0{,}1172 \text{ atau } 11{,}72\%.$$

5.3. Padatan Probabilitas

Variabel random yang bersifat kontinyu memiliki karakteristik probabilitas utama adalah sebagai berikut :

$$F(X) = \int_{-}^{x} f(t)\ d_t$$

<u>Contoh-5.15</u> : Padatan Probabilitas Variabel Random

Jika sebuah variabel random memiliki padatan probabilitas seperti berikut:

$$f(X) = \begin{cases} 2\ e^{-2x} & \text{untuk } X > 0 \\ 0 & \text{untuk } X < 0 \end{cases}$$

Hitung probabilitas variabel tersebut memiliki nilai :

(a). antara 1 sampai dengan 3,

(b). lebih besar dari 0,5.

Jawaban :

(a) $\int_{1}^{3} 2\ e^{-2x}.\ dX = e^{-2} - e^{-6} = 0{,}135$

b. $\int_{0{,}5}^{+} 2\ e^{-2x}.\ dX = e^{-1} = 0{,}368$

1. Padatan Probabilitas Normal

Formula padatan Normal adalah :

$$F(z) = \frac{1}{\sqrt{2\,\square}} \int_{-}^{z} e^{-1/2\ tz}\ .d_t \qquad \text{di mana} - \quad < X <$$

. . . . . (5.13)

<u>Contoh-5.16</u> : Padatan Probabilitas Normal-1

Hitung probabilitas sebuah variabel random yang berdistribusi normal standard memiliki nilai:

(a). antara 0,87 sampai dengan 1,28

(b). antara -0,34 sampai dengan 0,62

(c). lebih besar dari 0,85

(d). lebih besar dari -0,65

Jawaban :

(a). F(1,28) – F(0,87) = 0,8997 – 0,8078 = 0,0919

(b). F(0,62) – F(-0,34) = 0,7324 – (1 – 0,6331) = 0,3655

(c). 1 – F(0,85) = 1 – 0,8023 = 0,1977

(d). 1 – F(-0,65) = 1 – [1 – F(0,65)] = F(0,65) = 0,742

Syafiadi Rizki Abdila, S.T., Ph.D, Prof. Dr. Ir. Drs. Syafwandi, M.Sc
Dr. Yulasmi, S.E., M.M, Dr. Robby Dharma, S.E., M.M
Dr. Syafri Arlis, S.Kom, M.Kom

Nilai probabilitas normal, bisa dilihat pada Tabel Fungsi Distribusi Normal. Contoh: F(1,28), lihat kolom z pada 1,2 dan baris z pada 0,08; nilainya = 0,8997.

$$P(a < X < b) = F\left(\frac{b - \mu}{\sigma}\right) - F\left(\frac{a - \mu}{\sigma}\right) \qquad .....(5.14)$$

<u>Contoh-5.17</u> : Padatan Probablitas Normal-2

Jika radiasi kosmik yang mengenai seseorang pada saat ia melakukan penerbangan merupakan variabel random yang berdistribusi normal dengan μ = 4,35 mrem dan σ = 0,59 mrem. Hitung proibabilitas seseorang akan menerima radiasi kosmik pada saat melakukan penerbangan :

(a). antara 4,00 sampai dengan 5,00 mrem.

(b). paling sedikit 5,50 mrem.

Jawaban :

(a). $F\left(\frac{5 - 4{,}35}{0{,}59}\right) - F\left(\frac{4 - 4{,}35}{0{,}59}\right) = F(1{,}10) - F(-0{,}59)^{*}$

$= 0{,}8643 - (1 - 0{,}7224) = 0{,}5867$

* Untuk nilai z negatif, maka F(-z) = 1 – F(z).

(b). $1 - F\left(\frac{5{,}50 - 4{,}35}{0{,}59}\right) = 1 - F(1{,}95) = 1 - 0{,}9744 = 0{,}0256$

Untuk kasus ini, paling sedikit 5,50 mrem, berarti sampai dengan + (di mana probabailitas total = 1).

<u>Contoh-5.18</u> : Padatan Probabilitas Normal-3

Banyaknya volume bubuk kopi yang diisi oleh mesin otomatis di *setup* = 4 ons per karton. Diperkirakan bahwa volume bubuk kopi tersebut merupakan variabel yang berdistribusi normal, dengan σ = 0,04 ons. Jika hanya 2,00% dari karton-karton terserbut berisi kurang dari 4 ons, berapa rata-rata pengisian bubuk kopi ke dalam karton oleh mesin otomatis tersebut?

Jawaban :

$$F\left(\frac{4-\mu}{0{,}04}\right) = 0{,}02$$

Maka

$$F\left(-\frac{4-\mu}{0{,}04}\right) = 1- 0{,}02 = 0{,}98$$

Lihat Tabel Kurve Normal, nilai probabilitas yang paling dekat dengan 0,98 adalah 0,9798, yaitu pada z = 2,05

$$-\left(\frac{4-\mu}{0{,}04}\right) = 2{,}05$$

Dengan demikian μ dapat dihitung = 4,082 ons.

<u>Contoh-5.19</u> : Pendekatan Normal Untuk Distribusi Binomial

Jika n besar dan p mendekati 0,50, maka dapat dilakukan pendekatan normal untuk distribusi Binomial. Teorema yang berlaku adalah:

Jika X adalah sebuah nilai dari sebuah variabel random yang berdistribusi Binomial dengan paramter n, p, dan jika

$$z = \frac{X - np}{\sqrt{np(1-p)}}$$

maka bentuk terbatas fungsi distribusi dari variabel random terstandard, di mana n → ∞ adalah :

$$F(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-1/2t^2}\, dt \quad -\infty < z < \infty$$

Syafiadi Rizki Abdila, S.T., Ph.D, Prof. Dr. Ir. Drs. Syafwandi, M.Sc
Dr. Yulasmi, S.E., M.M, Dr. Robby Dharma, S.E., M.M
Dr. Syafri Arlis, S.Kom, M.Kom

Rata-rata kerusakan diode yang dibuat sebuah pabrik ternyata 20,00%. Berapa probabilitas dari 100 unit diode yang dipilih secara random:

(a). Paling banyak 15 diode yang rusak?

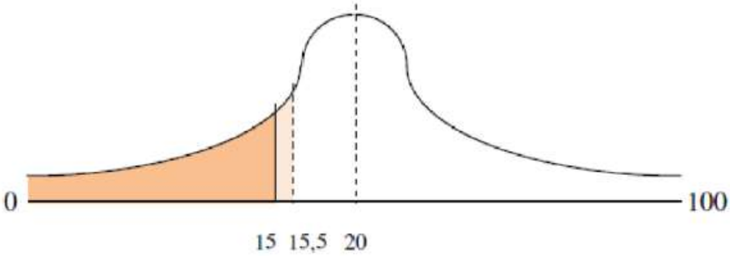
(b). 15 diode yang rusak?

Jawab :

Rata-rata kerusakan = 20,00% x 100 diode = 20 unit diode. s = $\sqrt{np(1-p)}$ =

$\sqrt{(100)(0,20)(0,80)} = 4$

Pendekatan normal untuk distribusi Binomial ini adalah:

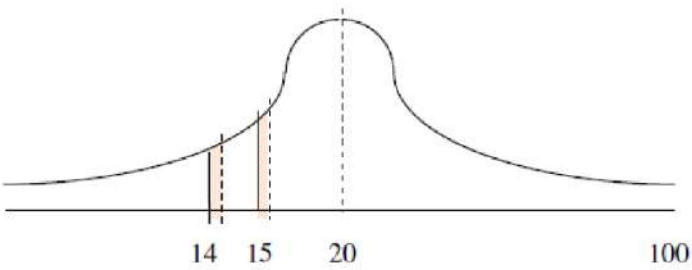


Gambar 5.7 Pendekatan Normal Untuk Distribusi Binomial

Agar nilai 15 masuk dalam area di bawah kurva normal, maka perlu ditambah dengan 0,5, selanjutnya diberlakukan rumus probabilitas normal:

$$F = \left(\frac{(15,50 - 20)}{4}\right) = F(-1,13) = 1 - F(1,13)$$

= 1 – 0,8708 = 0,1292. Probabilitas paling banyak 15 diode rusak dari 100 diode = 12,92%.



Gambar 5.8 Pendekatan Normal Untuk Distribusi Binomial.

$$F\left(\frac{15{,}50 - 20}{4}\right) - F\left(\frac{14{,}5 - 20}{4}\right) = F(-1{,}13) - F(-1{,}38)$$

$= \{1 - F(1{,}13)\} - \{1 - F(1{,}38)\}$

$= 0{,}9162 - 0{,}8708 = 0{,}0454$. Probabilitas 15 diode rusak dari 100 diode = 4,54%.

5.4. Rangkuman

Distribusi probabilitas adalah sebuah penyebaran probabilitas sebuah variabel random memiliki sebuah nilai pada suatu rentang nilai. Variabel radom bisa berbentuk variabel kontinyu atau berbentuk diskret. Variabel diskret adalah variabel dengan nilai yang terbatas (*finite*), contoh: data tingkat absensi buruh yang diukur dalam hari per bulan merupakan variabel diskret, tetapi berat seseorang merupakan variabel kontinyu (*infinite*). Untuk variabel diskret, distribusi probabilitas disebut sebagai ‘distribusi probabilitas’; sedangkan untuk variabel kontinyu disebut “padatan probabilitas”.

Formula yang digunakan untuk menghitung distribusi probabiltas:

(a) Distribusi Binomial:

$$b(n;x;p) = \binom{n}{x} p^x (1-p)^{n-x} \text{ untuk } x = 0, 1, 2, \ldots n \qquad \ldots.(5.1)$$

Teorema yang berlaku pada Distribusi Binomial adalah :

(1) b(x;n;p) = b(n-x;n;1-p)

(2) B(x;n;p) = 1 – B(n-x-1;n;1-p)

(3) b(x;n;p) = B(x;n;p) – B(x-1;n;p)

(4) b(x;n;p) = B(n-x;n;1-p) – B(n-x-1;n;1-p)

(b) Distribusi Hipergeometrik:

$$h(x;n;a;N) = \frac{\binom{a}{X}\binom{N-a}{n-X}}{\binom{N}{n}} \text{ untuk } x = 0, 1, 2, \ldots, n \qquad . \qquad \ldots . (5.4)$$

(c) Distribusi Poisson:

$$f(X;\lambda) = \frac{\lambda^X e^{-\lambda}}{X!} \quad \text{untuk } X = 0, 1, 2, \ldots \qquad \ldots\ldots (5.7)$$

(d) Distribusi Multinomial:

$$f(x_1, x_2, \ldots, x_k) = \frac{n!}{x_1!\, x_2!\, x_3! \ldots x_k!} p_1^{x1}.p_2^{x2}.p_3^{x3} \ldots p_k^{xk}$$

(5.12)

(e) Distribusi Geometrik:

$$g(X;p) = p(1-p)^{X-1} \text{ untuk } X = 1, 2, 3, 4, \ldots.$$

$\ldots\ldots (5.10)$

Formula yang digunakan untuk menghitung padatan probabilitas:

(a) Padatan Probabilitas Normal:

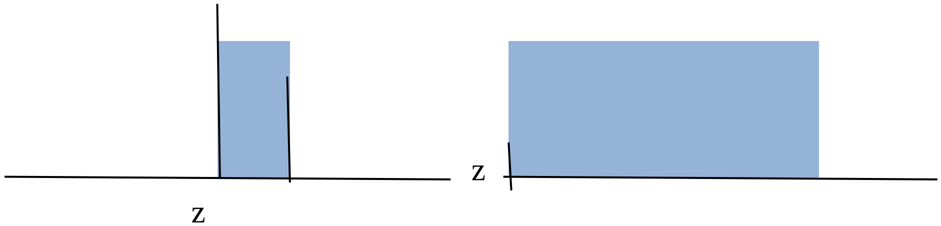
$$F(z) = \frac{1}{\sqrt{2□}} \int_{-}^{z} e^{-1/2\ tz} .d_t \quad \text{di mana} - \ < X < \qquad \ldots\ldots (5.13)$$

(b) Padatan Probabiltas Normal untuk range:

$$P(a < X < b) = F\left(\frac{b - \mu}{\sigma}\right) - F\left(\frac{a - \mu}{\sigma}\right) \qquad \ldots\ldots (5.14)$$

5.5. Diskusi

Pilihan penggunaan formula untuk menghitung probabilitas suatu kejadian. Tergantung kepada jenis kejadian dan asumsi-asumsi yang mendasari. Sekarang ini, telah tersedia tabel distribusi dan padatan probabilitas untuk mempermudah peneliti. Namun seandainya tabel tersebut tidak tersedia, aplikasi Excel dapat digunakan dengan mudah. Khusus untuk tabel distribusi normal, ada dua versi yang ada pada buku Statistik:



Luas area di bawah kurve pada gambar sebelah kiri, dimulai dari 0,500; sedang gambar sebelah kanan dimulai dari 0,000. Keduanya sama benar. Kalau menggunakan gambar sebelah kiri, maka nilai probabilitas ditambah dengan 0,500.

Syafiadi Rizki Abdila, S.T., Ph.D, Prof. Dr. Ir. Drs. Syafwandi, M.Sc
Dr. Yulasmi, S.E., M.M, Dr. Robby Dharma, S.E., M.M
Dr. Syafri Arlis, S.Kom, M.Kom

5.6. Referensi

Bowen, Earl, K., and Martin K. Starr, 2002, ***Basic Statistics for Business andEconomics,*** McGraw Hill Book Company, Tokyo.

Doane, David, P., and Lori E. Seward, 2007, ***Applied Statistics in Business andEconomics***, McGraw Hill/Irwin Series, New York.

Freund, John E., and Irwin Miller, 2004, ***Probability and Statistics for Engineers***, Fifth Edition,Prentice-Hall of India, Private Limited, New Delhi.

Hines, William, H., and Douglas C. Montgomey, 2000, ***Probability and Statistics for Engineers inEngineering and Management Science,*** John Wiley ans Sons,Inc., New York.

5.7. Latihan Soal

Gunakan Tabel Distribusi Poisson untuk menghitung F(4;7); f(4;7) dan

$$\sum_{k=6}^{19} f(4;8)$$

Pemerintah Daerah menetapkan bahwa 70,00% dari calon PNS memiliki peluang lulus dalam ujian seleksi untuk menjadi PNS jika memiliki IQ = 90. Gunakan Tabel Distribusi Binomial untuk menghitung probabilitas dari 15 orang calon PNS dengan IQ = 90:

(a) Paling sedikit 12 orang calon PNS lulus dalam ujian seleksi tersebut.

(b) Paling banyak 6 orangcalon PNS lulus dalam ujian seleksi tersebut.

(c) Sebanyak 6 orangcalon PNS lulus dalam ujian seleksi tersebut.

Sebuah koperasi petani melon menyatakan bahwa 90,00% hasil panennya berkualitas baik dan segar. Hitung probabilitas dari 18 unit kemasan yang siap dikirim ke toko buah:

(a) Seluruh kemasan melon itu benar-benar berkualitas baik dan segar.

(b) Paling sedikit 16 kemasan melaon benar-benar berkualitas baik dan segar.

(c) Paling banyak 14 kemasan melon benar-benar berkualitas baik dan segar.

Gunakan pendekatan distribusi Poisson untuk menghitung probabilitas distribusi Binomial b(3;100;0,03).

Peluang lampu LED memiliki daya pakai paling sedikit 500 jam = 0,85. Pada pemeriksaan 20 buah lampu LED, hitung probabilitas:

(a) 18 buah lampu LED memiliki daya pakai paling sedikit 500 jam.

(b) Paling sedikit 15 buah lampu LED memiliki daya pakai paling sedikit 500 jam.

(c) Paling sedikit 2 buah lampu LED tidak memiliki daya pakai paling sedikit 500 jam.

BAB 6

TEORI ESTIMASI ATAU PENDUGAAN

Estimasi parameter populasi dapat dilakukan dengan menggunakan parameter sampel, namun perlu dilibatkan peluang tingkat kesalahan yang mungkin terjadi. Persentase peluang tingkat kesalahan tergantung kepada tingkat keyakinan (*confidence interval*) analis dalam mengestimasi interval paramter populasi.

Kemampuan Akhir yang Diharapkan::
- Mahasiswa mampu memahami manfaat taksiran parameter populasi berdasar parameter sampel, - Mampu melakukan pendugaan mean, standard pentimpangan populasi melalui parameter yang sama dari sampel.

6.1 Pendahuluan

Untuk mengetahui parameter sebuah populasi, peneliti menggunakan parameter sampel. Melalui teknik penaksiran atau pendugaan, parameter populasi dapat diperkirakan nilainya melalui parameter sampel. Pembahasan mengenai penaksiran ini merupakan statistik inferensial.

Statistik inferensial memiliki dua elemen, yaitu: (a) inferen, dan (b) ukuran kesesuaiannya. Pada tataran praktek, teknik pendugaan ini banyak digunakan untuk mengestimasi ukuran rata-rata dan standad penyimpangan. Contoh: pemerintah ingin mengetahui rata-rata aliran uang ke luar negeri per tahun; broker saham ingin mengetahui kebiasaan-kebiasaan dalam pasar saham, dan ahli logam ingin mengetahui seberapa jauh logam tertentu akan berubah setelah menerima pemanasan sampai derajat temperatur tertentu.

Peningkatan angka pengangguran dari 6,9% menjadi 7,9% (atau sebesar 1,00%) dapat mempengaruhi harga saham di pasar saham, menjadi pemicu peralihan dana yang besar ke daerah-daerah dengan tingkat pengangguran yang tinggi. Bahkan secara politis, hal ini dapat mempersulit dalam mempertahankan kekuasaannya. Namun, pada faktanya tidak seorangpun yang tahu berapa peningkatan angka pengangguran itu sebenar-benarnya. Hasil peningkatan itu diperoleh dari sebuah survey dengan menggunakan sampel. Ini merupakan estimasi sampel yang diusahakan untuk mengukur parameter pengangguran dari populasi.

Ada dua jenis prosedur dalam estimasi, yaitu: (i) estimasi titik (*point estimation*); dan (ii) estimasi interval (*interval estimation*). Estimasi titik merupakan nilai estimasi tunggal, seperti: rata-rata nilai matakuliah Statistik mahasiswa pada fakultas tertentu = 65. Estimasi interval merupakan nilai estimasi dengan rentang tertentu, misal rata-rata nilai matakuliah Statistik adalah antara 65 – 85. Ada pertanyaan yang sering muncul di kalangan peneliti pemula, yaitu: metode inferensial yang mana akan dipakai; apakah nilai parameter akan diestimasi atau apakah akan melakukan uji hipotesis berkaitan dengan nilai parameter itu? Jawabannya, tergantung kepada kesimpulan logis yang dianggap benar oleh si peneliti.

6.2.Ciri-ciri Estimator yang Baik

Estimator parameter populasi ditaksir melalui parameter sampel. Syarat estimator yang baik adalah[10]:

a. Estimator ($\hat{}$) dikatakan estimator yang tidak bias jika rata-rata semua kemungkinan harga $\hat{}$ memiliki E{$\hat{}$} = 0. Sebaliknya, estimator dianggap bias, jika E{$\hat{}$} ≠ 0.
b. Konsisten. Estimator yang konsisten merupakan penduga yang berkonsentrasi secara sempurna pada parameter, jika sampel sebenarnya bertambah secara tidak terhitung.
c. Interval estimator minimum.

6.3. Estimasi Titik dan Estimasi Interval

1. Estimasi Titik (*point estimator*).

Sebuah estimator titik adalah estimator dengan nilai tunggal. Serangkaian data sampel berikut: 1 2 4 5 7 11, dapat dihitung rata-rata sampel ($\bar{X}$) :

$\bar{X}$ = (© X_t)/n = 30/6 = 5

Nilai rata-rata sampel = 5 ini merupakan estimator rata-rata populasi, ⌠ yang tidak diketahui. Jika angka 2 dan 4 merupakan simbul ‘kesuksesan’, maka proporsi sukses ($\bar{p}$) = 2/6 = 1/3, ini merupakan estimator titik untuk proporsi kesuksesan populasi.

[10]Setyawan, 2009:57

Demikian pula, jika ingin mengetahui standard penyimpangan populasi (ʃ$_x$); estimasinya menggunakan standard penyimpangan sampel (S_x): di mana $S_x = \sqrt{\{(© X_t - X)^2/(n-1)\}}$. Untuk rangkaian data sampel di atas, S_x = 3,633; ini merupakan estimator titik untuk ʃ$_x$ yang tidak diketahui.

Jika ukuran sampel kurang dari 5,00% dari populasi, maka estimasi standard penyimpangan sampel menggunakan:

$$\boxed{S_x' = \frac{S_x}{\sqrt{n}}} \qquad \ldots\ldots (6.1)$$

2. Estimator Interval (*Interval Estimator*)

Estimator interval menjelaskan sebuah rentang nilai di mana parameter yang diestimasi tersebut berada. Kembali pada data: 1 2 4 5 7 dan 11 di atas, yang dipilih secara random dari sebuah populasi di mana ⌠ tidak diketahui; Rata-rata sampel = 5 merupakan estimator titik untuk ⌠ (rata-rata populasi). Jika peneliti memiliki keyakinan bahwa rentang di mana X itu berada adalah: X $\pm$ 1, maka estimator interval untuk rata-rata populasi: 4 $\leq$ ⌠ $\leq$ 6.

6.4. Interval Keyakinan (*confidence interval*) untuk Estimasi ⌠

Interval keyakinan estimasi ⌠ merupakan sebuah interval nilai a ke b di mana ⌠ diduga berada di dalamnya. Interal ini merupakan inferen berdasar pada: (i) nilai rata-rata sampel random, X, yang dipilih dari populasi; dan (ii) distribusi ʃ diketahui. Peneliti tidak dapat memastikan bahwa interval itu benar 100,00% karena sampel hanya bagian kecil dari populasi. Untuk itu diperlukan tingkat keyakinan peneliti berkaitan dengan tingkat kebenaran pendugaan interval, contoh: 90,00% atau 95,00% atau lainnya. Pilihan metode untuk menentukan tingkat keyakinan untuk μ tergantung kepada: “apakah populasinya berdistribusi normal atau tidak, dan apakah σ_x diketahui atau tidak”.

Tingkat Keyakinan Estimasi Interval μ, Populasi Berdistribusi Normal dan σ_X Diketahui

Sebuah populasi berdistribusi normal memiliki μ dan σ_X. Variabel normal standardnya adalah:

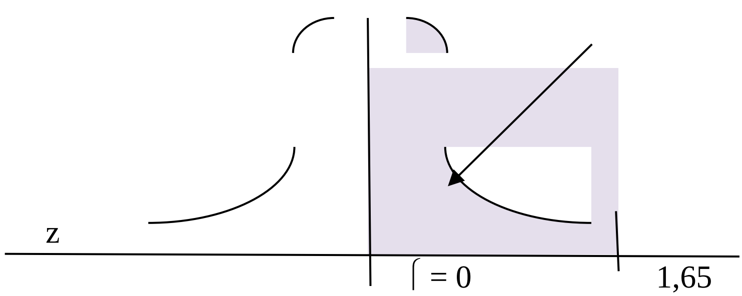
$$\boxed{Z = \frac{\bar{X} - \mu}{\sigma_{X'}}} \qquad \text{. . . . . (6.2)}$$

Di mana $\sigma_{X'} = \sigma_X/\sqrt{n}$.

Jika C = 0,95 Pada titik $\bar{x} = \mu + 1{,}65\ \sigma_{X'}$;

$$Z = \frac{(\mu + 1{,}65\ \sigma_{X'}) - \mu}{\sigma_{X'}} = \frac{1{,}65\ \sigma_{X'}}{\sigma_{X'}} = 1{,}65$$

p(0 s/d 1,65) = 0,4505



Gambar 6.1 Distribusi Sampel μ

Gambar tersebut di atas berlaku pada probabilitas pernyataan tidak benar (α) = 5,00%. Proporsi pernyataan itu benar = C. C + α = 1 atau α = 1 – C.

<u>Contoh 6.1</u>: Estimasi μ, Populasi berdistribusi normal, σ_X diketahui.

Sebuah populasi berdistribusi normal, memiliki standard penyimpangan, $\sigma_X = 10$. Sampel sebanyak 25 dipilih secara random dari poplasi, memiliki rata-rata sampel = 50,0. Pada tingkat keyakinan = 0,95; berapa estimasi μ populasi tersebut?

Jawabnya: n = 25; $\sigma_X = 10$, $\bar{X} = 50{,}0$. Hitung standard penyimpangan μ (*standard estimation of mean,* σ_X'):

$$\boxed{\sigma_X' = \frac{\sigma_X}{\sqrt{n}}} \qquad \text{. . . . . (6.3)}$$

$\sigma_X' = 10/\sqrt{25} = 2$,

Pada tingkat keyakinan (C) = 0,95:

$\bar{X} - 1{,}65\ \sigma_X' \leq \mu \leq \bar{X} + 1{,}65\ \sigma_X'$

$50 - (1{,}65)(2) \leq \mu \leq 50 + (1{,}65)(2)$

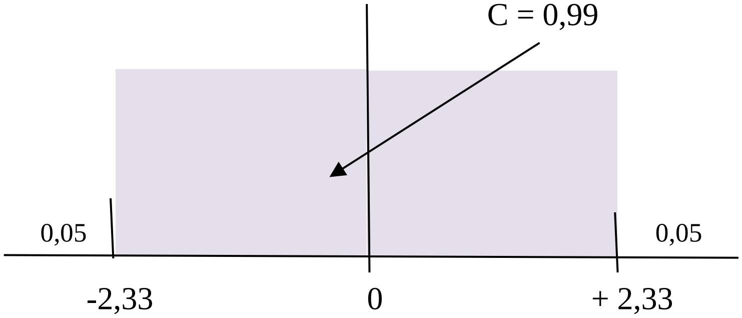
$46{,}7 \leq \mu \leq 53{,}3$

Dengan demikian dapat dinyatakan bahwa pada tingkat keyakinan = 0,95; estimasi rata-rata populasi (μ) berada di antara 46,7 sampai dengan 53,3.

<u>Contoh 6.2</u>: Estimasi μ, Populasi berdistribusi normal, σ_X diketahui.

Ketebalan lempengan ubin marmer yang diproses oleh mesin otomatis, diperkirakan berdistribui normal, dan memilik standard penyimpangan populasi = 0,02 mm. Sebanyak 4 buah lempengan marmer yang dihasilkan dipilih sebagai sampel untuk diukur ketebalannya. Rata-rata ketebalan sampel tersebut = 25,01 mm. Pada tingkat keyakinan 0,95; estimasi μ ketebalan populasi ubin marmer yang.dihasilkan mesin tersebut adalah: 24,99 mm ≤ ⌠ ≤ 25,03 mm.

Untuk menentukan nilai z, dapat menggunakan Tabel Distribusi Normal. Misal C = 0,99:



Gambar 6.2 Luas Area Dibawah Kurve Normal

Nilai z = 2,33 pada Tabel Distribusi Normal ditentukan sebagai berikut:

Tabel 6.1 Fungsi Distribusi Normal

Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	**0,08**	0,09
0,1										
0,2										
...										
2,3				0,9901						
...										
6,0										

Nilai dalam sel tabel yang mendekati 0,99 adalah 0,9901. Lihat pada baris-z = 2,3; pada kolom-z = 0,03, maka nilai z = 2,33.

Makin kecil nilai C, makin kecil standard penyimpangan populasi, dan makin besar ukuran sampel, maka makin presisi hasil estimasi parameter populasi. Ini bisa dilihat pada kasus berikut:

<u>Contoh 6.3</u>: Perbandingan Presisi Hasil Estimasi Parameter Populasi.

Diasumsikan sebuah populasi berdistribusi normal. Skenario A: (dengan C = 0,95), n_1 = 25, dengan X_1 = 50, dan σ_X = 10. Skenario B: (dengan C = 0,90), n_2 = 100, X_2 = 50, dan σ_X = 4.

Skenario A (C = 0,95): z = 1,65. X_1 = 50. σ_X' = 10/√25 = 2

$50 – (1,65)(2) \leq \mu \leq. 50 + (1,65)(2)$

$46,7 \leq \mu \leq 53,3$

Skenario B (C = 0,90): z = 1,28. X_2 = 50. σ_X' = 4/√100 = 0,4

$50 – (1,28)(0,4) \leq \mu \leq. 50 + (1,28)(0,4)$

$49,49 \leq \mu \leq 50,51$

Hasil estimasi μ populasi pada skenario A memiliki interval yang lebih lebar dibandig interval skenario B. Dapat dinyatakan bahwa skenario B menghasilkan estimasi μ populasi yang lebih presisi.

Tingkat Keyakinan Estimasi Interval μ, Populasi Berdistribusi Normal dan σ_X Tidak Diketahui

Kadang-kadang, parameter populasi (μ dan σ) tidak diketahui. Dengan demikian, σ_X harus diestimasi juga. Pendugaan σ_X dapat dilakukan dengan formula:

$$S_X = \sqrt{\frac{\Sigma (X_t - \bar{X})^2}{n - 1}} \qquad \text{. . . . . (6.4)}$$

Pendugaan μ dilakukan dengan formula:

$$\bar{X} - t_{\alpha/2,v}\frac{S_X}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\alpha/2,v}\frac{S_X}{\sqrt{n}} \qquad \text{. . . .. (6.5)}$$

<u>Contoh 6.4</u>: Estimasi μ, Populasi berdistribusi normal, σ_X tidak diketahui.

Sebuah pabrik besar yang menghasilkan limbah poluta n (*sulfur oxide*) ingin mengetahui rata-rata jumlah polutan tersebut dalam setiap harinya. Karena biaya pengukuran polutan itu mahal, maka hanya 10 hari kerja pabrik yang diamati. Hasil pengukuran polutan dalam 10 hari tersebut (dalam satuan ton) adalah:

8 7 10 15 11 6 8 5
13 12

Jika limbah polutan yang dihasilkan itu berdistribusi normal, gunakan tingkat keyakinan = 0,95 untuk mengetahui μ. Rata-rata sampel ($\bar{X}$) = 95/10 = 9,5 ton per hari.

Estimasi σ_X dengan formula (6.4):

$$S_X = \sqrt{\frac{94{,}5}{10-1}} = 3{,}24 \text{ ton}$$

v = *degree of freedom* = n – 1 = 9. α = 1 – C = 1 – 0,95 = 0,05.

Nilai $t_{\alpha/2,v}$ pada Tabel-t adalah: 2,26. Selanjutnya lakukan pendugaan μ dengan formula (6.5):

$9{,}5 - (2{,}26)(3{,}24)/\sqrt{10} \leq \mu \leq 9{,}5 + (2{,}26)(3{,}24)/\sqrt{10}$

$7{,}184 \leq \mu \leq 11{,}816$

Jika standard penyimpangan emisi polutan itu diketahui secara pasti, yaitu = 3,24 ton per hari; maka σ_X tidak perlu ditaksir dengan formula (6.4). Gunakan formula (6.3) untuk menaksir σ_X' = 3,24/$\sqrt{10}$ = 1,025. Selanjutnya lakukan estimasi μ dengan formula: $X - z_{\alpha/2}\,\sigma_X' \leq \mu \leq X + z_{\alpha/2}\,\sigma_X'$

$$9{,}5 - (1{,}28)(1{,}025) \le \mu \le 9{,}5 + (1{,}28)(1{,}025)$$

$$8{,}188 \le \mu \le 10{,}812$$

Perbandingan antara interval estimasi μ, jika σ_X diketahui dengan jika σ_X tidak diketahui; menunjukkan bahwa dengan diketahuinya σ_X dengan pasti, akan lebih presisi, karena interval estimasinua makin kecil.

Pada tataran praktek, kadang-kadang peneliti tidak mengetahui apakah populasi berdistribusi normal atau tidak. Untuk itu, biasanya ukuran sampel ditingkatkan sehingga n > 30. Dengan sampel besar ini, diasumsikan bahwa populasi berdistribusi normal, walaupun pada kenyataannya tidak normal. Para ahli statistik kemudian bersepakat (*rule of thumb*) seperti ini: jika σ_X tidak diketahui, namun ukuran sampelnya besar (n > 30), maka σ_X ditaksir dengan S_X. Gunakan estimasi interval:

$$\bar{X} - z_{\alpha/2,v}\frac{S_X}{\sqrt{n}} \le \mu \le \bar{X} + z_{\alpha/2,v}\frac{S_X}{\sqrt{n}} \qquad \ldots\ldots (6.6)$$

6.5. Interval Keyakinan (*confidence interval*) Untuk Proporsi Populasi

Seringkali peneliti bermaksud mengetahui proporsi populasi yang tidak diketahui, contoh: 'berapa proporsi pemirsa menyaksikan program televisi tertentu'. Proporsi populasi disimbulkan sebagai 'p'. Pemirsa yang tidak menyaksikan program itu disimbulkan sebagai 'q', di mana q = 1 – p. Untuk mengestimasi p, digunakan $\hat{p}$:

$$\hat{p} = \frac{\text{Banyaknya peristiwa terjadi dalam ukuran sampel-n}}{n}$$

$$q = 1 - \hat{p}$$

Pendekatan distribusi hipergeometrik atau distribusi binomial adalah pendekatan yang digunakan untuk mengestimasi proporsi populasi, namun perhitungannya cukup rumit dan membutuhkan waktu. Untuk itu, digunakan pendekatan distribusi normal, yaitu: nÞ $\geq$ 15 dan nq $\geq$ 15.

Distribusi sampel p memiliki *mean* = p, dan *standard error* :

$$\boxed{ʃ_p = \sqrt{(pq/n)}} \qquad \text{. . . . . (6.7)}$$

Jika p tidak diketahui, formula tersebut menjadi:

$$\boxed{S_p = \sqrt{(Þq/n)}} \qquad \text{. . . . . (6.8)}$$

Untuk pendugaan proporsi populasi, formula yang digunakan adalah:

$$\boxed{Þ - z_{〈/2,v}\, S_p \leq p \leq Þ + z_{〈/2,v}\, S_p} \qquad \text{. . . . . (6.9)}$$

<u>Contoh 6.5</u>: Estimasi p, Populasi berdistribusi normal.

Pada suatu tahun, dari sampel random sebanyak 400 orang buruh anggota serikat buruh ternyata 32 orang yang tidak bekerja. Pada C = 0,95, buat estimasi interval untuk populasi.

n = 400, proporsi sampel = Þ = 32/400 = 0,08. Nilai z pada C = 0,95 adalah 1,65. Dengan demikian estimasi interval p adalah:

0,08 – 1,65(√{(0,08)(0,92)/400} $\leq$ p $\leq$ 0,08 + 1,65(√{(0,08)(0,92)/400}

0,058 $\leq$ p $\leq$ 0,102

Jika pendugaan proporsi populasi ini dianggap kurang akurat, maka ukuran sampel diperbesar menjadi 500, C diturunkan menjadi 0,90. Dari sampel sejumlah itu ternyata ada

35 orang buruh yang tidak bekerja Maka estimasi interval untuk proporsi populasi:

Þ = 35/500 = 0,07 q = 1 – Þ = 1 – 0,07 = 0,93 z = 1,28

0,07 – 1,28(√{(0,07)(0,93)/500} ≤ p ≤ 0,08 + 1,28(√{(0,07)(0,93)/500}

0,055 ≤ p ≤ 0,085

Terlihat bahwa interval pendugaan proporsi populasi makin kecil, dengan kata lain lebih akurat.

6.6. Penentuan Ukuran Sampel untuk Estimasi Parameter Populasi[11]

Penggunaan sampel dari suatu populasi adalah karena alasan biaya. Namun sampel yang besar juga menimbulkan biaya yang besar pula. Untuk menekan biaya, peneliti harus menekan ukuran sampel sekecil mungkin.

1. Ukuran Sampel untuk Estimasi Rata-rata Populasi

$$\boxed{n\,ʃ = \left(\frac{Z_{\langle/2}\,ʃ x^{2}}{e}\right)} \qquad \text{. . . . . (6.10)}$$

di mana

$$e = Z_{\langle/2}\,\frac{ʃ x}{\sqrt{n}}$$

ʃx = (nilai tertinggi – nilai terendah)/4

[11]Bowen and Starr;2002:326

Syafiadi Rizki Abdila, S.T., Ph.D, Prof. Dr. Ir. Drs. Syafwandi, M.Sc
Dr. Yulasmi, S.E., M.M, Dr. Robby Dharma, S.E., M.M
Dr. Syafri Arlis, S.Kom, M.Kom

Contoh 6.6: Penentuan Ukuran Sampel Untuk Estimasi σ

Sebuah sampel harus ditentukan untuk mengestimasi rata-rata gaji buruh dalam $\pm$ \$ 500 pada C = 0,99. Pimpinan serikat buruh menyatakan bahwa gaji buruh adalah \$ 26,000 s/d \$ 40,000. Berapa sampel yang dibutuhkan untuk menduga rata-rata gaji buruh dari populasinya?

$C = 0{,}99 \qquad \alpha = 1 - 0{,}99 = 0{,}01 \qquad Z_{\alpha/2} = 2{,}33$

$e = 500$

$\sigma_X = (40.000 - 26.000)/4 = 3.500$

$n_\sigma = \{(2{,}33)(3.500)/500\}^2 = 266{,}02$, dibulatkan menjadi 266 orang buruh.

2. Ukuran Sampel untuk Estimasi Proporsi Populasi

$$\boxed{n_p = \left(\frac{Z_{\alpha/2}^2\, pq}{e^2}\right)} \qquad \ldots\ldots (6.11)$$

di mana

$$e^2 = Z_{\alpha/2}^2 \frac{pq}{n}$$

Contoh 6.7: Penentuan Ukuran Sampel Untuk Estimasi Proporsi

Seorang manajer promosi memperkirakan bahwa proporsi konsumen yang menyaksikan iklan perusahaannya pada televisi berkisar antara 0,65 – 0,85. Pada C = 0,98, berapa sampel konsumen harus diteliti, agar interval estimasi proporsi berada dalam $\pm$ 5,00%?

$C = 0{,}98 \qquad Z_{\alpha/2} = 2{,}05 \qquad e = 0{,}05$

$n_p = (2{,}05)^2(0{,}65)(0{,}35)/(0{,}05)^2 = 382{,}427$ dibulatkan menjadi 382 konsumen.

6.7. Rangkuman

Untuk mengetahui parameter sebuah populasi, peneliti menggunakan parameter sampel. Melalui teknik penaksiran atau pendugaan, parameter populasi dapat diperkirakan nilainya melalui parameter sampel. Pembahasan mengenai penaksiran ini merupakan statistik inferensial. Ada dua jenis prosedur dalam estimasi, yaitu: (i) estimasi titik (*point estimation*); dan (ii) estimasi interval (*interval estimation*). Formula yang digunakan untuk menduga interval ⌠ sebuah populasi, dengan sampel kecil ($n \leq 30$):

$$\bar{X} - t_{⟨/2,v} \frac{S_X}{\sqrt{n}} \leq ⌠ \leq \bar{X} + t_{⟨/2,v} \frac{S_X}{\sqrt{n}} \qquad \ldots\ldots (6.5)$$

Untuk $n > 30$, formula yang digunakan adalah:

$$\bar{X} - z_{⟨/2,v} \frac{S_X}{\sqrt{n}} \leq ⌠ \leq \bar{X} + z_{⟨/2,v} \frac{S_X}{\sqrt{n}} \qquad \ldots\ldots (6.6)$$

Untuk pendugaan proporsi populasi, formula yang digunakan adalah:

$$Þ - z_{⟨/2,v} S_p \leq p \leq Þ + z_{⟨/2,v} S_p \qquad \ldots\ldots (6.9)$$

Ukuran Sampel untuk Estimasi Rata-rata Populasi

$$n_{⌠} = \left(\frac{Z_{⟨/2}\, ⌠_X}{e} \right)^2 \qquad \ldots\ldots (6.10)$$

Ukuran Sampel untuk Estimasi Proporsi Populasi

$$n_p = \left(\frac{Z_{⟨/2}^2\, pq}{e^2} \right) \qquad \ldots\ldots (6.11)$$

6.8. Diskusi

Estimator parameter populasi ditaksir melalui parameter sampel. Syarat estimator yang baik adalah: (a) Estimator ($\hat{\theta}$) dikatakan estimator yang tidak bias jika rata-rata semua kemungkinan harga $\hat{\theta}$ memiliki E{$\hat{\theta}$} = 0. Sebaliknya, estimator dianggap bias, jika E{$\hat{\theta}$} $\neq$ 0; (b) Konsisten. Estimator yang konsisten merupakan penduga yang berkonsentrasi secara sempurna pada parameter, jika sampel sebenarnya bertambah secara tidak terhitung; dan (c) Interval estimator minimum. Pada tataran praktek, kadang-kadang peneliti tidak mengetahui apakah populasi berdistribusi normal atau tidak. Untuk itu, biasanya ukuran sampel ditingkatkan sehingga n > 30. Dengan sampel besar ini, diasumsikan bahwa populasi berdistribusi normal, walaupun pada kenyataannya tidak normal. Para ahli statistik kemudian bersepakat (*rule of thumb*) seperti ini: jika σ_X tidak diketahui, namun ukuran sampelnya besar (n > 30), maka σ_X ditaksir dengan S_X.

6.9. Referensi

Liviatan, N., 1981, ***Consistent Estimation of Distributed Lags***, InternationalEconomicReview, vol. 4, Januari.

___________, 1982, ***Instumental Variables in Simultaneous Equations Model***, JohnWiley and Son Coy., Tokyo.

Miller, Irwin and John E. Freund, 2008, ***Probability And Statistics For Engineers***,7^{th} Edition, Prentice-Hall of India Private Limited, New Delhi.

Setyawan, HB, 2009. Statistika (Deskriptif dan Diferensial), Perpustakaan Nasional:Katalog dalm Terbitan ISBN 978-602-98107-0-7, Yayasan Dharma Nusantara, Jember.

6.10. Latihan Soal

Berikut data jarak tempuh yang dihasilkan BBM/liter pada mobil merk tertentu adalah:

15,4	18,3	19,1	20,4	21,0	21,6	19,3
16,3	18,7	19,4	20,5	21,3	21,8	20,8
16,7	18,7	19,4	20,6	21,4	22,5	22,0
17,2	19,0	19,8	20,9	21,5	23,2	20,1
18,1	19,0	20,4	21,0	21,6	25,0	17,8

C = 0,95. Berapa ⌠ jarak tempuh mobil yang dihasilkan BBM/liter?

Sebuah tim peneliti pasar mengamati 400 orang kosumenyang menggunakan kosmetik merk tertentu. Rata-rata penggunaan kosmetik dari sampel tersebut diketahui 14,2 pon per bulan, dan standard penyimpangannya = 1,8 pon per bulan. Buatlah estimasi interval untuk rata-rata populasi penggunaan kosmetik tersebut pada C = 95.

Standard penyimpangan pengisian minuman ringan ke dalam botol oleh mesin pengisi botol = 1,8 ml. Dipilih secara random 8 botol yang telah terisi sebagai sampel, diukur isinya, hasilnya adalah (dalam ml):

481 479 482 480 477 478 481 482

Jika isian ini berdistribsi normal, pada tingkat keyakinan = 0,90 berapa interval estimasi rata-rata isian populasinya?

Sebelum melakukan produksi pada kapasitas maksimum, perusahaan meneliti 400 konsumen yang dipilih secara random. Dari sampel itu, ternyata 80 orang konsumen yang membeli produk perusahaan. Dengan tingkat keyakinan = 0,95, estimasikan proporsi dari populasi konsumen akan membeli produk perusahaan.

Pada tingkat keyakinan = 0,95, berapa ukuran sampel yang diperlukan untuk mengestimasi interval proporsi dalam rentang $\pm$ 0,02 ?

BAB 7

UJI HIPOTESIS

Hipotesis adalah sebuah kesimpulan yang sementara dianggap benar, sehingga perlu dilakukan uji untuk meyakini kebenaran itu. Hipotesis dibangun berdasar landasan teori, hasil penelitian empiris atau dugaan analis.

Kemampuan Akhir yang Diharapkan::
- Mahasiswa mampu memahami bagaimana menyusun hipotesis, - Mampu melakukan uji hipotesis dengan metode parametrik maupun non parametrik.

7.1 Uji Hipotesis dengan Metode Parametrik

Hipotesis: adalah sebuah kesimpulan sementara yang perlu diuji kebenarannyamelalui uji statistik. Ada dua jenis hipotesis, yaitu : (a) hipotesis nol (*null hypothesis* atauH_0); dan (b) hipotesis alternatif (*alternative hypothesis* atau H_a). Hipotesis nol merupakanpernyataan yang berisi kesamaan, ketidak bedaan, atau pernyataan-pernyataan lain yang menyiratkan kenihilan. Hipotesis altenatif merupakan komposit dan bersifat *mutuallyexclusive* terhadap hipotesis nol. Komposit mengartikan bahwa tidak ada ruang sampel yang terbuang atau hilang, sedang *mutually exclusive* mensyaratkan kedua jenis hipotesis tersebut saling mematikan. Jika hipotesis nol yang diterima, maka hipotesis alternatif harus ditolak, dan sebaliknya.

Selain mengestimasi nilai sebuah parameter, persoalan-persoalan lain penting adalahbagaimana menentukan kebenaran sebuah pernyataan tentang sebuah parameter, untuk itulah perlu diuji hipotesis tentang parameter tersebut. Pengujian hipotesis ini bisa menghasilkan empat macam keputusan, yaitu:

- hipotesis benar dan diterima,
- hipotesis salah dan ditolak,
- hipotesis benar dan ditolak,
- hipotesis salah dan diterima.

	Menerima Hipotesis	Menolak Hipotesis
Hipotesis benar	Keputusan benar	Kesalahan tipe I (α)
Hipotesis salah	Kesalahan tipe II (β)	Keputusan benar

Ada lima langkah umum yang dilakukan dalam menguji hipotesis dengan sistimatis,yaitu:

(1) memformulasikan atau mendefinisikan H_0 dan H_a. Ada dua kemungkinan pengujian statistik untuk hipotesis, yaitu : uji satu sisi atau uji dua sisi. Uji satu sisi jika H_a menyatakan lebih besar atau lebih kecil terhadap parameter yang diuji (>, <, >, <); dan uji dua sisi jika H_a menyatakan ketidak samaan ($\neq$).

(2) Menentukan probabilitas kesalahan tipe I (〈). Penentuan nilai 〈 sangat dipengaruhi harapan si penguji hipotesis terhadap hasil uji. Untuk uji hipotesis dalam bidang-bidang ilmu sosial, umumnya nilai 〈 beriksar antara 5,00% – 10,00%. Prinsipnya, makin kecil 〈, makin akurat hasil uji hipotesisnya, sebab makin kecil peluang terjadinya kesalahan tipe I.

(3) Berdasar distribusi sampel dari uji statistik yang dilakukan, tentukan kriteria ujinya.

(4) Hitung dari data nilai-nilai statistik yang dijadikan basis keputusan uji.
(5) Buat keputusan menolak atau menerima hipotesis yang diuji.

7.2 Contoh-contoh Aplikasi Metode Parametrik

1. Uji hipotesis yang berkaitan dengan satu rata-rata.

Untuk sampel besar (n > 30), daerah kritis pengujian:

Tabel 7.1 Daerah Kritis Uji Hipotesis Sebuah Rata-rata.

H_0	Tolak H_0 jika:
$\mu < \mu_0$	$Z < -Z_\alpha$
$\mu > \mu_0$	$Z > Z_\alpha$
$\mu \neq \mu_0$	$Z < -Z_{\alpha/2}$ atau $Z > Z_{\alpha/2}$

Di mana,

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \qquad \text{. . . . . (7.1)}$$

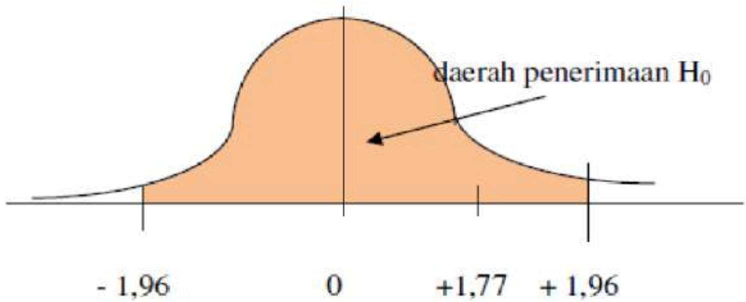
<u>Contoh-7.1</u> : Uji Hipotesis (satu sisi) Untuk Sebuah Rata-rata.

Standard konduksi termal dari sejenis batu bara adalah = 0,340. Dari hasil produksi batu bara dipilih sejumlah (n) = 35 unit batu bara untuk diuji apakah rata-rata konduksi termalnya memenuhi standard pada tingkat keyakinan (α) = 0,05. Informasi mengenai σ konduksi termal dari sebuah penelitian yang pernah dilakukan, diketahui = 0,010. Rata-rata konduksi termal batu bara sampel dihitung, hasilnya = $\bar{X}$ = 0,343.

$H_0 : \mu = 0,340$

$H_a : \mu \neq 0,340$

$$Z = \frac{0{,}343 - 0{,}340}{0{,}010/\sqrt{35}} = 1{,}77$$



Gambar 7.1 Daerah Penerimaan H_0

Z_{hitung} berada dalam daerah penerimaan H_0, maka H_0 diterima dan H_a ditolak. Sehingga dapat disimpulkan bahwa sampel batu bara yang diteliti tidak memenuhi standard konduksi termal yang diharapkan.

(Angka + 1,96 bisa dilihat pada Tabel Z pada 〈 = 0,025, karena ini uji dua sisi).

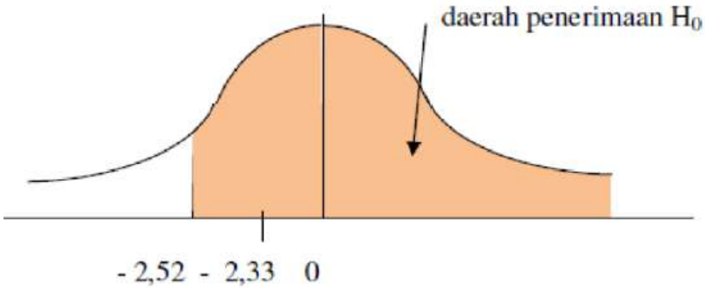
Contoh-7.2 : Uji Hipotesis (dua sisi) Untuk Sebuah Rata-rata

Sebuah perusahaan angkutan barang meragukan pernyataan pabrik ban yang menyatakan bahwa ban hasil produksi pabrik tersebut memiliki rata-rata daya pakai paling sedikit 28.000 km. Untuk menguji pernyataan tersebut, perusahaan mencoba menggunakan 40 buah ban jenis itu untuk truk armadanya, dan hasil percobaannya menunjukkan bahwa rata-rata daya pakai ban = 27.463 km dengan standard deviasi = 1.348 km. Dengan 〈 = 0,01.

$H_0 : \mu > 28.000$

$H_a : \mu < 28.000$

$$Z = \frac{27.463 - 28.000}{1.348/\sqrt{40}} = -\,2{,}52$$



Gambar 7.2 Daerah Penerimaan H_0.

H_0 ditolak, H_a diterima; artinya pernyataan pabrik ban tentang daya pakai ban hasil produksinya tidak terbukti.

Untuk sample kecil (n < 30), prosedur uji hipotesisnya sama, dengan perbedaan pada kriteria ujinya, yaitu:

$$\boxed{t = \frac{\overline{X} - \mu_0}{s/\sqrt{n}}} \qquad \text{. . . . . (7.2)}$$

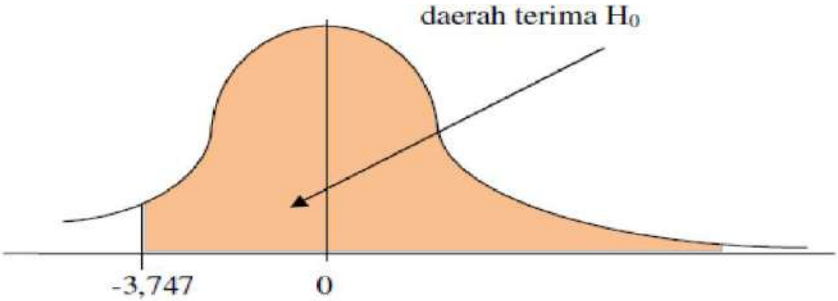
<u>Contoh–7.3</u> : Uji Hipotesis Untuk Sebuah Rata-rata (sampel kecil)

Spesifikasi pita yang memenuhi standard kualitas adalah jika pita tersebut mampu menahan beban sampai 180 pon. Jika lima helai pita yang dipilih secara random dari gudang diuji kekuatannya, dihasilkan bahwa rata-rata dapat menahan beban = 169,5 pon dengan standard deviasi = 5,7 pon. α yang digunakan adalah = 0,01.

H_0 : ⌠ = 180 pon

H_a : ⌠ < 180 pon → indikasi uji sisi sebelah kiri

$$t = \frac{169{,}5 - 180}{5{,}7/\sqrt{5}} = -\,4{,}12$$



Gambar 7.3 Daerah Penerimaan H0 untuk Uji Satu Sisi

`

t_{hitung} = - 4,12 > $t_{0,01;4}$ (lihat Tabel-t, sebesar = 3,747), maka H_0 ditolak dan menerima H_a.

2. Uji hipotesis yang berkaitan dengan dua rata-rata.

Untuk sampel besar (n > 30), populasi normal, σ_1 dan σ_2 diketahui, daerah kritis pengujian adalah:

Tabel 7.2. Daerah Kritis Uji Hipotesis Dua Rata-rata.

H_0	Tolak H_0 jika:
⌠$_1$-⌠$_2$ <™	Z< -Z$_〈$
⌠$_1$- ⌠$_2$ >™	Z > Z$_〈$
⌠$_1$- ⌠$_2$ $\overline{™}$	Z < - Z$_{〈/2}$ atau Z > Z$_{〈/2}$

Formula uji hipotesis perbedaan dua rata-rata :

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - ™}{\sqrt{(⌠_1^2/n_1 + ⌠_2^2/n_2)}} \qquad . . . . . (7.3)$$

<u>Contoh-7.4</u> : Uji Hipotesis untuk Dua Rata-rata

Untuk menguji pernyataan bahwa resistensi kabel listrik bisa dikurangi sampai lebih dari 0,050 ohm dengan cara membuat campuran logamnya; dipilih sampel sebanyak 32 kabel listrik tanpa logam campuran dan 32 jenis kabel listrik dengan logam campuran. Rata-rata resistensi untuk kelompok-1, $\overline{X}_1$ = = 0,136 ohm dengan s_1 = 0,004 ohm, X_2 = 0,083 ohm dengan s_2 = 0,005 ohm, Pada 〈 = 0,05, ujilah apakah pernyataan tersebut benar?

H_0 : ⌠$_1$ - ⌠$_2$ = 0,050
Ha : ⌠$_1$- ⌠$_2$ > 0,050

$$Z = \frac{0{,}136 - 0{,}083 - 0{,}050}{\sqrt{(0{,}0042/32 + 0{,}0052/32)}} = 2{,}65$$

Karena $Z_{hitung} > Z_{0,05}$ (2,65 > 1,645), maka H_0 ditolak dan menerima H_a. Pernyataan itu terbukti benar secara statistik.

<u>Contoh-7.5</u> : Uji Hipotesis untuk Dua Rata-rata (sampel besar)

Sebuah perusahaan lampu yang terkemuka menyatakan bahwa bola lampu hasil produksinya lebih tahan lama daripada produksi perusahaan pesaingnya. Untuk itu perlu diuji dengan mengambil sampel bola lampu hasil produksinya sebanyak 40 buah dan 40 buah bola lampu hasil produksi perusahaan pesaing. $\overline{X}_1$ = 647 jam dengan s_1 = 27 jam, $\overline{X}_2$ = 638 jam dengan s_2 = 31 jam.

〈 = 0,05

H_0 : ⌠$_1$- ⌠$_2$ = 0
H_a : ⌠$_1$- ⌠$_2$ > 0

$$Z = \frac{647 - 638 - 0}{\sqrt{(0{,}272/40 + 312/40)}} = 1{,}38$$

Karena $Z_{hitung} < Z_{0,05}$ (1,38 < 1,645), maka H_a ditolak dan menerima H_0. Pernyataan itu tidak terbukti benar secara statistik.

Untuk sampel kecil (n < 30), maka formula ujinya menjadi :

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - \text{™}}{\sqrt{[(n_1-1)s_1^2 + (n_2-1)s_2^2]}} \sqrt{\frac{n_1.n_2\,(n_1 + n_2 - 2)}{n_1 + n_2}}$$

…. (7.4)

<u>Contoh-7.6</u> : Uji Hipotesis untuk Dua Rata-rata (sampel kecil)

Berikut ini adalah data sampel random kapasitas panas batubara (kalori/ton) dari dua tambang batubara yang berbeda.

Tambang-1 : 8.260 8.130 8.350 8.070 8.340

Tambang-2 : 7.950 7.890 7.900 8.140 7.920 7.840

Ujilah apakah kapasitas panas kedua jenis batubara itu berbeda pada 〈 = 0,01.

$H_0 : \bar{X}_1 - \bar{X}_2 = 0$

$H_a : \bar{X}_1 - \bar{X}_2 \neq 0$

$\bar{X}_1$ = 8.230 dengan s1 = 63.000/4 = 15.750

$\bar{X}_2$ = 7.940 dengan s2 = 54.600/5 = 10.920

$$t = \frac{(8.230 - 7.940) - 0}{\sqrt{(63.000 + 54.000)}} \sqrt{\frac{(5)(6)(5 + 6 - 2)}{5 + 6}}$$

$= 4{,}19$

Nilai $t_{0.05}$ dengan derajat bebas = $n_1 + n_2 - 2 = 5 + 6 - 2 = 9$, sama dengan = 3,250 (periksa Tabel-t). $t_{hitung} > t_{0.05}$ (4,19 > 3,230), maka H_0 ditolak dan menerima H_a.

Uji hipotesis untuk dua rata-rata ini dapat juga diaplikasi untuk data berpasangan. Formula uji berubah menjadi :

$$t = \frac{\overline{X} - \mu_0}{s/\sqrt{n}} \qquad \text{. . . . . (7.5)}$$

<u>Contoh-7.7</u> : Uji Hipotesis untuk Dua Rata-rata Berpasangan (sampel kecil)

Berikut ini data rata-rata jam kerja yang hilang karena adanya kecelakaan kerja dalampabrik pada 10 perusahaan, pada saat sebelum program keselamatan kerja diberlakukan dan sesudah diberlakukan.

45 dan 36 73 dan 60 46 dan 44 124 dan 119 33 dan 35

57 dan 51 83 dan 77 34 dan 29 26 dan 24 17 dan 11

Ujilah pada α = 0,05, apakah program keselamatan kerja itu efektif menurunkan kecelakaan kerja dalam pabrik.

$H_0 : \mu = 0$

$H_a : \mu > 0$

Selisih jam kerja yang hilang pada setiap pabrik dihitung sebagai data tunggal. Rata-rata selisih kehilangan jam kerja = 5,2 jam dengan standard deviasi, s = 4,08 jam.

$$t = \frac{5,2 - 0}{4,08/\sqrt{10}} = 0,403$$

$t_{hitung} < t_{0,05}$ pada derajat bebas n - 1 = 10 – 1 = 9 (= 1,833); maka tolak H_a dan menerima H_0. Dengan demikian program keselamatan kerja tersebut terbukti tidak efektif dalam menurunkan kecelakaan kerja dalam pabrik.

3. Uji Hipotesis untuk Sebuah Varians.

Untuk sampel random (n < 30) dari populasi normal, σ_0^2 diketahui sebagai nilai darivariabel random yang berdistribusi χ^2 dengan derajat bebas = n - 1, daerah kritis pengujian:

Tabel 7.3 Daerah Kritis Uji Hipotesis Sebuah Varians

H_0	Tolak H_0 jika:
$\sigma^2 < \sigma_0^2$	$\chi^2 < \chi^2_{1-\alpha}$
$\sigma^2 > \sigma_0^2$	$\chi^2 > \chi^2_{\alpha}$
$\sigma^2 \neq \sigma_0^2$	$\chi^2 < \chi^2_{1-\alpha/2}$ atau $\chi^2 > \chi^2_{\alpha/2}$

Formula uji hipotesis sebuah varians:

$$\boxed{\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}} \qquad \text{. . . . .(7.6)}$$

<u>Contoh-7.8</u> : Uji Hipotesis untuk Sebuah Varians (sampel kecil)

Pada proses *lapping* yang digunakan untuk menghaluskan kotak silikon tertentu agar memiliki ketebalan standard, dapat dikatakan baik jika standard deviasi ketebalan silikon tersebut paling banyak 0,50 mm. Pada α = 0,05; ujilah pada 15 buah kotak silikon yang memiliki standard deviasi ketebalan = 0,64 mm.

H_0 : σ = 0,50
H_a : σ > 0,05

$$\chi^2 = \frac{(15 - 1)(0{,}642)}{0{,}50^2} = 22{,}94$$

$\chi^2_{0,05}$ pada tabel χ^2 dengan derajat bebas 15 – 1 = 14, sama dengan = 23,683; maka hipotesis nol diterima dan H_a ditolak.

Dengan demikian poses *lapping* itu memenuhi standard karena σ ketebalan silikon yang dihasilkan tidak melebihi 0,50 mm.

Untuk sampel besar (n > 30), formula uji hipotesis yang digunakan adalah:

$$\boxed{Z = \frac{s - \sigma_0}{\sigma_0/\sqrt{2n}}} \qquad \text{. . . . . (7.7)}$$

<u>Contoh-7.9</u> : Uji Hipotesis untuk Sebuah Varians (sampel besar)

Spesifikasi teknis pada produksi massa gir roda sepeda motor, mensyaratkan bahwa kualitas gir dianggap baik jika standard deviasi diameter tidak melebihi 0,0040 cm. Dari sampel sebanyak 35 gir, diketahui bahwa standard deviasi diameter = 0,0053 cm. Pada α = 0,01, ujilah kualitas gir tersebut apakah memenuhi spesifikasi atau tidak.

H_0 : σ = 0,0040

H_a : σ > 0,0040

$$Z = \frac{0{,}0053 - 0{,}0040}{0{,}0040/\sqrt{2(35)}} = 2{,}70$$

$Z_{0,01}$ pada tabel Z = 2,33, maka terima H_a dan tolak H_0. Artinya kualitas gir tidak memenuhi spesifikasi yang disyaratkan; karena standard deviasi diameter terbukti melebihi 0,0040 cm.

4. Uji hipotesis untuk dua varians.

Formula uji untuk dua varians adalah :

$$\boxed{F = \frac{s_1^2}{s_2^2}} \qquad \text{. . . . . (7.8)}$$

Daerah kritis pengujian:

Tabel 7.4 Daerah Kritis Uji Hipotesis Dua Varians

H_0	Uji Statistik	Tolak H_0 jika:
$\sigma^2 < \sigma_0^2$	$F = s_2^2/s_1^2$	$F > F_{\alpha};(n_2\text{-}1,\ n_1-1)$
$\sigma^2 > \sigma_0^2$	$F = s_1^2/s_2^2$	$F > F_{\alpha};(n_1\text{-}1,\ n_2-1)$
$\sigma^2 \neq \sigma_0^2$	$F = s_M^2/s_m^2$	$F > F_{\alpha/2};\ (n_M\text{-}1,\ n_m-1)$

<u>Contoh-7.10</u> : Uji Hipotesis Untuk Dua Varians

Pelapisan perak untuk kaleng bahan kimia padat yang dilakukan oleh Perusahaan A dan Perusahaan B diyakini memiliki variasi ketebalan yang berbeda. Untuk itu dipilih masing-masing 12 sampel hasil pelapisan perak dari kedua perusahaan itu. Standard penyimpangan ketebalan pelapisan Perusahaan A (s_1) terhitung = 0,035 mm, pada Perusahaan B (s_2) terhitung = 0,062 mm.

$H_0 : s_1^2 = s_2^2$

$H_a : s_1^2 < s_2^2$

Pada α = 0,05:

$$F = \frac{0{,}062^2}{0{,}035^2} = 3{,}14$$

$F_{0,05}$ pada tabel F dengan numerator = $n_1 - 1 = 11$ dan denominator = $n_2 - 1 = 11$ adalah sebesar 2,85. Karena $F_{hitung} > F_{0,05}$ pada tabel tersebut, maka H_0 ditolak dan H_a diterima; artinya variabilitas ketebalan lapisan perak hasil produksi Perusahaan A lebih rendah daripada variabilitas lapisan perak hasil produksi Perusahaan B.

Syafiadi Rizki Abdila, S.T., Ph.D, Prof. Dr. Ir. Drs. Syafwandi, M.Sc
Dr. Yulasmi, S.E., M.M, Dr. Robby Dharma, S.E., M.M
Dr. Syafri Arlis, S.Kom, M.Kom

<u>Contoh-7.11</u> : Uji Hipotesis Untuk Dua Varians

Kembali pada Contoh-7.6, data sampel random kapasitas panas batubara (kalori/ton) dari dua tambang batubara yang berbeda adalah:

Tambang-1 : 8.260 8.130 8.350 8.070 8.340

Tambang-2 : 7.950 7.890 7.900 8.140 7.920 7.840

s_1^2 = 15.750 dan s_2^2 = 10.920. Ujilah pada 〈 = 0,02; apakah variabilitas kapasitas panas batubara pada Tambang-1 berbeda dengan batubara pada Tambang-2.

$H_0 : s_1^2 = s_2^2$

$H_a : s_1^2 \neq s_2^2$

$F_{0,02}$ dalam Tabel F pada derajat bebas numerator = 4 dan denominator = 5 sebesar 11,4.

$$F = \frac{15.750}{10.920} = 1,44$$

Karena $F_{hitung} < F_{0,02}$ pada Tabel F tersebut, maka H_0 diterima dan H_a ditolak; artinya variabilitas kapasitas panas batubara pada Tambang-1 sama dengan variabilitas kapasitas panas batubara pada Tambang-2.

5. Uji Hipotesis untuk Sebuah Proporsi.

Banyak metode yang digunakan untuk pemeriksaan dalam kendali mutu, dan uji kehandalan dengan berbasis pada uji hipotesis nol yang menyatakan proporsi sama dengan konstanta tertentu.

Untuk sampel besar digunakan formula uji sebagai berikut :

$$z = \frac{X - np_0}{\sqrt{np_0 (1 - p_0)}} \qquad \text{. . . . . (7.9)}$$

<u>Contoh-7.12</u> : Uji Hipotesis Untuk Sebuah Proporsi (sampel besar)

Dalam sebuah penelitian yang didesain untuk menguji apakah detonator merk tertentu yang digunakan untuk peledakan pada tambang dapat berfungsi dengan benar paling sedikit 90,00% dari seluruh detonator yang dipakai. Dari 200 detonator ternyata 174 buah detonator yang berfungsi baik. Ujilah apakah tujuan penelitian itu terjawab pada α = 0,05.

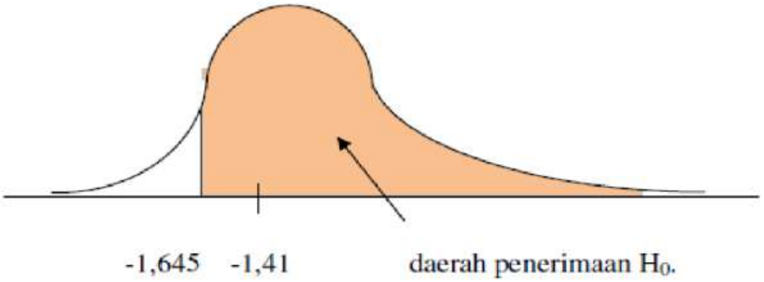
$H_0 : p = 0,90$

$H_a : p < 0,90$

Pada α = 0,05, nilai z = -1,645

$$z = \frac{174 - 200(0,90)}{\sqrt{200(0,90)(1 - 0,90)}} = -1,41$$

Karena $z_{hitung} > -1,645$, maka H_0 diterima, dan H_a ditolak. Dengan demikian, dapat dinyatakan bahwa paling sedikit 90,00% detonator merk tertentu tersebut berfungsi dengan baik.



Gambar 7.4 Daerah Penerimaan H_0 untuk Uji Sebuah Proporsi

6. Uji Hipotesis untuk Banyak Proporsi

Formula uji:

$$\boxed{\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(o_{ij} - e_{ij})^2}{e_{ij}}} \quad \ldots\ldots (7.10)$$

$H_0 : p_{i1} = p_{i2} = \ldots . p_{ic}$ (i = 1, 2, . . . r).

$H_a : p_{i1} \neq p_{i2} \neq \ldots . p_{ic}$

Syafiadi Rizki Abdila, S.T., Ph.D, Prof. Dr. Ir. Drs. Syafwandi, M.Sc
Dr. Yulasmi, S.E., M.M, Dr. Robby Dharma, S.E., M.M
Dr. Syafri Arlis, S.Kom, M.Kom

Di mana p_{ij} = probabilitas data pada baris-i dan kolom-j. $\overset{r}{\underset{i=1}{©}} p_{ij} = 1$

$e_{ij} = n_j . p = (n_j . X)/n$

Penyusunan tabel kontingensi menurut aturan : variabel yang dipengaruhi diletakkan sebagai baris, sedang yang mempengaruhi diletakkan sebagai kolom. Koefisien kontingensi merupakan ukuran keeratan saling ketergantungan kedua variabel (*dependency*) dapat dihitung dengan rumus:

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

makin tinggi nilai C makin tinggi keeratan saling ketergantungannya.

<u>Contoh-7.13</u> : Uji Hipotesis Untuk Banyak Proporsi-1

Sampel dari tiga jenis bahan, dipanasi sampai temperatur yang tinggi. Data perubahan secara fisik terhadap bahan-bahan tersebut adalah:

Tabel 7.5 Perubahan Fisik Bahan Setelah Pemanasan

	Bahan A	Bahan B	Bahan C	Total
Berkerut	41	27	22	90
Tidak Berubah	79	53	78	210
Total	120	80	100	300

Ujilah pada 〈 = 0,05, apakah pemanasan dengan temperatur tinggi tersebut
memberikan kesamaan kecenderungan pengkerutan bahan..

$H_0 : p_1 = p_2 = p_3$

$H_a : p_1 \neq p_2 \neq p_3$

$|^2$ pada 〈 = 0,05 dengan derajat bebas = k – 1 = 3 – 1 = 2, adalah = 5,991.

Hitung terlebih dahulu e_{ij} :

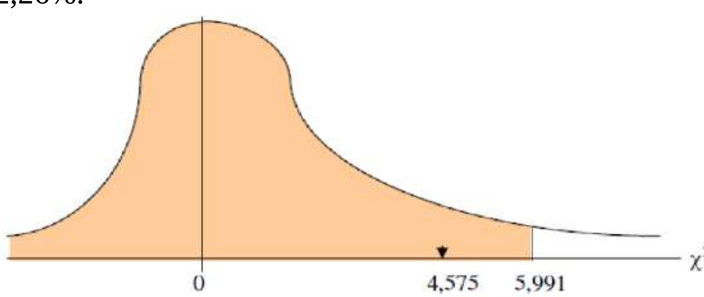
e_{11} = (90)(120)/300 = 36 e_{12} = (90)(80)/300 = 24 dan seterusnya. Hasil perhitungan e_{ij} adalah:

	Bahan A	Bahan B	Bahan C
Berkerut	36	24	30
Tidak Berubah	84	56	70

$$|^2 = \frac{41\text{-}36)^2}{36} + \frac{(27\text{-}24)^2}{24} + \frac{(22\text{-}30)^2}{30} + \frac{(79\text{-}84)^2}{84} + \frac{(53\text{-}56)^2}{56} + \frac{(78\text{-}70)^2}{70}$$

= 4,575

C = koefisien kontigensi = √(4,575)?(4,575+300) = 0,1226 atau 12,26%.



Gambar 7.5 Daerah Penerimaan H_0 Uji Hipotesis Banyak Proporsi

Kesimpulan: $|^2$ hasil perhitungan < $|^2$ pada Tabel $|^2$, maka H_0 diterima; artinya kemungkinan berkerut karena perubahan temperatur pada ketiga jenis bahan tersebut, sama. Atau lebih ekstrim lagi dapat dikatakan bahwa kualitas ketiga jenis bahan adalah sama. Nilai koefisien kontingensinya tidak berarti.

<u>Contoh-7.14</u> : Uji Hipotesis Untuk Banyak Proporsi-2

Untuk menentukan apakah ada hubungan antara kinerja pegawai dengan prestasi hasil pendidikan dan pelatihan yang dilakukan oleh kantor, digunakan sampel pegawai sebanyak 400 orang yang telah mengikuti program DIKLAT yang dimaksud. Hasil observasi pada file kepegawaian (400 *records*) adalah:

Tabel 7.6 Prestasi Pada Program DIKLAT dan Kinerja Pegawai

Kinerja Pegawai	Prestasi DIKLAT			Total
	Di bawah Rata-rata	Rata-rata	Di atas Rata-rata	
Baik	23	60	29	112
Sedang	28	79	60	167
Buruk	9	49	63	121
Total	60	188	152	400

H_0 : tidak ada hubungan antara prestasi DIKLAT dengan kinerja pegawai.
H_a : ada hubungan antara prestasi DIKLAT dengan kinerja pegawai.

Pada α = 0,01 dan df = (r-1)(k-1) = 4, nilai χ^2 adalah = 13,277 (lihat Tabel χ^2). Hasil perhitungan e_{ij} dapat ditabel sebagai berikut :

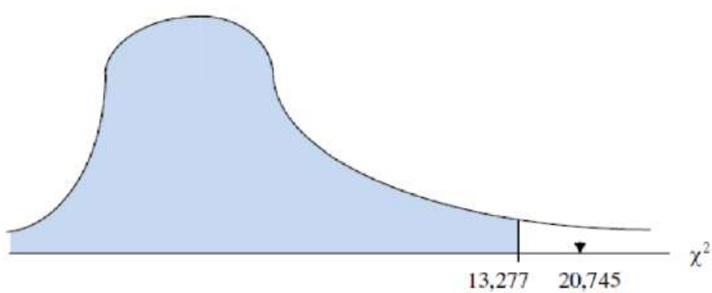
Tabel 7.7 Hasil Perhitungan e_{ij}

Kinerja Pegawai	Prestasi DIKLAT		
	Di bawah Rata-rata	Rata-rata	Di atas Rata-rata
Baik	16,80	52,64	42,56
Sedang	25,05	78,49	63,46
Buruk	18,15	56,87	45,98

$$\chi^2 = \frac{(23-16{,}80)^2}{16{,}80} + \frac{(60-52{,}64)^2}{52{,}64} + \frac{(29-42{,}56)^2}{42{,}56} + \frac{(28-25{,}05)^2}{25{,}05} +$$

$$\frac{(79-78{,}49)^2}{78{,}49} + \frac{(60-63{,}46)^2}{63{,}46} + \frac{(9-18{,}15)^2}{18{,}15} + \frac{(49-58{,}87)^2}{58{,}87} + \frac{(63-45{,}98)^2}{45{,}98}$$

$$= 20{,}745.$$



Gambar 7.6 Daerah Penerimaan H_0 Uji Hipotesis Banyak Proporsi

Kesimpulan : χ^2 hasil perhitungan > χ^2 pada tabel, maka H_a diterima; artinya adahubungan antara prestasi hasil Program Pendidikan & Pelatihan dengan kinerja pegawai. Tingkat keeratan hubungan = 0,2220482 atau 22,20%.

<u>Contoh-7.15</u> : Uji Hipotesis Untuk Banyak Proporsi-3

Untuk menguji kualitas empat merk ban dilakukan pengamatan pada laboratoriumterhadap 800 ban, hasilnya adalah (o_{ij}) :

Tabel 7.8 Data Kerusakan Ban

	Tingkat Kerusakan Ban				Total
	Merk A	Merk B	Merk C	Merk D	
Rusak sebelum 20.000 mil	26	23	15	12	96
Rusak setelah 20.000 mil s/d 30.000 mil	118	93	116	121	448
Rusak setelah 30.000 mil	56	84	69	47	256
Total	200	200	200	200	800

α = 0,01, df = (3-1)(4-1) = 6, nilai χ^2 pada Tabel χ^2 =16,812

Hasil perhitungan nilai ekspektasinya (e_{ij}) adalah :

Tabel 7.9 Nilai Ekspektasi Data Kerusakan Ban

	Tingkat Kerusakan Ban			
	Merk A	Merk B	Merk C	Merk D
Rusak sebelum 20.000 mil	24	24	24	24
Rusak setelah 20.000 mil sampai 30.000 mil	112	112	112	112
Rusak setelah 30.000 mil	64	64	64	64

$$\chi^2 = \left(\frac{(26-24)^2}{24} + \frac{(23-24)^2}{24} + \frac{(15-24)^2}{24} + \frac{(32-24)^2}{24} + \frac{(118-112)^2}{112} + \frac{(93-112)^2}{112} + \frac{(116-112)^2}{112} + \frac{(121-112)^2}{112} + \frac{(56-64)^2}{64} + \frac{(84-64)^2}{64} + \frac{(69-64)^2}{64} + \frac{(47-64)^2}{64} \right) = 16{,}567$$

χ^2 hasil perhitungan < χ^2 pada tabel, maka H_0 diterima, artinya proporsi kerusakan ban pada keempat merk ban tersebut adalah sama. Dengan kata lain, bahwa kualitas ban tidak berhubungan dengan merk ban.

7. *Goodness of Fit Test*

<u>Contoh-7.16</u> : *Goodness of Fit* (Distribusi Poisson).

Pada menara kontrol di lapangan terbang, dalam rentang waktu 5 menitan, diamati 400 kali panggilan radio untuk mengetahui apakah panggilan radio tersebut berdistribusi Poisson dengan rata-rata (λ) = 4,6. Hasil pengamatannya adalah:

Tabel 7.10 Data Frekuensi Pannggilan Radio

Jumlah Penggilan Radio	Frekuensi Aktual (o_{ij})	Probabilkitas Poisson	Ekspektasi Frekuensi (e_{ij})
0	3 18	0,010	4,00 22,40
1	15	0,046	18,40
2	47	0,107	42,80
3	76	0,163	65,20
4	68	0,187	74,80
5	74	0,173	69,20
6	46	0,132	52,80
7	39	0,087	34,80
8	15	0,050	20,00
9	9	0,025	10,00
10	5	0,012	4,80
11	2 8	0,005	2,00 8,00
12	0	0,002	0,80
13	1	0,001	0,40
Total	400		400

Penjelasan :

Frekuensi yang terjadi (o_{ij}) merupakan hasil observasi. Probabilitas Poisson dapat dilihat pada Tabel Distribusi Poisson, dan merupakan probabilitas akumulasi pada ⎣ = 4,6. Ekspekstasi frekuensi : e_{ij} = o_{ij} x probabilitas. Pada kolom o_{ij}, tidak diperkenankan nilai di bawah 5, untuk itu perlu dikumulatifkan agar > 5.

H_0 : frekuensi panggilan radio berdistribusi Poisson dengan λ = 4,6

H_a : frekuensi panggilan radio tidak berdistribusi Poisson dengan λ = 4,6

$$\chi^2 = \frac{(18-22{,}4)^2}{22{,}4} + \frac{(47-42{,}80)^2}{42{,}80} + \frac{(76-65{,}20)^2}{65{,}20} + \frac{(68-74{,}80)^2}{74{,}80} + \frac{(74-69{,}20)^2}{69{,}20} + \frac{(46-52{,}80)^2}{52{,}80} + \frac{(39-34{,}80)^2}{34{,}80} + \frac{(15-20{,}00)^2}{20{,}00} + \frac{(9-10{,}00)^2}{10{,}00} + \frac{(8-8{,}00)^2}{8{,}00}$$

$= 6{,}749$,

Pada α = 0,01 dan df = k – m = 10 – 1 = 9 (di mana k = banyaknya observasi yang diperbandingkan dan m = adalah banyaknya variabel), χ^2 pada tabel = 16,919. χ^2 hasil perhitungan < χ^2 pada tabel, maka H_0 diterima; artinya panggilan radio yang datang ke menara kontrol tersebut berdistribusi Poisson dengan λ = 4,6.

Contoh-7.17 : *Goodness of Fit* (Distribusi Normal).

Armor Carpet Store adalah sebuah toko yang menjual karpet, memperkirakan bahwarata-rata penjualan karpet per minggu (μ) = 4.200 yard dengan standard deviasi (σ) = 1.232yard. Parameter ini diperoleh dari pengamatan selama 10 minggu terakhir:

Tabel 7.11 Data Penjualan Karpet.

Minggu	Penjualan (yard)	$(X_t - \bar{X})$	$(X_t - \bar{X})^2$
1	2.900	-1.300	1.690.000
2	5.400	1.200	1.440.000

3	3.100	-1.100	1.210.000
4	4.700	500	250.000
5	3.800	-400	160.000
6	4.300	100	10.000
7	6.800	2.600	6.760.000
8	2.900	-1.300	1.690.000
9	3.600	-600	360.000
10	4.500	300	90.000
Rata-rata	4.200	Total	13.660.000

Varians = 1.517.777,78 yard, maka standard deviasi = 1.231,981, dibulatkan = 1.232 yard. Diasumsikan bahwa data penjualan karpet berdistribusi normal. Untuk itu dibutuhkan pengamatan yang lebih banyak, yaitu = 200 minggu, hasilnya adalah frekuensi *range* penjualan sebagai berikut::

Tabel 7.12 Distribusi Data Penjualan

No.	*Range* Penjualan/minggu	Frekuensi
1	0 – 999	2
2	1.000 – 1.999	5
3	2.000 – 2.999	22
4	3.000 – 3.999	50
5	4.000 – 4.999	62

6	5.000 – 5.999	40
7	6.000 – 6.999	15
8	7.000 – 7.900	3
9	$\geq$ 8.000	1
	Total	200

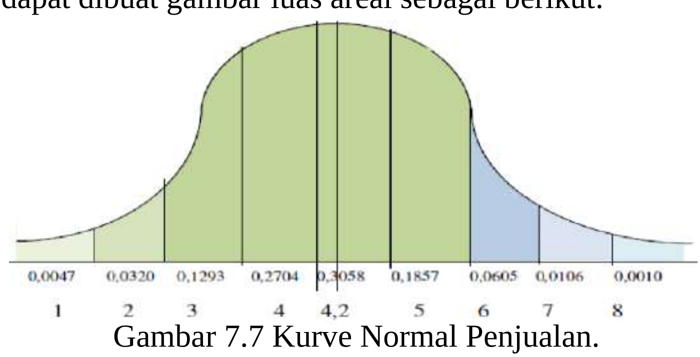
H_0 : data penjualan berdistribusi normal.

H_a : data penjualan tidak berdistribusi normal.

langkah-1: menghitung probabilitas setiap *range* penjualan dengan menggunakan μ = 4.200 dan σ = 1.232. Contoh untuk menghitung probabilitas penjualan = 1.000 yard:

$$Z = \frac{1.000 - 4.200}{1.232} = -2{,}60 \rightarrow P\{1.000\} = 0{,}4953$$

Luas areal di bawah kurve bisa dilihat pada Tabel Distribusi Normal. Luas = 0,4953 merupakan luas areal di bawah kurve normal untuk penjualan dari 1.000 yard sampai dengan rata-ratanya. Sehingga luas areal di bawah kurve normal untuk penjualan < 1.000 yard = 0,5000 – 0,4953 = 0,0047. Analog dengan cara menghitung luas areal di bawah kurve normal tersebut dapat dibuat gambar luas areal sebagai berikut:



Gambar 7.7 Kurve Normal Penjualan.

Tabel 7.13 Perhitungan Frekuensi Normal

Range	Z	Area $X_t - \bar{X}$	Range Area	Frekuensi Normal
0 – 999	-	0,5000	0,0047	0,9940
1.000 – 1.999	-2,60	0,4953	0,0320	6,40
2.000 – 2.999	-1,9	0,4633	0,1293	25,86
3.000 – 3.999	-0,97	0,3340	0,2704	54,08
4.000 – 4.999	-0,16	0,0636	0,3058	61,16
	0,65	0,2422		
5.000 – 5.999	1,46	0,4279	0,1857	37,14
6.000 – 6.999	2,27	0,4884	0,0605	12,10
7.000 – 7.900	3,08	0,4990	0,0106	2,12
$\geq$ 8.000	-	0,5000	0,0010	0,20

(2) langkah-2 : menghitung nilai χ^2 :

Untuk menghitung nilai χ^2, perlu dilakukan penyesuaian terlebih dahulu terhadap tabel frekuensi ini, yaitu setiap sel *normal frequency* harus > 5, sehingga tabel frekuensi yang telah disesuaikan dengan ketentuan ini menjadi:

Tabel 7.14 Perhitungan Nilai χ^2

Range	f_o	f_e	$(f_o - f_e)^2$	$(f_o - f_e)^2/f_e$
0 – 1.999	7	7,34	0,12	0,0163
2.000 – 2.999	22	25,86	14,90	0,5762

3.000 – 3.999	50	54,08	16,65	0,3079
4.000 – 4.999	62	61,16	0,71	0,0116
5.000 – 5.999	40	37,14	8,18	0,2202
$\geq$ 6.000	19	14,42	20,98	1,4549
			Total	2,5871

(3) langkah-3 : membandingkan nilai χ^2 dengan χ^2 pada tabel dengan df = k-p-1 (k = banyaknya kelas frekuensi, = 6; p = banyaknya parameter yang digunakan, = 2) = 6 – 2 – 1 = 3. Pada α = 0,05, nilai χ^2 tabel = 7,815. Nilai hitung χ^2 < nilai χ^2 tabel, maka H_0 diterima, artinya distribusi data frekuensi penjualan pada setiap *range* penjualan berdistribusi normal.

Tabel 7.15 Perhitungan Nilai χ^2

Range	f_o	f_e	$(f_o – f_e)^2$	$(f_o – f_e)^2/f_e$
0 – 1.999	7	7,34	0,12	0,0163
2.000 – 2.999	22	25,86	14,90	0,5762
3.000 – 3.999	50	54,08	16,65	0,3079
4.000 – 4.999	62	61,16	0,71	0,0116
5.000 – 5.999	40	37,14	8,18	0,2202
$\geq$ 6.000	19	14,42	20,98	1,4549
			Total	2,5871

(3) langkah-3 : membandingkan nilai χ^2 dengan χ^2 pada tabel dengan df = k-p-1 (k =banyaknya kelas frekuensi, = 6; p = banyaknya parameter yang digunakan, = 2) = 6 –2 – 1 = 3. Pada α= 0,05, nilai χ^2 tabel = 7,815. Nilai hitung χ^2< nilai χ^2

tabel, makaH_0 diterima, artinya distribusi data frekuensi penjualan pada setiap *range* penjualan berdistribusi normal.

8. *Analisys of Variance* (ANOVA)

Ini merupakan perbandingan banyak rata-rata (lebih dari dua rata-rata) secara serentak dan mencoba merunut sumber variasi terhadap faktor eksplanatori. Mempersiapkan rancangan percobaan (*experimental design*) bisa lebih efisien untuk penggunaan data terbatas. ANOVA dikembangkan oleh Ronald A. Fisher (1890 – 1962), dalam penelitian agrikultur (khususnya faktor-faktor yang mempengaruhi pertumbuhan tanaman), selanjutnya juga digunakan di bidang psikologi, enginering, farmasi, pemasaran dan bidang-bidang lainnya. Tujuan ANOVA adalah mencari sumber variasi pada sebuah variabel dependen numerik, Y. Variasi dalam variabel tersebut terhadap sekitar rata-ratanya, itu dijelaskan oleh sebuah atau lebih variabel independen dan sebagian tidak dijelaskan (merupakan *error* yang bersifat random).

Variasi Y = Variasi yang dapat menjelaskan + Variasi yang tidak dapat menjelaskan
(sekitar (oleh faktor) (merupakan *error*
random) rata-rata)

ANOVA merupakan perbandingan rata-rata. Setiap keungkinan nilai faktor atau kombinasi faktor merupakan perlakuan (*treatment*). Ini merupakan uji apakah sebuah faktor memiliki efek terhadap Y, bahkan menguji bagaimana efek antar faktor-faktor.

Asumsi pada ANOVA :

a. Data Y harus independen.

b. Populasi harus berdistribusi normal.

c. Populasi harus memiliki varians yang sama.

Ada tiga bentuk data untuk ANOVA, yaitu : (1) *One Factor ANOVA* (*completely; randomized model*); (2) *Two Factor*

ANOVA tanpa replikasi (*randomized block model*); dan (3) *Two Factor ANOVA* dengan replikasi (*full factorial model*).

Contoh-7.18 : *One Factor ANOVA*

Perusahaan kosmetik memiliki empat bagian pembungkusan dengan karton untuk kemudian dikirim ke agen-agen pengecer. Pada setiap bagian, ada dua pekerja yang bertugas memasukkan produk ke bungkus karton dan memberikan label. Setiap bagian biasanya bisa membungkus sampai dengan 200 karton/hari, dan seringkali lebih. Data pembungkusan dengan karton dalam seminggu dapat dilihat sebagai berikut:

Tabel 7.16 Data Pembungkusan Dengan Karton.

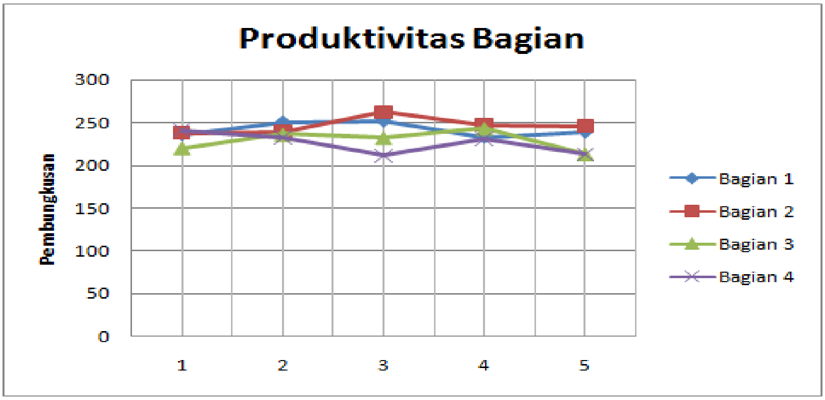
Hari	Bagian 1	Bagian 2	Bagian 3	Bagian 4
1	236	238	220	241
2	250	239	236	233
3	252	262	232	212
4	233	247	243	231
5	239	246	213	213
Jumlah	1210	1232	1144	1130
Rata-rata	242	246,4	228,8	226
SD	8,515	9,607	12,153	12,884
N	5	5	5	5

Langkah pertama adalah membuat plot data untuk melihat pola data dari hari ke hari; hasilnya sebagai berikut:

Tabel 7.17 Data Pembungkusan Dengan Karton.

Hari	Bagian 1	Bagian 2	Bagian 3	Bagian 4
1	236	238	220	241
2	250	239	236	233
3	252	262	232	212
4	233	247	243	231
5	239	246	213	213
Jumlah	1210	1232	1144	1130
Rata-rata	242	246,4	228,8	226
SD	8,515	9,607	12,153	12,884
N	5	5	5	5

Langkah pertama adalah membuat plot data untuk melihat pola data dari hari ke hari; hasilnya sebagai berikut:



Gambar 7.8. Plot Produktivitas Bagian.

Selanjutnya nyatakan hipotesis :

H_0 : $\mu_1 = \mu_2 = \mu_3 = \mu_4$□rata-rata produktivitas ke empat bagian, sama.

H_a : tidak semua rata-rata sama, paling tidak ada sebuah rata-rata yang berbeda.

Tabel *One Factor ANOVA* dapat dijelaskan sebagai berikut:

Sumber Variasi	Jumlah Kuadrat	Derajat Bebas	Kuadrat Rata-2	Statistik-F
Perlakuan (antar kelompok)	$SSA = \sum_{j=1}^{c} n_j(\bar{y}_j - \bar{y})^2$	c - 1	$MSA = \dfrac{SSA}{c-1}$	$F = \dfrac{MSA}{MSE}$
Error (dalam kelompok)	$SSE = \sum_{j=1}^{c} \sum_{i=1}^{nj} (y_{ij} - \bar{y}_j)^2$	N - c	$MSE = \dfrac{SSE}{n-c}$	
Total	$SST = \sum_{j=1}^{c} \sum_{i=1}^{nj} (y_{ij} - \bar{y})^2$	N - 1		

Di mana, c = kelompok, N = banyaknya data, dalam hal ini = 5 x 4 = 20 Lakukan perhitungan nilai-F:

Sumber Variasi	Jumlah Kuadrat	Derajat Bebas	Kuadrat Rata-2	Statistik-F
Perlakuan (antar kelompok)	1.479,2	3	493,067	F = 4,12
Error (dalam kelompok)	1.914,0	16	119,625	
Total	3.393,20	19		

Selanjutnya, bisa dilihat pada Tabel F; bahwa pada 〈 = 0,05, numerator = derajat bebas c - 1 = 4 – 1 = 3, dan denominator = N – c = 20 – 4 = 16; nilai F_{kritis} = 3,24. Karena F hasil perhitungan > F_{kritis}, maka H_a diterima, artinya rata-rata produktivitas antar bagian memang berbeda. Atau dengan membandingkan probabilitas nilai F hasil perhitungan dengan aplikasi Excel sebagai berikut: '=FDIST(4,12;3;16)'; hasilnya adalah = 0,0241. Angka ini lebih kecil daripada 〈 (= 0,05); maka H_a diterima.

<u>Contoh-7.19</u> : *Two Factor ANOVA without Replication*

Setiap supir mengharapkan mengendarai mobil dengan akselerasi yang baik. Sebuah uji terhadap derajat rotasi pedal akselerator (faktor A) pada berbagai kecepatan awal (faktor B), hasil uji coba itu adalah:

Tabel 7.18 Percepatan Maksimum Laju Kendaraan

Rotasi Pedal (A_j)	Kecepatan Awal (B_k)			
	10 mph	20 mph	40 mph	55 mph
5°	0,35	0,19	0,14	0,10
8°	0,37	0,28	0,19	0,19
10°	0,42	0,30	0,29	0,23

Dalam teori fisika, percepatan merupakan fungsi dua faktor yaitu derajat rotasi pedal akselerator (A) dan kecepatan awal (B) :

Akselerasi = f{rotasi pedal, kecepatan awal}, atau dapat dituliskan sebagai sebuah fungsi linier: $Y_{jk} = \mu + A_j + B_k + e_{jk}$. Hipotesis uji adalah :

Faktor Rotasi Pedal:
$H_0 : A_1 = A_2 = A_3 = A_4 = 0$
H_a : tidak semua $A_j = 0$

Faktor Kecepatan Awal:
$H_0 : B_1 = B_2 = B_3 = B_4 = 0$
H_a : tidak semua $B_k = 0$

Tabel *Two Factor ANOVA without replication* dapat dijelaskan sebagai berikut:

Sumber Variasi	Jumlah Kuadrat	Derajat Bebas	Kuadrat Rata-2	Statistik-F
Faktor A (efek baris)	$SSA = c\sum_{j=1}^{r} (\bar{y}_j - \bar{y})^2$	r - 1	$MSA = \frac{SSA}{r-1}$	$F_A = \frac{MSA}{MSE}$
Faktor B (efek kolom)	$SSB = r\sum_{k=1}^{c} (\bar{y}_k - \bar{y})^2$	c – 1	$MSB = \frac{SSB}{c-1}$	$F_B = \frac{MSB}{MSE}$
Error	$SSE = \sum_{j=1}^{r}\sum_{k=1}^{c} (y_{jk} - \bar{y}_j - \bar{y}_k + \bar{y}_k)^2$	(r – 1)(c – 1)	$MSE = \frac{SSE}{(r-1)(c-1)}$	
Total	$SST = \sum_{j=1}^{c}\sum_{k=1}^{r} (y_{jk} - \bar{y})^2$	rc – 1		

Di mana r = banyaknya baris, c = banyaknya kolom. Numerator untuk F_A = (r – 1) = 3 – 1 = 2, dan denominator = (r – 1)(c – 1) = (2)(3) = 6; numerator F_B = (c – 1) = 4 – 1 = 3, dan denominator = (r – 1)(c – 1) = (2)(3) = 6.

Sumber Variasi	Jumlah Kuadrat	Derajat Bebas	Kuadrat Rata-rata	Statistik F
Fakor A (efek baris)	0,026517	2	0,013258	22,83732
Faktor B (efek kolom)	0,073892	3	0,024631	42,42584
Error	0,003483	6	0,000581	
Total	0,103892	11		

Probabilitas (F_A = 22,83732) dihitung dengan aplikasi Excel, "=FDIST(22,83732;2;6)" = 0,001565 < 0,05; maka H_0 ditolak, artinya ada efek faktor A (derajat rotasi pedal ekselerasi) terhadap akselerasi laju mobil. Probabilitas (F_B = 42,42584, juga dihitung dengan aplikasi Excel "=FDIST(42,42584;3;6)" = 0,000196 < 0,05; maka H_0 juga ditolak, artinya ada efek faktor B (kecepatan awal) terhadap akselerasi laju mobil. Perbandingan menunjukkan bahwa efek kecepatan awal mobil lebih kuat daripada efek rotasi pedal terhadap akselerasi laju mobil. Percepatan mobil pada jalan tol bebas hambatan lebih tinggi daripada mobil yang baru bergerak.

Contoh-7.20 : *Two Factor ANOVA with Replication*

Sebuah institusi kesehatan memesan peralatan kesehatan untuk dikirim kepada empat klinik yang dimiliki. Ada lima suplier yang biasa memenuhi permintaan tersebut. Data waktu pengiriman barang dari suplier sampai tiba di klinik pada kuartal yang lalu adalah:

Tabel 7.19 Waktu Pengiriman Barang (dalam satuan hari)

	Suplier 1	Suplier 2	Suplier 3	Suplier 4	Suplier 5
Klinik A	8 – 8 - 10 – 13	14 – 9 -14 – 11	10 – 15 – 10 – 7	8 – 7 – 13 – 10	17 – 12 – 9 – 10
Klinik B	13 – 14 – 12 -13	9 – 9 – 7 – 8	12 – 10 – 10 – 11	6 – 10 – 12 – 8	15 – 12 – 12 – 10
Klinik C	11 – 10 – 12 – 14	8 – 9 – 11 - 12	12 – 10 – 13 – 10	10 – 11 – 7 – 10	14 – 13 – 10 – 12
Klinik D	7 – 10 – 10 – 13	8 – 13 - 9 - 12	7 – 5 – 6 - 5	8 – 5 – 11 - 4	14 – 13 – 8 - 11

Waktu Pengiriman = f{Klinik; Suplier; Interaksi Klinik – Suplier} , atau dapat ditulis sebagai sebuah fungsi linier: $Y_{jk} = \mu + A_j + B_k + AB_{jk} + e_{ijk}$. Hipotesis uji adalah:

Faktor Klinik:

H_0 : $A_1 = A_2 = A_3 = A_4 = 0$ (rata-rata waktu penerimaan peralatan kesehatan antar klinik, sama).

H_a : tidak seluruh Aj = 0 (rata-rata waktu penerimaan peralatan kesehatan antar klinik, berbeda).

Faktor Suplier:

H_0 : $B_1 = B_2 = B_3 = B_4 = B_5 = 0$ (rata-rata waktu pengiriman peralatan kesehatan antar suplier, sama)

H_a : tidak seluruh $B_k = 0$ (rata-rata waktu pengiriman peralatan kesehatan antar suplier, berbeda).

Faktor Ineraksi :

H_0 : seluruh $AB_{jk} = 0$ (tidak ada efek interaksi)

H_a : tidak seluruh $AB_{jk} = 0$ (ada efek interaksi).

Tabel *Two Factor ANOVA with replication* dapat dijelaskan sebagai berikut :

Sumber Variasi	Jumlah Kuadrat	Derajat Bebas	Kuadrat Rata-2	Statistik-F
Faktor A (efek baris)	$SSA = cm\sum_{j=1}^{r} (\overline{y}_j - \overline{y})^2$	r - 1	$MSA = \frac{SSA}{r-1}$	$F_A = \frac{MSA}{MSE}$
Faktor B (efek kolom)	$SSB = rm\sum_{k=1}^{c} (\overline{y}_k - \overline{y})^2$	c – 1	$MSB = \frac{SSB}{c-1}$	$F_B = \frac{MSB}{MSE}$
Interaksi (A x B)	$SSI = m\sum_{j=1}^{r}\sum_{k=1}^{c} (\overline{y}_{jk} - \overline{y}_j - \overline{y}_k + y)^2$	(r – 1)(c – 1)	$MSI = \frac{SSI}{(r-1)(c-1)}$	$FI = \frac{MSI}{MSE}$
Error	$SSE = \sum_{i=1}^{m}\sum_{j=1}^{r}\sum_{k=1}^{c} (y_{ijk} - \overline{y}_{jk})^2$	rc(m – 1)	$MSE = \frac{SSE}{(r-1)(c-1)}$	
Total	$SST = \sum_{i=1}^{m}\sum_{j=1}^{r}\sum_{k=1}^{c} (y_{ijk} - \overline{y})^2$	rcm – 1		

Dari data pada Tabel 7.19 dapat diisikan pada Tabel ANOVA berikut:

Sumber Variasi	Jumlah Kuadrat	Derajat Bebas	Kuadrat Rata-rata	Statistik F
Fakor A (efek baris)	51,35	3	17,117	3,43
Faktor B (efek kolom)	104,43	4	26,106	5,24
Interkasi (A x B)	102,78	12	8,565	1,72
Error	299,0	60	4,983	
Total	557,56	79		

Di mana r = banyaknya baris, c = banyaknya kolom, dan m = banyaknya replikasi. Numerator untuk $F_A = (r – 1) = 4 – 1 =$

3, dan denominator = (r)(c)(m-1) = (4)(5)(3) = 60; numerator F_B = (c – 1) = 5 – 1 = 4, dan denominator = (r)(c)(m-1) = (4)(5)(3) = 60. Numerator untuk faktor interaksi (F_I) = (r – 1)(c – 1) = 3 x 4 = 12, dan denominator = (r)(c)(m-1) = (4)(5)(3) = 60.

Probabilitas (F_A = 3,43) dihitung dengan aplikasi Excel: '=FDIST(3,43;3;60)' = 0,0226 < 0,05; maka H_0 ditolak, artinya rata-rata waktu penerimaan antar klinik, berbeda. Probabilitas (F_B = 5,24) dihitung dengan '=FDIST(5,24;4;60)' = 0,011 < 0,05; H_0 ditolak, artinya rata-rata waktu pengiriman antar suplier, berbeda. Probabilitas (F_I = 1,72) dihitung dengan: '=FDIST(1,72;12;60) = 0,0849 > 0,05; H_0 diterima, artinya tidak terbukti ada perbedaan interkasi waktu pengiriman antar suplier dan waktu penerimaan antar klinik.

7.3 Uji Hipotesis dengan Metode Non Parametrik

1. Pendahuluan

Sebagian besar metode yang digunakan dalam statistik inferensial didasari dengan asumsi bahwa data yang diobservasi berasal dari populasi dengan distribusi normal. Jika hal ini benar, maka metode-metode tersebut menyimpulkan seluruh informasi yang ada dari sejumlah sampel dan biasanya kualitas informasi tersebut memiliki kemungkinan ketelitian yang terbaik.

Namun karena banyak situasi yang tidak dapat memenuhi asumsi di atas, para ahli statistik kemudian mengembangkan suatu metode lain yang tidak membutuhkan asumsi semacam itu. Metode-metode ini kemudian dikenal sebagai **metode non parametrik**, yang pada awalnya dikemukakan Wolfowitz pada tahun 1942. Metode ini dapat digunakan pada kondisi yang lebih umum, dan seringkali lebih mudah untuk dijelaskan dan dimengerti; karena metode-metode tersebut diawali dari pemikiran "cepat dan mudah" atau "teknik jalan pintas".

Metode nonparametrik merupakan metode statistik yang mengabaikan asumsi-asumsi pada metode statistik parametrik, khususnya yang berkaitan dengan asumsi distribusi normal; sehingga ada beberapa istilah yang dikaitkan dengan metode ini, yaitu statistik bebas sebaran distribusi (*distribution-free statistics*) dan uji bebas asumsi (*assumption-free test)*. Statistik nonparametrik digunakan untuk menganalisis data yang berskala nominal atau ordinal karena pada umumnya data berjenis nominal dan ordinal tidak menyebar normal. Dari segi jumlaH data, pada umumnya statistik non parametrik digunakan untuk data dalam jumlah kecil ($n < 30$).

Beberapa metode non parametrik yang umum dan populer dalam statistik inferensial adalah :

a. Uji chi-square (χ^2).
b. The Wilcoxon Signed Rank Test,
c. The Wilcoxon Rank-Sum Test (Mann-Whitney Test),
d. The Kruskal-Wallis Test,
e. The Sign Test,
f. The Sprearman’s Rank Correlation,
g. The Test of Randomness,
h. The Kolmogorov-Smirnov Test,
i. The Kendall Test of Concordance.

2. Keunggulan dan Kelemahan Metode Statistik Non parametrik

Berikut ini dapat disebutkan beberapa keunggulan metode non parametrik jika dibandingkan dengan metode parametrik :

(1) Asumsi dalam metode non parametrik relatif lebih longgar dibanding dengan metode parametrik.
(2) Stokastik perhitungan lebih mudah dan lebih cepat.
(3) Untuk pemahaman konsep dan metodenya tidak membutuhkan pemahaman ilmu matematika atau ilmu statistik yang mendalam.

(4) Metode non parametrik dapat diterapkan untuk data dengan pengukuran yang kurang terstandard (misal : nominal atau ordinal).
(5) Efisiensi statistik non parametrik lebih tinggi dibandingkan dengan metode parametrik untuk jumlah sampel yang kecil.

Beberapa kelemahan metode non parametrik adalah :
(1) Jika asumsi pada uji statistik parametrik dapat terpenuhi, penggunaan metode ini menyebabkan pemborosan informasi.
(2) Untuk ukuran sampel yang besar, tingkat efisiensi metode ini relatif lebih rendah dibandingkan dengan metode parametrik.

7.4. Contoh-contoh Aplikasi Metode Statistik Non Parametrik

1. Uji Chi Square (χ^2)

Uji χ^2 merupakan uji independensi yang hanya digunakan untuk data diskrit. Yang dimaksud dengan independensi adalah bahwa sebuah variabel tidak dipengaruhi atau tidak memiliki hubungan dengan variabel lainnya. Uji ini digunakan untuk mengestimasikemungkinan adanya variabel atau faktor lain (selain *sampling error*) yang menyebabkan adanya hubungan atau pengaruh. Jika H_0 menyatakan bahwa tidak ada hubungan (independen), maka tujuan analisisnya adalah untuk menguji apakah hubungan pada data itudisebabkan oleh *sampling error*.

Hipotesis yang diuji adalah :

H_0 : kedua variabel, independen.

H_a : kedua variabel, dependen.

Untuk menghitung χ^2digunakan formula sebagai berikut:

$$\boxed{\chi^2 = \Sigma \frac{(f_e - f_o)^2}{f_e}} \qquad \text{. . . . . (7.11)}$$

di mana,

f_o = frekuensi riil hasil observasi

f_e = ekspektasi frekuensi.

Untuk menghitung f_e rumus yang digunakan adalah :

$$f_e = \frac{(\Sigma c_i)(\ \Sigma r_j)}{N} \qquad \text{. . . . . (7.12)}$$

di mana,

Σc_i = jumlah kolom

Σr_j = jumlah baris

N = banyaknya sampel yang diamati

Kriteria pengujian adalah :

H_0 diterima : $\chi^2 < \chi^2_{df,\langle}$

H_0 ditolak : $\chi 2 > \chi^2_{\ df,\langle}$

<u>Contoh-7.21</u> : Uji Independensi Dengan $|^2$

Dari 150 orang mahasiswi pada sebuah fakultas terdiri atas tiga ras, yaitu : Asia, Kaukasia dan Melanesia. Penelitian bertujuan membuktikan apakah variabel intensitas kunjungan ke salon perawatan kulit selama periode tertentu berhubungan dengan ras. H_a yang diuji adalah bahwa intensitas kunjungan ke salon perawatan kulit tidak independen terhadap ras. Data yang diperoleh dari survey adalah:

Tabel 7.20 Data Kunjungan Ke Salon Perawatan Kulit (f_o)

Ras	Frekuensi Kunjungan			Total Baris
Asia	4	32	29	65
Kaukasia	9	29	20	58
Melanesia	5	16	6	27
Total Kolom	18	77	55	150

Syafiadi Rizki Abdila, S.T., Ph.D, Prof. Dr. Ir. Drs. Syafwandi, M.Sc
Dr. Yulasmi, S.E., M.M, Dr. Robby Dharma, S.E., M.M
Dr. Syafri Arlis, S.Kom, M.Kom

Dari Tabel 7.20 di atas, dapat dihitung ekspektasi frekuensi (f_e) dengan menggunakan formula (7.12), pada setiap sel seperti terlihat pada Tabel 7.21 :

Tabel 7.21 Ekspektasi Frekuensi (f_e).

Ras	Frekuensi Kunjungan			Total Baris
Asia	8	33	24	65
Kaukasia	7	30	21	58
Melanesia	3	14	10	27
Total Kolom	18	77	55	150

Dari Tabel 7.20 dan Tabel 7.21, dapat dihitung nilai χ^2 dengan menggunakan formula (7.11) sebagai berikut:

$$\chi^2 = \frac{(4-8)^2}{8} + \frac{(9-7)^2}{7} + \frac{(5-3)^2}{3} + \frac{(32-33)^2}{33} + \frac{(29-30)^2}{30} + \frac{(16-14)^2}{14} + \frac{(29-24)^2}{24} + \frac{(20-21)^2}{21} + \frac{(6-10)^2}{10}$$

$$= 6{,}54$$

Derajat bebas (*degree of freedom*, df) = (r -1)(c – 1) = 2 x 2 = 4

Pada Tabel χ^2, untuk α = 0,05 dan df = 4, nilai χ^2 = 9,49.

Karena χ^2 hasil perhitungan < χ^2 pada tabel, maka H_0 diterima, artinya variabel ras independen terhadap variabel intensitas kunjungan ke salon perawatan kulit.

Selanjutnya, jika tujuan penelitian berubah menjadi : apakah ada hubungan antara ras Asia saja dengan intensitas kunjungan ke salon perawatan kulit, maka kedua ras lain (Kaukasia dan Melanesia) digabungkan datanya menjadi ras non Asia. Tabel 7.20 berubah menjadi:

Tabel 7.22 Data Kunjungan Ke Salon Perawatan Kulit (f_o).

Ras	Frekuensi Kunjungan			Total Baris
Asia	4	32	29	65
Non Asia	14	45	26	85
Total Kolom	18	77	55	150

Aplikasi formula (7.12) menghasilkan tabel berikut :

Tabel 7.23 Ekspektasi Frekuensi (f_e).

Ras	Frekuensi Kunjungan			Total Baris
Asia	8	33	24	65
Non Asia	10	44	31	85
Total Kolom	18	77	55	150

Dari Tabel 7.22 dan Tabel 7.23, dapat dihitung nilai χ^2 dengan menggunakan formula (7.11) sebagai berikut :

$$\chi^2 = \frac{(4-8)^2}{8} + \frac{(14-10)^2}{10} + \frac{(32-33)^2}{33} + \frac{(45-44)^2}{44} + \frac{(29-24)^2}{24} + \frac{(26-31)^2}{31}$$

$$= 5{,}34$$

Derajat bebas (*degree of freedom*, df) = (r -1)(c – 1) = 1 x 2 = 2

Pada Tabel χ^2, untuk α= 0,05 dan df = 4, nilai χ^2 = 5,99.

Karena χ^2 hasil perhitungan < χ^2 pada tabel, maka H_0 diterima, artinya variabel ras Asia independen terhadap variabel intensitas kunjungan ke salon perawatan kulit.

Untuk kasus matriks uji χ^2 berdimensi 2 x 2 (df = 1), χ^2 dihitung dengan formula yang lebih singkat, yaitu:

$$\chi^2 = \frac{n[(a)(d) - (b)(c)]^2}{(a + b)(c + d)(a + c)(b + d)} \quad \text{. . . . . (7.13)}$$

<u>Contoh-7.22</u> : Uji Independensi Dengan χ^2

Sebuah penelitian bertujuan untuk mengetahui apa ada keterkaitan antara program Diklat karyawan dengan kinerjanya. Untuk itu dipiih sampel karyawan secara random sebanyak 150 orang (n = 150). Data yang dikumpulkan adalah sebagai berikut:

Tabel 7.24 Data Peserta Program Diklat dan Perubahan Kinerja.

	Sebelum	Sesudah	Total Baris
Pernah ikut Program DIKLAT	65 (a)	24 (b)	89
Tidak pernah ikut Program DIKLAT	37 (c)	27 (d)	61
Total Kolom	99	51	150

Perhitungan χ^2 :

$$\chi^2 = \frac{150\,[(65 \text{ x } 27) - (24 \text{ x } 37)]^2}{(65 + 24)(37 + 27)(65 + 37)(24 + 27)}$$

$= 4{,}82$

Pada tabel, nilai χ^2 pada $\alpha = 0{,}05$ dan df = 1 sama dengan = 3,841. Karena χ^2 hasil perhitungan > $\chi^2_{0,05;1}$, maka H_0 ditolak dan menerima H_a. Artinya karyawan yang pernah mengikuti program DIKLAT dependen terhadap kinerjanya.

Jika pada matriks analisis χ^2 dengan dimensi 2 x 2 ini ada sel dengan nilai < 10, formula perhitungan χ^2 (formula-(7.13) harus dikoreksi. Koreksi yang dikembangkan oleh Yates adalah sebagai berikut:

$$\chi^2 = \frac{n\,[(a)(d) - (b)(c) - n/2]^2}{(a + b)(c + d)(a + c)(b + d)} \qquad \text{.. . . . . (7.14)}$$

Jika data pada Tabel 7.24 di atas berubah seperti Tabel 7.25 di bawah ini:

Tabel 7.25 Data Peserta Program Diklat dan Perubahan Kinerja.

	Sebelum	Sesudah	Total Baris
Pernah ikut Program DIKLAT	70 (a)	8 (b)	78
Tidak pernah ikut Program DIKLAT	9 (c)	63 (d)	72
Total Kolom	79	71	150

Perhitungan χ^2:

$$\chi^2 = \frac{150\,[(70 \text{ x } 63) - (8 \text{ x } 9) - 150/2]^2}{(70 + 8)(9 + 63)(70 + 9)(8 + 63)}$$

$= 86{,}54$

Pada tabel, nilai χ^2 dengan $\alpha = 0{,}05$ dan $df = 1$ sama dengan $= 3{,}841$. Karena χ^2 hasil perhitungan $> \chi^2_{0,05;1}$, maka H_0 ditolak dan menerima H_a. Artinya program DIKLAT karyawan berkaitan dengan kinerjanya.

2. *The Wilcoxon Signed Rank Test.*

Median dari populasi kontinyu adalah suatu jumlah η di mana dari sejumlah X observasi yang dipilih secara random memiliki probabilitas sama untuk daerah yang lebih kecil atau lebih besar daripada η.

$P\{X<\eta\} = P\{X<\eta\} = 0{,}50$

Atau dengan kata lain, populasi simetri di sekitar median jika untuk setiap observasi a positif, X memiliki probabilitas yang sama antara daerah lebih kecil (η - a) dengan daerah lebih besar (η + a).

$P\{X<\eta\text{-}a\} = P\{X<\eta+a\}$

Metode Wilcoxon dapat diaplikasi untuk menguji : (a) median dari suatu populasi yang simetri dan (b) median dari perbedaan antara data berpasangan.

a. Wilcoxon Signed Rank Test Untuk Median.

H_0 : $\eta < \eta_0$ dan simetris.

H_a : $\eta > \eta_0$ dan simetris,

Jika populasi diasumsikan berdistribusi normal, maka uji-t dapat diaplikasi pada persoalan ini. Untuk populasi yang tidak berdistribusi normal, uji-t tidak tepat digunakan, maka metode Wilcoxon merupakan solusinya.

Prosedur dalam metode ini :

(1) Kurangkan median hipotesis (η_0) terhadap setiap data observasi.

$Y_t = X_t - \eta_0$

Untuk $Y_t = 0$ harus diabaikan (dikeluarkan) dari data.

(2) Buat urutan meningkat (*ascending sort*) Y_t dengan menggunakan nilai absolutnya.

(3) Tuliskan rank (R_t), dimulai 1, 2, …, n dengan tanda aljabar (+/-) yang sesuai dengan tanda aljabar Y_t.

(4) Hitung jumlah R_t (=W), atau:

$W = \Sigma R_t$.

(5) Kriteria uji: jika $W < \eta_0$, maka H_0 diterima, H_a ditolak dan sebaliknya.

<u>Contoh-7.23</u> : Wilcoxon Signed Rank Test dengan sampel kecil (n < 10).

ROR saham perusahaan tertentu selama empat tahun berturut tercatat (dalam %) = +8,40; -4,30; -0,80; dan +12,50. Hipotesis yang diuji adalah (di mana η_0 = +3,00%) :

$H_0 : \eta < 3,00$

$H_a : \eta \geq 3,00$

(1) Mengkurangkan data observasi dengan η_0 ($Y_t = X_t - \eta0$):

$Y_1 = +8,40 - 3,00 = +5,40$

$Y_2 = - 4,30 - 3,00 = -7,30$

$Y_3 = - 0,80 - 3,00 = -3,80$

$Y_4 = +12,50 - 3,00 = +9,50$

(2) Urutan peringkat berdasar nilai absolut Y_t :

-3,80 + 5,40 -7,30 + 9,50

(3) Urutan rank (R_t) :

- 1 + 2 - 3 + 4

(4) Jumlah rank (W) :

W = -1 + 2 – 3 + 4 = + 2

(5) Kriteria uji :

W < η_0, maka H_0 diterima. Artinya median R.O.R saham perusahaan tersebut < +3,00%.

Catatan : untuk n > 10, dalam menghitung W dapat dilakukan dengan penyesuaian melalui pendekatan distribusi normal. Rata-rata dan standard deviasi W dapat dihitung dengan:

$\int_W = 0$

$\sigma_W = \sqrt{\Sigma R_t^2}$

Variabel yang distandardisasi adalah W/σ_W. Untuk uji hipotesis pada α tertentu dengan satu sisi, berlaku aturan sebagai berikut :

$H_0 : \eta < \eta_0$ dan simetris,

$H_a : \eta > \eta_0$ dan simetris.

Kriteria uji untuk satu sisi :
- jika $W/\sigma_W > Z_\alpha$, maka H_0 ditolak dan H_a diterima dan sebaliknya, atau
- jika $W/\sigma_W < -Z_\alpha$, maka H_0 ditolak dan H_a diterima.

Kriteria uji untuk dua sisi :
- jika $W/\sigma_W > Z_{\alpha/2}$, maka H_0 ditolak dan H_a diterima dan sebaliknya, atau
- jika $W/\sigma_W < -Z_{\alpha/2}$, maka H_0 ditolak dan H_a diterima.

Untuk data yang sama, digunakan rata-rata rank. Contoh untuk data : 3, 5, 7, 7, dan 9, nilai rank yang seharusnya 1, 2, 3, 4 dan 5. Tetapi karena data no. 3 dan no. 4 sama, maka rank disesuaikan menjadi : 1, 2, 3,5, 3,5 dan 5. Angka 3,5 berasal dari (3 + 4)/2.

<u>Contoh-7.24</u> : Wilcoxon Signed Rank Test dengan sampel besar (n > 10).

Berikut ini data jawaban 12 orang responden tentang desain bungkus produk tertentu yang diperoleh dengan pembobotan. Bobot skornya : 1, 2, . . . , 7; untuk 1 = tidak suka dan 7 = sangat suka. Dua belas data skor tersebut adalah:

7	3	5	4	7	1	2	2
5	7	6	5				

Hipotesis yang diuji adalah median dari responden tentang desain bungkus produk ($\eta_0 = 4$) pada $\alpha = 0{,}05$.

$H_0 : \eta = 4$

$H_a : \eta \neq 4$

(1) Mengkurangkan data observasi dengan η_0 ($Y_i = X_i - \eta_0$) :

$Y_1 = 7 - 4 = 3$ $\quad$ $Y_7 = 2 - 4 = -2$

$Y_2 = 3 - 4 = -1$ $\quad$ $Y_8 = 2 - 4 = -2$

$Y_3 = 5 - 4 = 1$ $\quad$ $Y_9 = 5 - 4 = 1$

$Y_4 = 4 - 4 = 0$ $\quad$ $Y_{10} = 7 - 4 = 3$

$Y_5 = 7 - 4 = 3$ $\quad$ $Y_{11} = 6 - 4 = 2$

$Y_6 = 1 - 4 = -3$ $\quad$ $Y_{12} = 5 - 4 = 1$

Untuk Y_4 = 0, dikeluarkan dari data tersebut. Sehingga data menjadi:

3 -1 1 3 -3 -2 -2 1
3 2 1

(2) Urutan peringkat berdasar nilai absolut Y_t:

-1 1 1 1 -2 -2 2 3
3 -3 3

(3) Urutan rank (R_t):

-2,5 +2,5 +2,5 +2,5 -6 -6 6 +9,5
+9,5 -9,5 +9,5

(4) Jumlah rank (W):

W = -2,5 + 2,5 + 2,5 + 2,5 - 6 - 6 + 6 + 9,5 + 9,5 - 9,5 + 9,5 = 18

$\sigma_W = \sqrt{\Sigma R_t^2} = \sqrt{494} = 22{,}2$

$W_{standardized} = W/\sigma_W = 18/22{,}2 = 0{,}81$

(5) Kriteria uji:

$W_{standardized} < Z_{\alpha/2}$, maka H_0 diterima dan H_a ditolak. Artinya median jawaban responden = 4.

<u>Contoh-7.25</u> : *Wilcoxon Signed Rank Test* untuk Data Berpasangan.

Uji Wilcoxon dapat digunakan pula untuk menguji median selisih (perbedaan) data yang berpasangan, di mana perbedaan ditimbulkan oleh suatu *treatment* (perlakuan) tertentu. Data ke-satu merupakan hasil observasi dari perlakuan-1 dan data kedua merupakan hasil observasi dari perlakuan-2.

Hipotesis yang diuji adalah perlakuan-1 dan perlakuan-2 tidak menghasilkan efek

yang berbeda. Dengan kata lain, median selisih (X_t) = 0 dan simetris di sekitar angka = 0.

Sebuah perusahaan makanan ingin menguji apakah ada perbedaan efek terhadap jumlah kandungan bakteri pada susu yang dihasilkan dengan pemanasan sampai dengan temperatur tertentu. Hasil eksperimen pada berbagai tingkat pemanasan menunjukkan selisih jumlah bakteri sebagai berikut:

0,03	0,14	1,17	0,15	-0,02	-0,04	0,44	0,22
-0,01	0,13		0,19	0,70			

$H_0 : X_t = 0$

$H_a : X_t > 0$

(1) Mengkurangkan data observasi dengan η_0 ($Y_t = X_t - \eta_0$) :

Karena $\eta_0 = 0$, maka $Y_t = X_t$

$Y_1 = 0{,}03$	$Y_7 = 0{,}44$
$Y_2 = 0{,}14$	$Y_8 = 0{,}22$
$Y_3 = 1{,}17$	$Y_9 = -0{,}01$
$Y_4 = 0{,}15$	$Y_{10} = 0{,}13$
$Y_5 = -0{,}02$	$Y_{11} = 0{,}19$
$Y_6 = -0{,}04$	$Y_{12} = 0{,}70$

(2) Urutan peringkat berdasar nilai absolut Y_t :

-0,01	-0,02	0,03	-0,04	0,13	0,14	0,15	0,19
	0,22	0,44	0,70	1,17			

(3) Urutan rank (R_t) :

-1	-2	3	-4	5	6	7	8
9	10	11	12				

(4) Jumlah rank (W) :

$W = 64$

$\sigma_W = \sqrt{\Sigma R_t^2} = \sqrt{n(N+1)(2N+1)/6} = 25{,}5$

$W_{standardized} = W/\sigma_W = 64/25{,}5 = 2{,}51$

(5) Kriteria uji :

Pada $\alpha = 0{,}05$, $Z_{0,05} = 1{,}645$, sedangkan $W_{standardized} = 2{,}51$; maka H_0 ditolak dan Ha diterima. Artinya pemanasan sampai dengan temperatur tertentu menimbulkan efek penurunan jumlah kandungan bakteri dalam susu.

3. *The Wilcoxon Rank-Sum Test (Mann-Whitney Test).*

The Wilcoxon Rank-Sum Test merupakan uji non parametrik sebagai alternatif terhadap uji-t dua sampel (U test), dan uji alternatif terhadap *one way anova*, (H test). Dengan kata lain, H test ditujukan untuk menguji hipotesis nol yang menyatakan bahwa sejumlah-k sampel berasal dari populasi dengan rata-rata sama, melawan hipotesis alternatifnya yang menyatakan bahwa populasi memiliki rata-rata yang berbeda.

<u>Contoh-7.26</u> : *The Wilcoxon Rank-Sum Test* (Mann-Whitney Test atau U-test).

Data diameter dari dua jenis butiran plastik yang diamati adalah (dalam mm):

Platik-I : 0,63 0,17 0,35 0,49 0,18 0,43 0,12
0,20
0,47 1,36 0,51 0,45 0,84 0,32 0,40

Plastik-II : 1,13 0,54 0,96 0,26 0,39 0,88 0,92
0,53
1,01 0,48 0,89 1,07 1,11 0,58

Rata-rata kedua sampel adalah 0,46 mm dan 0,77 mm. Ujilah apakah kedua sampel tersebut memiliki rata-rata yang berbeda secara signifikan.

a. Urutkan keseluruhan data :

0,12 0,17 0,18 0,20 0,26 0,32 0,35 0,39
0,40 0,43
(I) (I) (I) (I) (II) (I) (I) (II)
(I) (I)
0,45 0,47 0,48 0,49 0,51 0,53 0,54 0,58
0,63 0,84
(I) (I) (II) (I) (I) (II) (II) (II)
(I) (I)
0,88 0,89 0,92 0,96 1,01 1,07 1,11 1,13
1,36

(II) (II) (II) (II) (II) (II) (II) (II)
(I)

b. Buat rank urutan dari hasil point-a tersebut:

1 2 3 4 29

c. Buatlah rank urutan dari masing-masing kelompok plastik:

Plastik-I : 1 2 3 4 6 7 9
10 11
12 14 15 19 20 29

Plastik-II : 5 8 13 16 17 18 21
22 23
24 25 26 27 28

Catatan: jika ada beberapa ada beberapa data yang sama, maka rank urutan data tersebut merupakan rata-ratanya, contoh jika rank urutan 3 dan 4 merupakan data yang sama, maka rank urutan keduanya = (3+4)/2 = 3,5.

d. Hipotesis

H_0: kedua sampel berasal dari populasi yang sama, dengan demikian memiliki rata-rata yang sama; $\mu_1 = \mu_2$.

Ha: kedua sampel berasal berasal dari dua populasi yang berbeda, $\mu_1 \neq \mu_2$.

e. $R_1 = \Sigma R_t$

$R_1 = 162$ dan $R_2 = 273$.

f. Hitung *U-statistic* dengan formula:

$$\boxed{U_1 = n_1 n_2 + \frac{(n_1)(n_1 + 1)}{2} - R_1} \qquad \text{. . . . . (7.15)}$$

atau

$$\boxed{U_2 = n_1 n_2 + \frac{(n_2)(n_2 + 1)}{2} - R_2} \qquad \text{. . . . . (7.16)}$$

Rata-rata dan varians *U-statistic* adalah:

$$\boxed{\int_{U1} = (n_1 .\ n_2)/2} \qquad \text{. . . . (7.17)}$$

$$\boxed{\sigma^2_{U1} = \frac{n_1 . n_2\ (n_1+n_2+1)}{12}} \qquad \text{. . . . . (7.18)}$$

g. Formula uji :

$$\boxed{z = \frac{U1 - \int_{U1}}{\sigma_{U1}}} \qquad \text{. . . . . (7.19)}$$

Kriteria uji :

- jika $z_{hitung} > z_{\alpha/2}$ (uji dua sisi), maka H_0 ditolak dan sebaliknya.

$U_1 = (15)(14) + (15)(16)/2 – 162 = 168$

$\int_{u1} = (15)(14)/2 = 105$

$\sigma^2_{U1} = (15)(14)(30)/12 = 525$

$z = (168 – 105)/\sqrt{525} = 2{,}74$

$z_{0,05/2} = 2{,}575$

$Z_{hitung} > z_{\alpha/2}$, maka H_0 ditolak dan H_a diterima, Artinya kedua jenis butir plastik tersebut berasal dari dua populasi yang berbeda.

<u>Contoh-7.27</u> : Uji Mann-Whitney Untuk Kepuasan Pelanggan Restoran

Sebuah penelitian ingin mengetahui apakah dengan pengeluaran yang lebih banyak pada sebuah restoran memberikan kepuasan yang lebih baik bagi konsumen. Untuk itu diamati pendapat konsumen yang sedang berada di restoran (sebanyak 29 restoran). Pada setiap restoran itu dipilih seorang konsumen. Data yang dikumpulkan (telah diurutkan menurut tingkat kepuasan konsumen) terlihat pada Tabel 7.24. Peringkat untuk data tingkat kepuasan yang sama, menggunakan rata-rata peringkat. Contoh, pengamatan pada konsumen no. 2 dan no.3, tingkat kepuasan keduanya = 76, peringkatnya 2, dan 3. Rata-rata peringkat = 5/2 = 2,5. Maka kedua pengamatan tersebut diberi peringkat masing-masing 2,5.

H_0 : tidak ada perbedaan tingkat kepuasan berdasar tingkat pembelanjaan

H_a : ada perbedaan tingkat kepuasan berdasar perbedaan tingkat pembelanjaan

Rumus yang digunakan adalah[12] :

$$Z = \frac{\overline{R_1} - \overline{R_2}}{(n_1 + n_2)\sqrt{\dfrac{n_1 + n_2 + 1}{(12)(n_1)(n_2)}}} \quad \ldots\ldots (7.20)$$

Di mana R_i adalah rata-rata rank kelompok-i.

Tabel 7.26 Pembelanjaan dan Kepuasan Konsumen di Restoran.

No.	Tingkat Pembelanjaan	Tingkat Kepuasan	Rank
1	L	73	1
2	L	76	2,5
3	H	76	2,5
4	L	77	4,5
5	L	77	4,5
6	L	78	7
7	H	78	7

[12] Frisch, 2004

8	H	78	7
9	L	79	10
10	L	79	10
11	L	79	10
12	L	80	13,5
13	L	80	13,5
14	L	80	13,5
15	L	80	13,5
16	L	81	16,5
17	L	81	16,5
18	H	82	18,5
19	H	82	18,5
20	L	83	20,5
21	L	83	20,5
22	H	84	22,5
23	H	84	22,5
24	H	85	25
25	H	85	25
26	H	85	25
27	H	86	28
28	H	86	28
29	H	86	28

Keterangan :

L = tingkat pembelanjaan rendah

H = tingkat pembelanjaan tinggi

Selanjutnya, dihitung rata-rata rank dari kedua jenis pembelanjaan, yaitu :

Tabel 7.27 Perhitungan Jumlah Rank Berdasar Pembelanjaan

No	Pem-belanjaan	Kepu-asan	Rank	No.	Pem-belanjaan	Kepu asan	Rank
1	L	73	1	1	H	76	2,5
2	L	76	2,5	2	H	78	7
3	L	77	4,5	3	H	78	7
4	L	77	4,5	4	H	80	13,5
5	L	78	7	5	H	82	18,5
6	L	79	10	6	H	82	18,5
7	L	79	10	7	H	84	22,5
8	L	79	10	8	H	84	22,5
9	L	80	13,5	9	H	85	25
10	L	80	13,5	10	H	85	25
11	L	80	13,5	11	H	85	25
12	L	81	16,5	12	H	86	28
13	L	81	16,5	13	H	86	28
14	L	83	20,5	14	H	86	28
15	L	83	20,5				

	n_1 =	15			n_2 =	14	
	© R_i =		164		© R_2 =		271
	Rata-rata Rank		10,93 3		Rata-rata Rank		19,35 7

$$Z = \frac{10{,}93333 - 19{,}35714}{(15+14)\sqrt{\dfrac{15+14+1}{(12)(15)(14)}}} = -2{,}662$$

Pada α = 0,01 (uji dua sisi), penerimaan H_0 adalah pada *range* -2,326 < z < + 2,326. Karena Z hasil perhitungan < - 2,326, maka H_0 ditolak dan menerima Ha. Artinya ada perbedaan tingkat kepuasan konsumen karena tingkat pembelanjaannya.

<u>Contoh-7.28</u> : Uji Kruskal Wallis (H-test) untuk Efektivitas Metode Preventif

Sebuah metode uji non parametrik yang lebih sederhana yang berkaitan dengan ANOVA adalah prosedur yang dikembangkan Kruskal-Wallis, dengan menggunakan ratio H, di mana *numerator* adalah nilai **SS bg(R)** dan *denominator* adalah porsi **SS bg(R)**, yaitu N(N + 1)/12.

$$\boxed{SS\ bg(R) = \Sigma \frac{R_i^2}{N_i} - \frac{TR}{N}} \qquad \text{. . . . . (7.21)}$$

Di mana SS bg(R) = jumlah penyimpangan kuadrat antar kelompok (*sum of squares deviates between group*), TR = total rank seluruh kelompok, N = banyaknya seluruh observasi.

Syafiadi Rizki Abdila, S.T., Ph.D, Prof. Dr. Ir. Drs. Syafwandi, M.Sc
Dr. Yulasmi, S.E., M.M, Dr. Robby Dharma, S.E., M.M
Dr. Syafri Arlis, S.Kom, M.Kom

$$H = \frac{SS\ bg(R)}{N(N+1)/12} \quad \ldots\ldots (7.22)$$

Sebuah penelitian melibatkan 21 orang penggemar anggur untuk membuat perbandingan kualitas rasa tiga jenis anggur. Kualitas anggur di beri skor 1 – 10 (1 = paling rendah dan 10 = paling tinggi). Ketiga jenis anggur itu sebenarnya sama, perbedaannya adalah pada jenis pertanyaan pada setiap kelompok. Pertanyaan untuk kelompok A diarahkan pada harapan kualitas yang tinggi, kelompok C diarahkan pada pertanyaan dengan harapan kualitas yang rendah, dan untuk kelompok B, adalah netral. Hasil *ranking* adalah :

Tabel 7.29 Hasil Skor Kualitas Anggur Oleh Setiap Kelompok

A	Skor	B	Skor	C	Skor
1	6,4	1	2,5	1	1,3
2	6,8	2	3,7	2	4,1
3	7,2	3	4,9	3	4,9
4	8,3	4	5,4	4	5,2
5	8,4	5	5,9	5	5,5
6	9,1	6	8,1	6	8,2
7	9,4	7	8,2		
8	9,7				
Rata-rata	8,2		5,5		4,9

Data skor tersebut dibuat peringkat, hasilnya adalah:

Tabel 7.30 Peringkat Kualitas Anggur Oleh Setiap Kelompok

A	Peringkat	B	Peringkat	C	Peringkat	Total
1	11	1	2	1	1	
2	12	2	3	2	4	
3	13	3	5,5	3	5,5	
4	17	4	8	4	7	
5	18	5	10	5	9	
6	19	6	14	6	15,5	
7	20	7	15,5			
8	21					
Jumlah	131		58		42	231
Rata-rata	16,38		8,29		7	
N	8		7		6	21

SS bg(R) = $(131)^2/8 + (58)^2/7 + (42)^2/6 - (231)^2/21 = 378{,}7$

H = $378{,}7/[(21 \times 22)/12] = 9{,}84$

Nilai χ^2 pada df = k – 1 = 2, dan $\alpha_{0,05}$ = 5,99; ini mengindikasikan bahwa H_a diterima, yang berati skor oleh ketiga kelompok responden berbeda signifikan.

Secara alternatif perhitungan H yang merupakan generalisasi dari U yang dapat diaplikasi untuk k-sampel. Prosedur uji sangat mirip dengan U-test, N = $\Sigma\, n_i$.

Formula uji untuk H alternatif :

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3\,(n+1) \qquad \text{. . . . . (7.23)}$$

***Catatan*:**

Untuk n > 5, untuk seluruh-i dan hipotesis nol benar, distribusi H-*statistic* ini didekati dengan distribusi χ^2 dengan derajat bebas = k – 1.

Sebuah percobaan dirancang untuk membandingkan tiga metode preventif terhadap karat. Hasil pengukuran kedalaman karat dari permukaan (dalam ukuran pits setiap seribu inci) adalah sebagai berikut:

Metode A :	77	54	67	74	71	66	
Metode B :	60	42	59	65	62	64	52
Metode C :	49	52	69	47	56		

Gunakan $\alpha = 0{,}05$ untuk menguji apakah ketiga sampel berasal dari populasi yang sama.

H_0 : ⌠$_1$ = ⌠$_2$ = ⌠$_3$

H_a : ⌠$_1$ ≠ ⌠$_2$ ≠ ⌠$_3$

$$H = \frac{12}{(18)\,(19)}\left(\frac{84^2}{6} + \frac{55{,}5^2}{7} + \frac{31{,}5^2}{5}\right) - (3)(19) = 6{,}7$$

Pada Tabel, nilai $\chi^2_{0,05,2} = 5{,}991$

Karena $\chi^2_{hitung} > \chi^2_{0,05,2}$; maka H_0 ditolak dan menerima H_a. Dengan demikian dapat dinyatakan bahwa ketiga metode preventif terhadap karat memiliki efektivitas yang berbeda.

<u>Contoh-7.29</u> : Uji Kruskal Wallis (H-test) untuk Pembelajaan di Toko Modern

Sebuah penelitian bertujuan untuk mengetahui apakah ada perbedaan jumlah pembelanjaan antar kelompok konsumen pada toko modern. Ada tiga kelompok konsumen yang diamati : (1) kelompok yang berbelanja karena merasa harga barangnya jelas tertulis pada label harga, (2) kelompok konsumen yang berbelanja karena merasa pilihan barangnya cukup banyak, dan (3) kelompok konsumen yang berbelanja karena merasa adanya kenyamanan waktu berbelanja. Hipotesis yang diuji:

H_0 : jumlah belanja ketiga kelompok konsumen, tidak berbeda

H_a : jumlah belanja ketiga kelompok konsumen, berbeda

Data yang terkumpul dari responden adalah:

Tabel 7.31 Data Belanja Tiga Kelompok Konsumen (dalam ribuan Rupiah).

	Res-1	Res-2	Res-3	Res-4	Res-5	Res-6	Res-7	Res-8
Kelompok 1	215	195	325	250	282	154		
Kelompok 2	200	225	235	207	190	175	205	225
Kelompok 3	150	180	210	220	157			

$n_1 = 6$, $n_2 = 8$, $n_3 = 5$, dan $N = n_1 + n_2 + n_3 = 6 + 8 + 5 = 19$

Buat urutan data dari terkecil sampai yang terbesar :

150	154	157	175	180	190	195
(1)	(2)	(3)	(4)	(5)	(6)	(7)
200	205	207	210	215	220	225
(8)	(9)	(10)	(11)	(12)	(13)	(14)
230	235	250	282	325		
(15)	(16)	(17)	(18)	(19)		

Angka dalam tanda kurung merupakan peringkat data, jika dijumlahkan berdasar kelompok konsumen:

$\Sigma R_1 = 75$ $\Sigma R_2 = 82$ $\Sigma R_3 = 33$

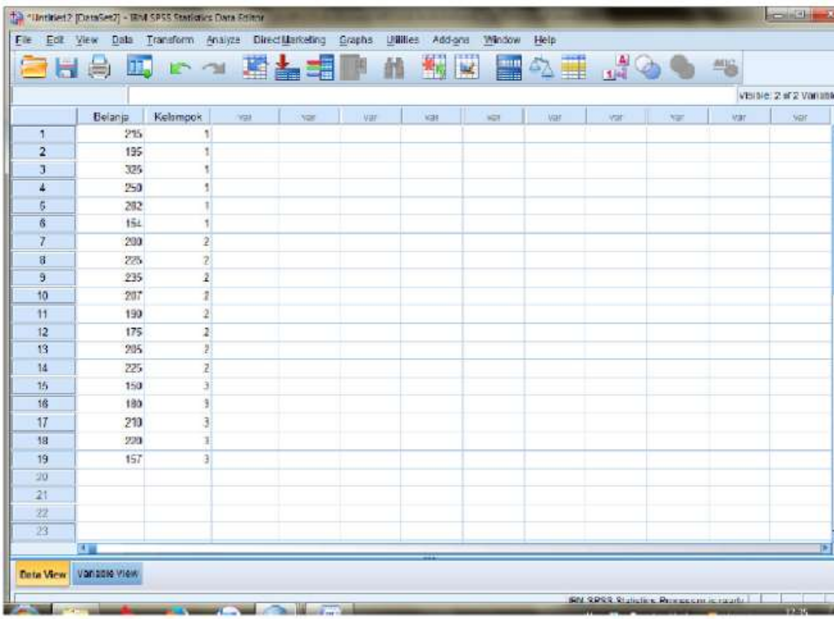
Aplikasikan formula (7.23) di atas:

$$H = \frac{12}{(19)\,(20)} \left(\frac{75^2}{6} + \frac{82^2}{8} + \frac{33^2}{5} \right) - (3)(20) = 3{,}028$$

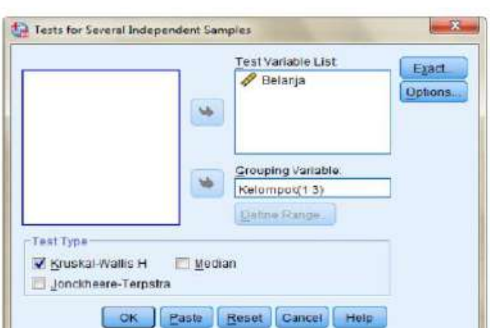
Nilai χ^2 pada α= 0,05 dan df = k-1 = 2, sama dengan = 5,991. Karena H < $\chi^2_{0,05;2}$; maka H_0 diterima. Artinya jumlah pembelanjaan ketiga kelompok konsumen itu tidak berbeda.

Aplikasi program SPSS untuk menyelesaikan kasus ini bisa juga dilakukan dengan langkah-langkah sebagai berikut :

(1) buat data input dengan SPSS:



(2) Aktifkan Uji Kruskal-Wallis, melalui '*Analyze – Non Parametric Test – Legacy Dialogs – K Indpendent Sample*'. Muncul tampilan sebagai berikut :



Klik Ok

(3) *Printout* SPSS adalah :

Ranks

	Kelompok	N	Mean Rank
Belanja	Kelompok-1	6	12,50
	Kelompok-2	8	10,25
	Kelompok-3	5	6,60
	Total	19	

Test Statistics[a,b]

	Belanja
Chi-Square	3,028
Df	2
Asymp. Sig.	,220

a. Kruskal Wallis Test
b. Grouping Variable: Kelompok

Nilai χ^2 = 3,028 dan Asymp. Sig. = 0,220 > α, maka H_0 diterima, artinya jumlah belanja dari ketiga kelompok itu tidak berbeda (kesimpulannya sama dengan hasil uji secara manual).

<u>Contoh-7.30</u> : Uji Kruskal Wallis (H-test) untuk Tingkat Produktivitas Buruh

Sebuah penelitian bertujuan mengetahui apakah perbedaan tingkat produktivitas buruh pada tiga divisi pabrik yang berbeda. Produktivitas buruh diukur per triwulan dan dinyatakan dalam persentase. Hasil pengumpulan datanya adalah:

Tabel 7.32 Data Produktivitias Buruh Per Divisi Pabrik (dalam %)

No. Buruh	Divisi 1	Divisi 2	Divisi 3
1	88,9	92,3	91,4
2	89,3	93,5	82,3
3	87,8	91,7	91,3

4	90,2	90,2	81,4
5	97,6	82,6	82,6
6	81,9	81,7	82,4
7	83,6	82,5	92,6
8	82,7	-	92,5
9	92,4	-	-
10	91,8	-	-

Hipotesis yang diuji :

H_0 : produktivitas buruh pada ketiga divisi, tidak berbeda

H_a : produktivitas buruh pada ketiga divisi, berbeda

$n_1 = 10$ $n_2 = 7$ $n_3 = 8$

Langkah pertama adalah mengurutkan data tanpa mempertimbangkan divisi-nya, hasilnya adalah :

Data	Rank	Divisi	Data	Rank	Divisi	Data	Rank	Divisi
81,4	(1)	3	87,8	(11)	1	92,4	(21)	1
81,7	(2)	2	88,9	(12)	1	92,5	(22)	3
81,9	(3)	1	89,3	(13)	1	92,6	(23)	3
82,3	(4)	3	90,2	(14,5)	1	93,5	(24)	2
82,4	(5)	3	90,2	(14,5)	2	97,6	(25)	1
82,5	(6)	2	91,3	(16)	3			
82,6	(7,5)	2	91,4	(17)	3			
82,6	(7,5)	3	91,7	(18)	2			
82,7	(9)	1	91,8	(19)	1			
83,6	(10)	1	92,3	(20)	2			

Untuk data yang sama, maka rank yang digunakan adalah rata-ratanya. Contoh, data 82,6 dengan rank = 7 dan 82,6 dengan rank 8; untuk keduanya rank diganti dengan rata-rata rank = (7 + 8)/2 = 7,5.

Jumlahkan nilai ranking pada masing-masing divisi, hasilnya adalah :

$R_1 = 137{,}5$ $\quad R_2 = 92$ $\quad R_3 = 95{,}5$

Aplikasikan formula (7.23) di atas :

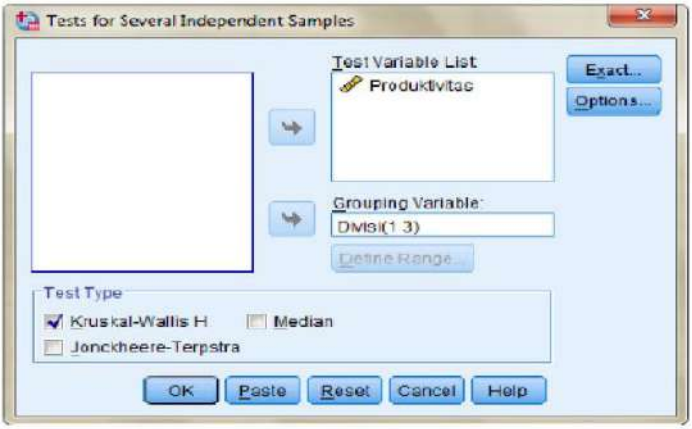
$$H = \frac{12}{(25)(26)}\left(\frac{137{,}5^2}{10} + \frac{92^2}{7} + \frac{95{,}5^2}{8}\right) - (3)(26) = 0{,}273$$

Nilai χ^2 pada $\alpha = 0{,}05$ dan df = k-1 = 2, sama dengan = 5,991. Karena $H < \chi^2_{0,05;2}$; maka H_0 diterima. Artinya produktivitas buruh ketiga kelompok divisi pabrik itu tidak berbeda.

Penggunaan SPSS untuk menyelesaikan kasus ini bisa juga dilakukan dengan langkah-langkah sebagai berikut :

(1) Buat data input dengan SPSS :

(2) Aktifkan Uji Kruskal-Wallis, melalui '*Analyze – Non Parametric Test – Legacy Dialogs – K Indpendent Sample*'. Muncul tampilan sebagai berikut :



(3) *Printout* SPSS adalah :

Ranks

	Divisi	N	Mean Rank
Produktivitas	Divisi-1	10	13,75
	Divisi-2	7	13,14
	Divisi-3	8	11,94
	Total	25	

Test Statistics[a,b]

	Produktivitas
Chi-Square	,273
df	2
Asymp. Sig.	,872

a. Kruskal Wallis Test
b. Grouping Variable: Divisi

Nilai χ^2 = 0,273 dan Asymp. Sig. = 0,872 > α, maka H_0 diterima, artinya produktivitas buruh dari ketiga divisi pabrik itu tidak berbeda (kesimpulannya sama dengan hasil uji cara manual).

<u>Contoh-7.31</u> : Uji Kruskal Wallis (H-test) untuk Tingkat Absensi Karyawan

Pada sebuah perusahaan ada tiga bagian penting : (1) bagian anggaran, (2) bagian gaji, dan (3) bagian akuntansi. Perusahaan ingin mengetahui apakah ada perbedaan tingkat absensi pada ketiga bagian itu. Dengan banyaknya karyawan yang berbeda, maka dipilih karyawan pada ketiga bagian itu dengan jumlah yang berbeda. Berikut hasil catatan tentang banyaknya hari mereka bekerja dalam tahun yang diamati :

Tabel 7.33 Banyaknya Hari Kerja Karyawan pada Tahun 2015.

Bagian	Hari Kerja (HK)									
Anggaran	278	260	265	245	258					
Gaji	205	270	220	240	255	217	266	239	240	228
Akuntansi	240	258	233	256	233	242	244	249		

H_0: tingkat absensi karyawan pada ketiga bagian itu tidak berbeda.

H_a: tingkat absensi karyawan pada ketiga bagian itu berbeda.

Peringkat data hari kerja tanpa melihat bagian dari karyawan adalah:

HK	Rank	Bagian	HK	Rank	Bagian	HK	Rank	Bagian
205	1	Gaji	242	11	Akuntansi	266	21	Gaji
17	2	Gaji	244	12	Akuntansi	270	22	Gaji
220	3	Gaji	245	13	Anggaran	278	23	Anggaran
228	4	Gaji	249	14	Akuntansi			
233	5,5	Akuntansi	255	15	Gaji			
233	5,5	Akuntansi	256	16	Akuntansi			
239	7	Gaji	258	17,5	Anggaran			
240	9	Gaji	258	17,5	Akuntansi			
240	9	Gaji	260	19	Anggaran			
240	9	Akuntansi	265	20	Anggaran			

Hitung jumlah *rank* pada setiap bagian, hasilnya adalah :

Tabel 7.34 Perhitungan Jumlah *Rank* Pada Setiap Bagian

Anggaran	Rank	Gaji	Rank	Akuntansi	Rank
278	23	205	1	240	9
260	19	270	22	258	17,5
265	20	220	3	233	5,5
245	13	240	9	256	16
258	17,5	255	15	233	5,5
		217	2	242	11
		266	21	244	12

		239	7	240	14
		240	9		
		228	4		
Jumlah	92,5		93		90,5
$n_1 = 5$		$n_2 = 10$		$n_3 = 8$	

Aplikasikan formula H (formula - 6.23):

$$H = \left(\frac{12}{(23)(24)} + \frac{92{,}5^2}{52} + \frac{93^2}{102} - \frac{90{,}5^2}{82} \right) (3)(24) = 6{,}259$$

Nilai χ^2 pada $\alpha = 0{,}05$ dan df = k-1 = 2, sama dengan = 5,991. Karena H > $\chi^2_{0,05;2}$; maka H_0 ditolak. Artinya tingkat absensi karyawan pada ketiga bagian itu berbeda.

4. Uji Random atau *Run Test*

Formula uji :

$$z = \frac{u - \int_U}{\sigma_u} \qquad \text{. . . . (7.24)}$$

Di mana *mean* ($\int_U$) dan *standard deviasi* (σ_U) adalah :

$$\int_U = \frac{2\,(n_1)(n_2)}{n_1 + n_2} + 1 \qquad \text{. . . . . (7.25)}$$

$$\sigma_U = \sqrt{\frac{2(n_1)(n_2)\,(2\,n_1,n_2 - n_1 - n_2)}{(n_1 + n_2)^2\,(n_1 + n_2 - 1)}} \qquad \text{. . . . . (7.26)}$$

<u>Contoh-7.32</u> : Uji Random untuk data non numerik.

Berikut ini data rusak (d, *defective*) dan tidak rusak (n, *non defective*) dari hasil produksi sebuah mesin tertentu:

n n n n n d d d d n n n n n n n n n n d d n n d d d d

1 2 3 4 5 6

Ujilah apakah data tersebut bersifat random pada $\alpha = 0{,}05$.

H_0 : data berdistribusi random

H_a : data tidak berdistribusi non random,

n_1 = banyaknya data *defective* = 10, dan

n_2 = banyaknya data *non defective* = 17.

Banyaknya rangkaian data berurutan yang sama = 6.

$\mu_U = [2\ (10)(17)]/27 + 1 = 13{,}59$

$$\sigma_U = \sqrt{\frac{2(10)(17)[(2)(10)(17) - 10 - 17}{(10+17)^2\ (10+17-1)}} = 2{,}37$$

$z_{hitung} = (6 - 13{,}59)/2{,}37 = -3{,}20$

$z_{0,05/2} = -2{,}575$ atau $+2{,}575$

Karena $z_{hitung} < -2{,}575$, maka H_a diterima dan H_0 ditolak. Artinya kerusakan hasil produksi tidak bersifat random. Maka diperkirakan kesalahan terjadi pada operator mesin bukan pada mesinnya.

***Catatan**:*

Run-test dapat juga diaplikasi untuk menguji random sampel dengan data numerik, melalui perhitungan banyaknya rangkaian di atas dan di bawah mediannya. Notasi data di atas median = a, sedang notasi data di bawah median = b, maka urutan a dan b tersebut dapat digunakan untuk menguji tingkat random data.

<u>Contoh-7.33</u> : *Run Test* untuk data numerik.

Seorang insinyur berkepentingan dengan kemungkinan terlalu banyaknya perubahan dalam setting mesin untuk menghasilkan produk standard. Berikut ini data diameter (dalam satuan inchi) dalam berbagai setting bagian mesin :

0,261	0,258	0,249	0,251	0,247	0,256	0,250	0,247	0,255	0,243
0,252	0,250	0,253	0,247	0,251	0,243	0,258	0,251	0,245	0,250
0,248	0,252	0,254	0,250	0,247	0,253	0,251	0,246	0,249	0,252
0,247	0,250	0,253	0,247	0,249	0,253	0,246	0,251	0,249	0,253

Gunakan α = 0,01 untuk menguji hipotesis yang menyatakan bahwa perubahan *setting* diameter bagian mesin itu bersifat random.

H_0 : data berdistribusi random.

H_a : data tidak berdistribusi random.

Median data = 0,250, maka data numerik di atas dapat dinyatakan dalam bentuk berikut:

a a b a b a b a b a a b a b a a b b a a b a a b b a b a b b a b a b a

a, adalah data dengan nilai di atas median, sedang b, adalah data dengan nilai di bawah median. Untuk data yang sama dengan median = 0,250, didrop.

n_1 = banyaknya data a = 19, n_2 = banyaknya data b = 16

U = banyaknya urutan data = 27

$\int_U = [2(19)(16)]/35 + 1 = 18{,}37$

$$\sigma_U = \sqrt{\frac{(2)(19)(16)\ (2{,}19{,}16 - 19 - 16)}{(19 + 16)2\ (19 + 16 - 1)}} = 2{,}89$$

$$z = \frac{27 - 18{,}37}{2{,}89} = 2{,}98$$

Pada tabel distribusi normal, $Z_{0,01/2} = 2{,}33$

Karena $Z_{hitung} > z_{0,01/2,}$, maka hipotesis nol ditolak dan menerima H_a. Artinya, bahwa *setting* mesin terlalu banyak berubah.

5. ***The Sign Test* (Uji Tanda)**

Uji Tanda merupakan uji hipotesis jika kondisi sampel tunggal untuk uji-t tidak dapat terpenuhi. Uji ini sederhana : jika sampelnya cukup besar, pendekatan yang digunakan adalah distribusi normal, jika sampel kecil bisa menggunakan pendekeatan distribusi Binomial. Jika datanya berpasangan, gunakan selisih pasangan data sebagai sampel tunggal.

$H_0 : \mu = \mu_0$

$H_a : \mu > \mu_0$ atau $\mu < \mu_0$

Prosedur uji adalah :

(1) Beri tanda ”+” untuk data yang lebih besar daripada μ_0, dan tanda ”-” untuk data yang lebih kecil daripada μ_0,

(2) Hitung banyaknya tanda ”+” untuk H_a yang berbunyi : $\mu > \mu_0$ atau hitung banyaknya tanda ”-” untuk untuk H_a yang berbunyi : $\mu < \mu_0$.

(3) Kriteria uji : terima H_0 jika $p(x) < t_{n,\alpha}$, di mana x = banyaknya tanda ”+” atau tanda ”-”.

<u>Contoh-7.34</u> : *The Sign Test* Untuk Sampel Kecil dengan Pendekatan -t

Data ukuran oktan pada BBM jenis tertentu adalah :

99,0 102,3 99,8 100,5 99,7 96,2 99,1
102,5
103,3 97,4 100,4 98,9 98,3 98,0

Ujilah hipotesis yang menyatakan bahwa rata-rata oktan jenis BBM tersebut adalah > 98,0 pada $\alpha = 0{,}01$

\+ + + + + - + +

\+ - + + + +

Tanda ”+” = x = 12. Pada Tabel-t, untuk n = 14 dan p = 0,50, probabilitas x > 12 = 1 – 0,9935 = 0,0065

Karena p(x > 12) = 0,0065 dan < α = 0,01; maka hipotesis nol ditolak, atau dapat dinyatakan secara statistik bahwa rata-rata oktan BBM tersebut di atas 98,00.

<u>Contoh-7.35</u> : *The Sign Test* Untuk Sampel Kecil dengan Pendekatan - **Binomial**

Sebuah penelitian pengujian efek analgesik Obat-A dan Obat-B terhadap artritis, dengan menggunakan 12 pasien. Hasil pengamatan terhadap 12 pasien berkaitan lamanya efek obat itu adalah :

Tabel 7.35 Waktu Efektif Obat-A dan Obat-B Terhadap Artritis (dalam satuan jam)

No.	Obat A	Obat B	No.	Obat A	Obat B
1	2,0	3,5	7	14,9	16,7
2	3,6	5,7	8	6,6	6,0
3	2,6	2,9	9	2,3	3,8
4	2,6	2,4	10	2,0	4,0
5	7,3	9,9	11	6,8	9,1
6	3,4	3,3	12	8,5	20,9

Apakah waktu efektif untuk meredakan artritis kedua obat itu berbeda pada α = 0,05?

H_0 : median selisih waktu efektif = 0

H_a : median selisih waktu efektif ≠ 0

Selisih waktu efektif kedua merk obat adalah : +1,5; +2,1; +0,3; −0,2; +2,6; −0,1; +1,8; −0,6; +1,5; +2,0; +2,3; dan +12,4. Median dari data ini = 1,65 jam. Tanda positif = 9, tanda negatif = 3, n = 12. Tanda terbanyak adalah tanda positif, yaitu = 9. Hasil perhitungan probabilitas Binomial untuk n = 12, probabilitas (X = 3) (uji dua sisi) = 0,146. Pada Excel, nilai ini dapat dihitung dengan perintah : '=BINOMDIST(3;12;0,50)', hasilnya = 0,072998. Karena merupakan uji dua sisi, maka nilai tersebut dikalikan 2 = 0,146.

Angka ini > 0,05; maka H_0 diterima, H_a ditolak; artinya kedua merk obat tersebut menghasilkan waktu efek peredaan artritis, sama.

<u>Contoh 7.36</u> : *The Sign Test* Untuk Sampel Kecil dengan Pendekatan – Binomial

Sebuah perusahaan menyatakan bahwa mereka menawarkan sebuah terapi untuk menekan turunnya memori pasien pelupa, sampai . Untuk menguji pernyataan ini mereka memeriksa 15 orang pasien pelupa untuk menghitung persentase turunnya memori setelah diterapi. α yang digunakan = 0,05. Hasilnya adalah:

Tabel 7.36 Tingkat Penurunan Memori Pasien Pelupa

	A	B	C
1	No. Pasien	Penurunan Memori (%)	Tanda
2	1	50	1
3	2	15	-1
4	3	25	1
5	4	12	-1
6	5	45	1
7	6	3	-1
8	7	45	1
9	8	8	-1
10	9	10	-1

Syafiadi Rizki Abdila, S.T., Ph.D, Prof. Dr. Ir. Drs. Syafwandi, M.Sc
Dr. Yulasmi, S.E., M.M, Dr. Robby Dharma, S.E., M.M
Dr. Syafri Arlis, S.Kom, M.Kom

11	10	8	-1
12	11	7	-1
13	12	20	0
14	13	9	-1
15	14	9	-1
16	15	12	-1

Median penurunan memori yang diharapkan = 20,00%. Data yang > 20,00% diberi tanda +1, < 20,00% diberi tanda -1. Data = 20,00%, diberi tanda = 0. Dari sampel tersebut median penurunan memori = 12,00%.

H_0 : median < 20, ini berarti terapi tidak efektif

H_a : median > 20, ini berarti terapi efektif. Ini merupakan uji satu sisi.

Dengan Excel, median data dari sampel dapat dicari dengan menggunakan formula: '=MEDIAN(B2:B16)'. Data dengan nilai tanda = 0, dikeluarkan dari pengamatan, sehingga pengamatannya menjadi 14. Banyaknya +1, dapat dihitung dengan: '=COUNTIF(C2:C16,1)', hasilnya = 4. Banyaknya -1, dihitung dengan '=COUNTIF(C4:C16,-1)', hasilnya = 10.

Gunakan perintah Excel untuk menghitung probabilitas Binomial, dengan mempertimbangkan H_a, yaitu menguji berapa probabilitas X = 4 pada n = 14, dengan probabilitas pada setiap percobaan = 0,50. Ini bisa menggunakan perintah '=BINOMDIST(4,14;0,5;TRUE)'. Hasilnya adalah = 0,0989. Nilai ini > 0,05, maka H_0 diterima; artinya terapi tersebut tidak efektif menekan penurunan memori pasien lebih dari 20,00%.

<u>Contoh 7-37</u> : *The Sign Test* **Untuk Sampel Besar Dengan Pendekatan χ^2**

Uji tanda dapat juga diaplikasi untuk sampel besar, dengan selisih antar pengamatan berpasangan. Misal data sesudah/sebelum perlakuan. Selisih atau beda dua buah variabel random disimbulkan sebagai (x – y) dengan median selisih = m. Uji tanda dilakukan untuk menguji apakah m = 0, yang berarti bahwa selisih itu tidak signifikan. Hipotesis yang diuji adalah :

$H_0 : m = 0$

$H_a : m \neq 0$

Atau dapat dinyatakan bahwa probabilitas tanda positif sama dengan probabilitas tanda negatif, dengan demikian probabilitas masing-masing tanda = 0,5. Hipotesis tersebut dapat juga dinyatakan sebagai:

$H_0 : P\{+\} = 0{,}50$

$H_a : P\{+\} \neq 0{,}50$

Formula pengujian dengan χ^2, di mana formula yang digunakan adalah :

$$\boxed{\chi^2 = \frac{(|n_1 - n_2| - 1)^2}{n_1 + n_2}} \qquad \text{. . . . (7.27)}$$

di mana

n_1 = banyaknya data dengan tanda positif

n_2 = banyaknya data dengan tanda negatif

Nilai di dalam tanda | | merupakan nilai absolut.

Kriteria uji adalah : H_0 diterima jika $\chi^2 < \chi^2_{\alpha,df}$ dan H_0 ditolak jia jika $\chi^2 > \chi^2_{\alpha,df}$.

Sebuah penelitian bertujuan untuk mengetahui apakah program DIKLAT untuk buruh dapat meningkatkan kinerja buruh. Untuk itu dilakukan pengamatan kepada 30 orang karyawan yang telah mengikuti program DIKLAT yang dimaksud. Kinerja diukur dalam format indeks produktivitas individual, α yang digunakan = 0,05. Datanya adalah:

Tabel 7.37 Kinerja Buruh Sebelum dan Sesudah Program DIKLAT

No. Pengamatan	Kinerja Sebelum	Kinerja Sesudah	Tanda
1	2,3	4,0	+
2	2,4	4,2	+
3	3,0	3,8	+
4	2,7	3,4	+
5	2,4	2,3	-
6	2,6	2,5	-
7	3,1	3,4	+
8	2,8	3,3	+
9	2,7	2,6	-
10	3,2	3,0	-
11	3,3	3,2	-
12	2,8	3,0	+
13	2,6	3,2	+
14	2,6	2,8	+
15	3,1	2,9	-
16	3,2	3,1	-
17	3,1	3,0	-

18	3,0	3,1	+
19	2,8	2,9	+
20	2,7	2,8	+
21	2,9	2,8	-
22	3,2	3,1	-
23	2,8	2,9	+
24	3,0	2,8	-
25	2,7	3,0	+
26	2,8	3,1	+
27	3,1	3,0	-
28	3,2	3,1	-
29	2,6	2,5	-
30	2,0	2,2	+

Tanda positif = $n_1 = 16$

Tanda negatif = $n_2 = 14$

Selanjutnya, dapat dihitung nilai χ^2 dengan formula-(6.27):

$$\chi^2 = \frac{(|16 - 14| - 1)^2}{16 + 14} = 0{,}333$$

Nilai kirits χ^2 pada $\alpha = 0{,}025$ (untuk dua sisi, $\alpha = 0{,}05/2 = 0{,}025$) dengan df = 1, sama dengan = 5,024, Nilai $\chi^2 < \chi^2_{\alpha,df}$, maka H_0 diterima, artinya : program DIKLAT itu tidak efektif terhadap kinerja buruh.

<u>Contoh-7.38</u> : *The Sign Test*

Dalam banyak penelitian dalam bidang pertanian, peneliti ingin membandingkan pengaruh hasil dua perlakuan. Untuk data yang berpasangan, satu sebagai hasil perlakuan A dan satu lagi hasil perlakuan B. Untuk membandingkan kedua hasil perlakuan (ditinjau dari rata-rata) itu dapat digunakan uji tanda. Uji ini sangat baik jika memenuhi syarat berikut:

a. Pasangan hasil pengamatan yang sedang dibandingkan bersifat independen.

b. Masing-masing pengamatan dalam tiap pasang terjadi karena pengaruh kondisi yang serupa.

c. Pasangan yang berlainan terjadi karena kondisi yang berbeda.

Sebagaimana namanya dinyatakan, uji tanda ini akan dilakukan berdasarkan tanda, yakni “+” dan “–“, yang didapat dari selisih nilai pengamatan. Untuk (x – y) = 0, diabaikan. Banyaknya pengamatan minimal = 6 ($n \geq 6$)

Hipotesis yang diuji :

H_0: tidak ada perbedaan pengaruh perlakuan

H_a: ada perbedaan pengaruh perlakuan

Kriteria uji : H_0 diterima, jika nilai H > Tabel H pada 〈 dan df = n

Ha diterima, jika nilai $H \leq$ Tabel H ada 〈 dan df = n

Sebuah penelitian bertujuan mengetahui apakah perlakuan terhadap hasil produksi kacang tanah berbeda atau tidak.

Tabel 7.38 Hasil Kacang Tanah Per Rumpun (dalam ons).

No.	X	Y	Tanda	No.	X	Y	Tanda
1	3,4	3,0	+	11	4,0	3,7	+
2	3,7	3,9	-	12	3,9	4,0	-
3	2,8	3,2	-	13	3,8	3,5	+

4	4,2	4,6	-	14	4,2	4,5	-
5	4,6	4,3	+	15	4,7	3,9	+
6	3,8	3,4	+	16	4,0	3,7	+
7	3,6	3,5	+	17	3,6	3,2	+
8	2,9	3,0	-	18	3,2	2,9	+
9	3,0	2,9	+	19	3,4	3,0	+
10	3,8	3,7	+	20	2,9	3,6	-

Tanda yang paling kecil jumlahnya adalah tanda negatif, H = 7. Perhatikan Tabel H pada df = 20 dan 〈 = 0,05.

Tabel 7.39 Tabel Kirits H untuk Uji Tanda

n	α		n	α		n	α	
	0,01	0,05		0,01	0,05		0,01	0,05
6	-	0	36	9	11	66	22	24
7	-	0	37	10	12	67	22	25
8	0	0	38	10	12	68	22	25
9	0	1	39	11	12	69	23	25
10	0	1	40	11	13	70	23	26
11	0	1	41	11	13	71	24	26
12	1	2	42	12	14	72	24	27
13	1	2	43	12	14	73	25	27
14	1	2	44	13	15	74	25	28
15	2	3	45	13	15	75	25	28
16	2	3	46	13	15	76	26	28
17	2	4	47	14	16	77	26	29
18	3	4	48	14	16	78	27	29
19	3	4	49	15	17	79	27	30
20	3	5	50	15	17	80	28	30
21	4	5	51	15	18	81	28	31
22	4	5	52	16	18	82	28	31
23	4	6	53	16	18	83	29	32
24	5	6	54	17	19	84	29	32
25	5	7	55	17	19	85	30	32

26	6	7		56	17	20		86	30	33
27	6	7		57	18	20		87	31	33
28	6	8		58	18	21		88	31	34
29	7	8		59	19	21		89	31	34
30	7	9		60	19	21		90	32	35
31	7	9		61	20	22		91	32	35
32	8	9		62	20	22		92	33	36
33	8	10		63	20	23		93	33	36
34	9	10		64	21	23		94	34	37
35	9	11		65	21	24		95	34	37

Nilai H pada 〈 = 0,05 dan df = n = 20 adalah = 5. Hasil perhatiungan H = 7, maka terima H_0 dan tolak H_a. Artinya tidak ada perbedaan hasil produksi karena perlakuan

6. *The Kolmogorov – Smirnov Test*

The Kolmogorov – Smirnov Test adalah uji hipotesis non paramertrik untuk perbedaan antara distribusi kumulatif. Uji untuk sebuah sampel berkaitan dengan nilai sampel dengan fungsi distribusi kontinyu tertentu. Jadi metode ini adalah metode *goodness of fit*.

Uji untuk dua sampel berkaitan dengan distribusi kumulatif yang diamati, menguji hipotesis apakah dua sampel independen berasal dari populasi yang sama, juga sensitif terhadap perbedaan penyebaran, atau kemencengan (*skewness*).

Metode ini untuk sebuah sampel secara umum lebih efisien daripada uji χ2 *goodness of fit* untuk sampel kecil (metode ini dapat diaplikasikan pada sampel yang sangat kecil, di mana uji χ^2 tidak dapat diaplikasikan), Tetapi perlu diingat, bahwa untuk distribusi data diskret, metode ini tidak dapat diaplikasikan; sedang analisis χ^2 bisa diaplikasikan. Uji satu sampel berdasarkan pada selisih maksimum absolut antara nilai distribusi kumulatif sebuah sampel random berukuran- n yang berdistribusi tertentu. Untuk menentukan apakah selisih tersebut lebih besar daripada nilai yang diharapkan seperti pada hipotesis dengan α dan n tertentu, dapat digunakan nilai-D pada Tabel Nilai-D.

<u>Contoh-7.39</u> : Uji Kolmogorov-Smirnov.

10 titik lubang pada sebuah lempengan tembaga diharapkan berada pada jarak penyebaran yang sama. Berikut ini jarak ke-10 titik lubang yang diamati (dalam satuan inci) pada sebuah lempengan tembaga selebar 30 inci :

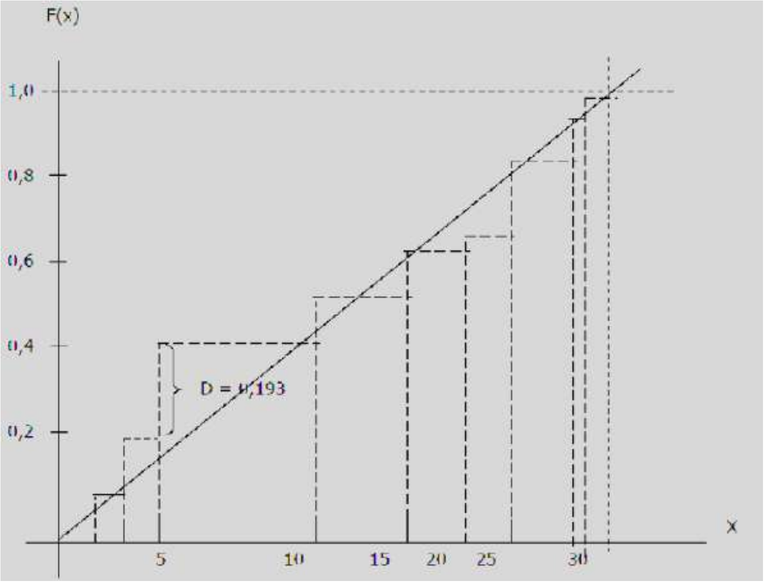
4,8 14,8 28,2 23,1 4,4 28,7 19,5 2,4
25,0 6,2

Hipotesis nol yang diuji pada $\alpha = 0{,}05$, adalah :

$$F(X) = \begin{cases} 0 & \text{untuk } x < 0 \\ x/30 & \text{untuk } 0 < x < 30 \\ 1 & \text{untuk } x > 30 \end{cases}$$

Di mana x adalah jarak titik lubang dari tepi lapisan tembaga.

Kriteria uji : tolak H_0 jika $D > 0{,}410$, di mana D adalah jarak maksimum antara distribusi kumulatif yang diamati dan distribusi kumulatif tersebut diasumsikan seperti pada hipotesis nol. Buat plot kedua distribusi kumulatif seperti pada gambar berikut:



Syafiadi Rizki Abdila, S.T., Ph.D, Prof. Dr. Ir. Drs. Syafwandi, M.Sc
Dr. Yulasmi, S.E., M.M, Dr. Robby Dharma, S.E., M.M
Dr. Syafri Arlis, S.Kom, M.Kom

Pada gambar di atas tampak bahwa jarak distribusi kumulatif yang terbesar terjadi pada x = 6,2. Jarak tersebut = D = 0,40 – 6,2/30 = 0,193. Pada Tabel D dengan n = 10, nilai D sebesar = 0,410. Karena $D_{Perhitungan} < D_{Tabel}$, maka H_0 diterima. Artinya, ke-10 titik lubang memiliki jarak dengan distribusi seragam.

<u>Contoh-7.40</u> : Uji K-S untuk Kesesuaian Distribusi Data Usia Kerusakan

Uji sebuah sampel Kolmogorov-Smirnov juga digunakan untuk mengetahui apakah data sampel berdistribusi teoritis tertentu (misal : normal, uniform, Poisson, dan eksponensial). Uji Kolmogorov-Smirnov sesuai untuk variabel kontinyu dengan sampel yang dipilih secara random. Atau dengan kata lain adalah bahwa uji ini hanya dapat digunakan bila variabel diukur paling tidak berupa skala ordinal. Secara umum, penerapan uji Kolmogorov–Smirnov dapat diterapkan pada dua kondisi, yaitu : (1) menguji apakah data sebuah sampel berdistribusi populasi teoritis tertentu; (2) menguji apakah dua buah sampel berasal dari dua populasi yang identik.

Uji Kolmogorov–Smirnov dilakukan dengan menghitung selisih absolut antara fungsi distribusi frekuensi kumulatif sampel (S[x]) dengan fungsi distribusi frekuensi kumulatif teoritis (F_0[x]) pada masing-masing kelas interval.

Hipotesis yang diuji adalah :

$$H_0 : F(x) = F_0(x), \text{ untuk } \forall \overset{+\infty}{\underset{-\infty}{+}} x$$

Ha : $F(x) \neq F_0(x)$ untuk paling sedikit sebuah x

Statistik uji Kolmogorov-Smirnov merupakan selisih absolut terbesar antara S[x] dan F_0[x], yang disebut deviasi maksimum D.

$$D = \text{Max } \{|S[x] - F_0[x]|\} \text{ untuk } i = 1, 2, \ldots, n$$

Selanjutnya, nilai D dibandingkan dengan nilai kritis pada Tabel D, pada ukuran sampel n dan α tertentu. Kriteria uji adalah : tolak H_0 jika nilai $D > D_{\alpha,n,}$. Penolakan H_0 mengindikasikan bahwa distribusi data tidak sama dengan distribusi teoritis.

Langkah-langkah dalam uji Kolmogorov-Smirnov adalah sebagai berikut :

(1) Urutkan tiap data yang diobservasi, dari nilai terkecil sampai nilai terbesar (*ascending*). Kemudian hitung proporsi berdasar rank-nya dalam nilai proporsi, S[x].

(2) Hitung nilai z untuk setiap data observasi, di mana $z = (Xt–X)/\sigma x$, Gunakan tabel distribusi normal standard untuk mengkonversi nilai z tersebut ke nilai luas areal di bawah kurve (nilai probabilitas), ini merupakan nilai $F_0[x]$.

(3) Hitung selisih absolut (D) = $S[x] - F_0[x]$ pada setiap pengamatan.

(4) Bandingkan nilai D dengan nilai D pada Tabel Kolmogorov-Smirnov.

Sebuah penelitian bertujuan menguji apakah data usia mesin fotokopi mengalami kerusakan pada pertama kalinya berdistribusi normal atau tidak. Tingkat α yang digunakan = 10,00%. Data yang dikumpulkan dari beberapa perusahaan jasa fotokopi adalah (telah diurutkan berdasar usia kerusakan awal) :

Tabel 7.40 Usia Kerusakan awal Pada Mesin Fotokopi

No.	Usia Pada Kerusakan Awal (tahun)	*Rank*
1	2,50	1
2	3,00	2
3	3,40	3
4	4,00	4

5	4,25	5
6	4,50	6
7	4,80	7
8	5,00	8

Hipotesis yang diuji adalah:

H_0 : usia kerusakan awal mesin fotokopi berdistribsi normal

H_a : usia kerusakan awal mesin fotokopi tidak berdistribsi normal

Tabel 7.41 Algoritma Perhitungan S[x] dan F_0[x]

No.	Usia pada Kerusakan Awal	Rank	Z	F_0(x)	S[X]	D = ∠S[X]-F_0(x) ∠
1	2,50	1	-1,61	0,054	0,125	0,071
2	3,00	2	-1,05	0,147	0,250	0,103
3	3,40	3	-0,60	0,275	0,375	0,100
4	4,00	4	0,08	0,531	0,500	0,031
5	4,25	5	0,36	0,640	0,625	0,015
6	4,50	6	0,64	0,739	0.750	0,011
7	4,80	7	0,98	0,836	0,875	0,039
8	5,00	8	1,20	0,886	1,000	0,114
Rata-rata X	3,931					
ʃX	0,888					

S[x] adalah proporsi kumulatif rank, contoh : pada observasi-1, rank = 1, maka S[x] = 1/8 = 0,125. Pada observasi-2, rank = 2, maka S[x] = 2/8 = 0,250.

Nilai z dihitung sebagai $(X_t - \overline{X})/\sigma_x$, Rata-rata data usia = 3,931, dengan standard deviasi = 0,888; contoh : pada observasi-1, usia = 2,50, maka z = (2,50 – 3,931)/0,888 = - 1,61 . Nilai D dihitung sebagai nilai absolut S[x] – F_0[x]. Nilai D maksimum = 0,114.

Selanjutnya bandingkan nilai D_{max} ini dengan nilai pada D_{tabel} pada n = 8 (dengan α/2 = 0,05 untuk uji dua sisi) :

Tabel 7.42 Tabel D Kolmogorov – Smirnov[13].

N	α = 0,20	α = 0,15	α = 0,10	α = 0,05	α = 0,01
1	0,900	0,925	0,950	0,975	0,995
2	0,684	0,726	0,776	0,842	0,929
3	0,565	0,597	0,642	0,708	0,828
4	0,494	0,525	0,564	0,624	0,733
5	0,446	0,474	0,510	0,565	0,669
6	0,410	0,436	0,470	0,521	0,618
7	0,381	0,405	0,438	0,486	0,577
8	0,358	0,381	0,411	0,457	0,543
9	0,339	0,360	0,388	0,432	0,514
10	0,322	0,342	0,368	0,410	0,490
11	0,307	0,326	0,352	0,391	0,468
12	0,295	0,313	0,338	0,375	0,450
13	0,284	0,302	0,325	0,361	0,433
14	0,274	0,292	0,314	0,349	0,418
15	0,266	0,283	0,304	0,338	0,404
16	0,258	0,274	0,295	0,328	0,392
17	0,250	0,266	0,286	0,318	0,381
18	0,244	0,259	0,278	0,309	0,371
19	0,237	0,252	0,272	0,301	0,363
20	0,232	0,265	0,294	0,329	0,352
21	0,226	0,259	0,287	0,321	0,344
22	0,221	0,253	0,281	0,314	0,337
23	0,216	0,247	0,275	0,307	0,330

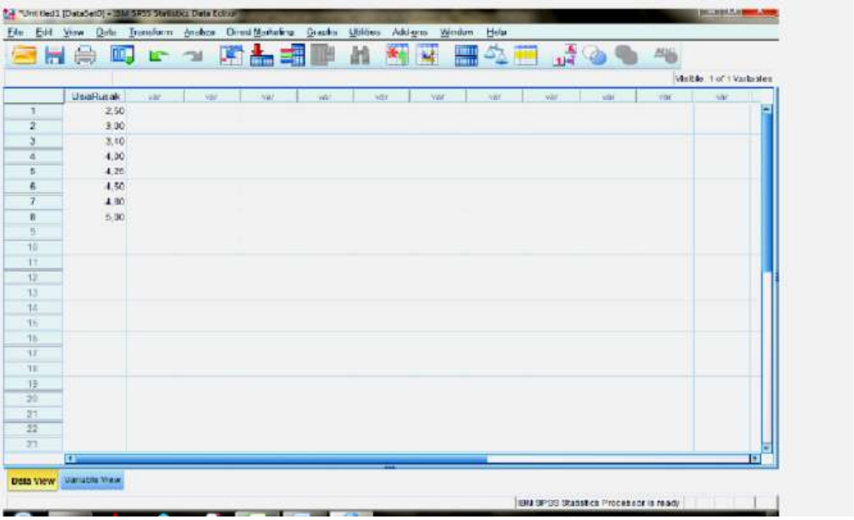
[13] Diadaptasi dari Tabachnick, 2000:705.

24	0,212	0,242	0,269	0,301	0,323
25	0,208	0,238	0,264	0,295	0,317
26	0,204	0,233	0,259	0,290	0,311
27	0,200	0,229	0,254	0,284	0,305
28	0,197	0,225	0,250	0,279	0,300
29	0,193	0,221	0,246	0,275	0,295
30	0,190	0,218	0,242	0,270	0,290
> 50	$1{,}07/\sqrt{n}$	$1{,}14/\sqrt{n}$	$1{,}22/\sqrt{n}$	$1{,}36/\sqrt{n}$	$1{,}63/\sqrt{n}$

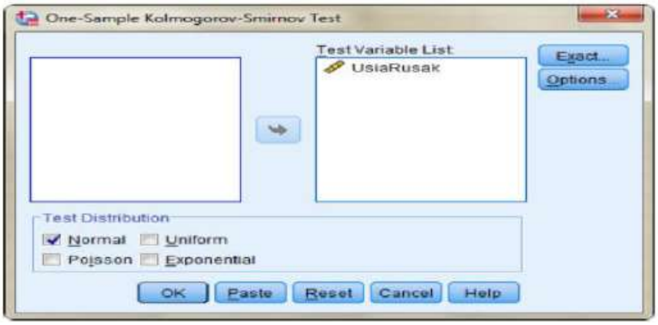
D_{max} hasil perhitungan = 0,114 (< 0,457), dengan demikian H_0 diterima. Artinya data usia kerusakan mesin fotokopi untuk pertama kalinya berdistribusi normal.

Aplikasi dengan SPSS, uji Kolmogorov-Smirnov dengan *One Sample* adalah:

(1) Membuat data dengan file SPSS :



(2) Pilih : '*Analyze – Non Paramteric Tests – Legacy Dialogs – 1 Sample K-S'* :



(3) Outputnya :

	N	Mean	Std. Deviation	Minimum	Maximum
UsiaRusak	8	3,9313	,88839	2,50	5,00

One-Sample Kolmogorov-Smirnov Test

		UsiaRusak
N		8
Normal Parameters[a,b]	Mean	3,9313
	Std. Deviation	,88839
Most Extreme Differences	Absolute	,156
	Positive	,114
	Negative	-,156
Kolmogorov-Smirnov Z		,441
Asymp. Sig. (2-tailed)		,990

a. Test distribution is Normal.

b. Calculated from data.

D_{max} hasil perhitungan = 0,114 (< 0,328), dengan demikian H_0 diterima. Artinya data usia kerusakan mesin fotokopi untuk pertama kalinya berdistribusi normal.

Keunggulan penggunaan uji Kolomogorov Smirnov dibanding kan dengan

penggunaan uji χ^2 adalah : (1) dalam uji K-S ini tidak dibutuhkan pengkatagorian data, sehingga seluruh informasi observasi dapat digunakan; (2) tidak ada batasan ukuran sampel minimum dalam uji K-S ini. Kelemahannya adalah : (1) uji K-S tidak bisa digunakan untuk mengestimasi parameter populasi, dan (2) uji K-S menggunakan asumsi bahwa distribusi populasi teoritis bersifat kontinyu.

<u>Contoh-7.41</u> : Uji K-S untuk Kesesuaian Distribusi Data Nilai UAS

Sebuah penelitian bertujuan mengetahui apakah data nilai UTS Matematik pada sebuah kelas berdistribusi normal. Untuk itu diamati nilai 27 mahasiswa yang telah mengikuti UTS, data yang terkumpul adalah:

Tabel 7.43 Nilai UTS Matematika

No.	Nilai	No	Nilai	No.	Nilai
1	87	11	67	21	86
2	88	12	68	22	98
3	85	13	59	23	67
4	65	14	82	24	67
5	57	15	69	25	87
6	80	16	70	26	54
7	89	17	75	27	56
8	75	18	81		
9	73	19	70		
10	72	20	76		

Data tersebut kemudian diurutkan dari terkecil ke terbesar:

Nilai	Rank	Nilai	Rank	Nilai	Rank	Nilai	Rank
54	1	67	8	75	15	86	22

56	2	68	9	75	16	87	23
57	3	69	10	76	17	87	24
59	4	70	11	80	18	88	25
65	5	70	12	81	19	89	26
67	6	72	13	82	20	98	27
67	7	73	14	85	21		

Tabel 7.44 Algoritma Perhitungan S[x] dan F_0[x]

No.	Nilai UTS	Rank	Z	F_0[x]	S[x]	$D = \angle S[x]-F_0(x) \angle$
1	54	1	-1,79	0,0370	0,037	0,000
2	56	2	-1,61	0,0537	0,074	0,020
3	57	3	-1,52	0,0641	0,111	0,047
4	59	4	-1,34	0,0894	0,148	0,059
5	65	5	-0,81	0,2081	0,185	0,023
6	67	6	-0,64	0,2624	0,296	0,034
7	67	7	-0,64	0,2624	0,296	0,034
8	67	8	-0,64	0,2624	0,296	0,034
9	68	9	-0,55	0,2920	0,333	0,041
10	69	10	-0,46	0,3212	0,370	0,047
11	70	11	-0,37	0,3555	0,444	0,088
12	70	12	-0,37	0,3555	0,444	0,088

13	72	13	-0,19	0,4233	0,481	0,058
14	73	14	-0,10	0,4582	0,519	0,060
15	75	15	0,07	0,5287	0,593	0,064
16	75	16	0,07	0,5287	0,593	0,064
17	76	17	0,16	0,5638	0,630	0,066
18	80	18	0,51	0,6966	0,667	0,030
19	81	19	0,60	0,7268	0,704	0,023
20	82	20	0,69	0,7555	0,741	0,015
21	85	21	0,96	0,8308	0,778	0,053
22	86	22	1,05	0,8522	0,815	0,037
23	87	23	1,13	0,8717	0,889	0,020
24	87	24	1,13	0,8717	0,889	0,017
25	88	25	1,22	0,8893	0,926	0,037
26	89	26	1,31	0,9051	0,963	0,058
27	98	27	2,11	0,9825	1,000	0,047

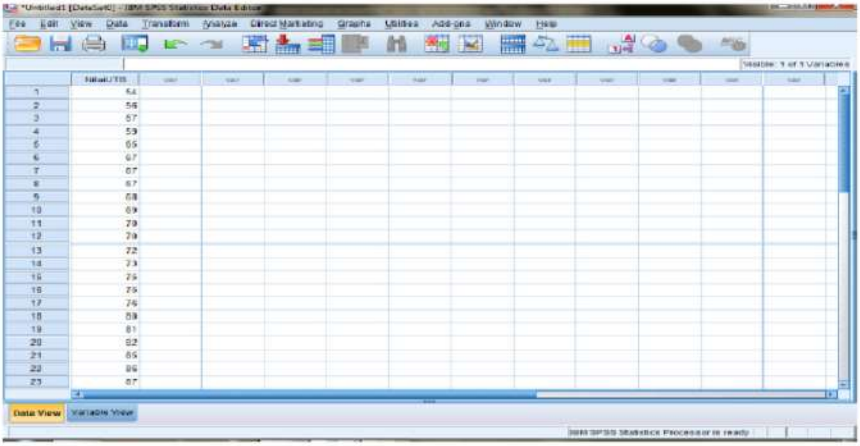
Rata-rata nilai (X) = 74,1852, dan standard deviasi = 11,2968

Nilai $F_0[x]$ dihitung sebagai proporsi R_t terhadap R terbesar (=27). Untuk data nilai yang sama, maka $F_0[x]$ yang digunakan adalah nilai proporsi terbesar. Contoh, pada observasi ke-6, 7, dan 8, nilai UTS = 67, maka $F_0[x]$ untuk ketiganya = 0,296. Nilai D_{max} = 0,089.

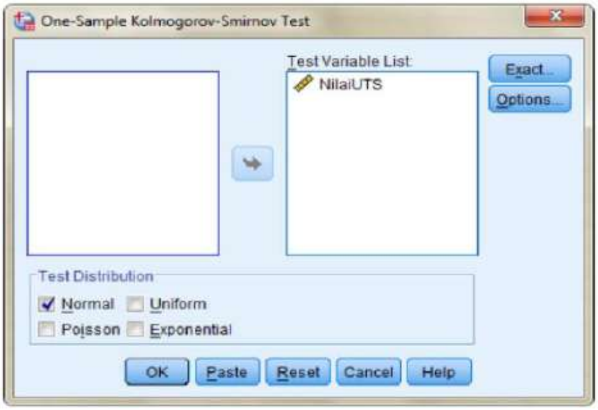
Bandingkan nilai D_{max} dengan D_{tabel} pada n = 27 dan $\alpha/2$ = 0,05, yaitu = 0,284. Nilai $D_{max} < D_{tabel}$; maka H_0 diterima, artinya nilai UTS ke-27 mahasiswa berdistribusi normal.

Aplikasi kasus ini dengan SPSS adalah sebagai berikut :

(1) Membuat file data SPSS:



(2) Pilih : '*Analyze – Non Parametric Tests – Legacy Dialogs – 1 Sample K-S*' :



(3) *Output* hasil :

One-Sample Kolmogorov-Smirnov Test

		NilaiUTS
N		27
Normal Parameters[a,b]	Mean	74,19
	Std. Deviation	11,297
Most Extreme Differences	Absolute	,089
	Positive	,089
	Negative	-,089
Kolmogorov-Smirnov Z		,468
Asymp. Sig. (2-tailed)		,981

a. Test distribution is Normal.

b. Calculated from data.

Hasil SPSS memperlihatkan bahwa D_{max} = 0,089. Atau juga bisa diindikasikan dari nilai Asymp. Sig. (2-tailed) = 0,981 > α yang digunakan, maka tolak H_a dan terima H_0.

<u>Contoh-7.42</u>: Uji K-S Untuk Kesesuaian Distribusi (pada Tabel Frekuensi)

Untuk menguji normalitas data pada sebuah tabel frekuensi (X, f), uji K-S tetap dapat digunakan dengan modifikasi pada perhitungan rata-rata dan standard deviasinya.

$$\text{Rata-rata} : \overline{X_t f_t} = \frac{\Sigma X_t f_t}{n}$$

$$\sigma_{Xf} = \sqrt{\frac{\{(\Sigma Xt^2 f_t)/N - [(\Sigma Xifi)/n]^2\}(n)}{(N-1)}} \qquad \text{di mana } n = \Sigma f_t$$

Berikut ini adalah data penelitian :

Tabel 7.45 Data Frekuensi Hipotetik Penelitian

Range (X_t)	100	200	300	400	500	600	700	800	900	1.000
Frek (f_t)	8	25	88	172	243	252	144	49	13	6

Ujilah apakah data pada tabel frekuensi di atas berdistribusi normal atau tidak pada α = 0,10; dengan Uji Kolmogorov-Smirnov.

Rata-rata dan standard deviasi data distribusi frekuensi tersebut dihitung terlebih dahulu seperti yang terlihat pada tabel di bawah ini :

Tabel 7.46 Perhitungan Rata-rata dan Standard Deviasi.

Mid Point (X_t)	f_t	X_t^2	X_tf_t	$X_t^2f_t$	Parameter	Nilai	Formula
50	8	2500	400	2000	n	1.000	= © f_t
150	25	22500	3750	562500	Rata-rata	481,40	= (©X_tf_t)/n
250	88	62500	22000	5500000	Kuadrat Rata-rata	231745,96	= [(©X_tf_t)/n]^2
350	172	122500	60200	21070000	Rata-rata Jumlah Kuadrat	255800,00	= (©$X_t^2f_t$)/n
450	243	202500	109350	49207500			
550	252	302500	138600	76230000	VARP	24054,04	(©$X_t^2f_t$)/n - [(©X_tf_t)/n]^2
650	144	422500	93600	60840000			
750	49	562500	36750	27562500	VAR	24078,12	=VARP(n)/(n-1)
850	13	772500	11050	9392500			
950	6	902500	5700	5415000	Standard Dviasi	155,17	=√VAR
Jumlah	1.000		481400	255800000			

Setelah rata-rata dan standard deviasi data pada tabel frekuensi terhitung, selanjutnya menghitung D dengan algoritma seperti biasa:

Tabel 7.47 Algoritma Perhitungan S[x] dan F_0[x]

No.	X_t	f_t	Kum. f_t	S[x]	F_0[x]	D = ∠S[x]-F_0(x) ∠
1	100	8	8	0,008	0,006987	0,00101
2	200	25	33	0,033	0,034879	0,00188
3	300	88	121	0,121	0,121196	0,00020
4	400	172	293	0,293	0,299937	0,00694
5	500	243	536	0,536	0,547706	0,01171
6	600	252	788	0,788	0,777661	0,01034
7	700	144	932	0,932	0,920548	0,01145
8	800	49	981	0,981	0,979974	0,00103
9	900	13	994	0,994	0,999509	0,00251
10	1.000	6	1.000	1,000	0,999584	0,00042

S[x] = merupakan proporsi kumulatif f_t; F_0[x] merupakan probabilitas kurve normal untuk X_t, dengan rata-rata = 481,40 dan standard deviasi = 155,17. D_{max} = 0,01171. Nilai D Kolmogorov-Smirnov untuk n = 1.000 dan α/2 = 0,05 didekati dengan $D_{1000,.05}$ = 1,36 /√1000 = 0,043007. Karena D_{max} < $D_{1000,0.05}$, maka H_0 diterima, artinya data pada tabel frekuensi berdistribusi normal.

7.5 Rangkuman

Ada dua metode dalam uji hipotesis, yaitu: (a) parametrik; dan (b) non parametrik. Perbedaan kedua metode terletak pada penggunaan parameter data atau tidak. Kedua metode menghasilkan kesimpulan uji yang sama.

1. Uji hipotesis yang berkaitan dengan satu rata-rata.

Untuk sampel besar (n > 30), daerah kritis pengujian:

H_0	Tolak H_0 jika:
$\mu < \mu_0$	$Z < -Z_{\alpha}$
$\mu > \mu_0$	$Z > Z_{\alpha}$
$\mu \neq \mu_0$	$Z < -Z_{\alpha/2}$ atau $Z > Z_{\alpha/2}$

Di mana,

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \qquad \text{.....}\ (7.1)$$

Untuk sample kecil (n < 30), prosedur uji hipotesisnya sama, dengan perbedaan padakriteria ujinya, yaitu:

$$t = \frac{X - \mu_0}{s/\sqrt{n}} \qquad \text{.....}\ (7.2)$$

2. Uji hipotesis yang berkaitan dengan dua rata-rata.

Untuk sampel besar (n > 30), populasi normal, σ_1 dan σ_2 diketahui, daerah kritis pengujian adalah:

H_0	Tolak H_0 jika:
$\mu_1 - \mu_2 < d_0$	$Z < -Z_{\alpha}$
$\mu_1 - \mu_2 > d_0$	$Z > Z_{\alpha}$
$\mu_1 - \mu_2 \neq d_0$	$Z < -Z_{\alpha/2}$ atau $Z > Z_{\alpha/2}$

Syafiadi Rizki Abdila, S.T., Ph.D, Prof. Dr. Ir. Drs. Syafwandi, M.Sc
Dr. Yulasmi, S.E., M.M, Dr. Robby Dharma, S.E., M.M
Dr. Syafri Arlis, S.Kom, M.Kom

Formula uji hipotesis perbedaan dua rata-rata :

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - \text{™}}{\sqrt{(\sigma_1^2/n_1 + \sigma_2^2/n_2)}} \qquad \text{......(7.3)}$$

Untuk sampel kecil (n < 30), maka formula ujinya menjadi :

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - \text{™}}{\sqrt{[(n_1-1)s_1^2 + (n_2-1)s_2^2]}} \sqrt{\frac{n_1.n_2\,(n_1 + n_2 - 2)}{n_1 + n_2}} \qquad \text{..... (7.4)}$$

3. Uji Hipotesis untuk Sebuah Varians.

Untuk sampel random (n < 30) dari populasi normal, σ_0^2 diketahui sebagai nilai darivariabel random yang berdistribusi χ^2 dengan derajat bebas = n - 1, daerah kritis pengujian:

H_0	Tolak H_0 jika:
$\sigma^2 < \sigma_0^2$	$\chi^2 < \chi^2_{1-\alpha}$
$\sigma^2 > \sigma_0^2$	$\chi^2 > \chi^2_{\alpha}$
$\sigma^2 \neq \sigma_0^2$	$\chi^2 < \chi^2_{1-\alpha/2}$ atau $\chi^2 > \chi^2_{\alpha/2}$

Formula uji hipotesis sebuah varians:

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} \qquad \text{. . . . .(7.6)}$$

Untuk sampel besar (n > 30), formula uji hipotesis yang digunakan adalah:

$$Z = \frac{s - \sigma_0}{\sigma_0/\sqrt{2n}} \qquad \text{. . . . . (7.7)}$$

4. Uji hipotesis untuk dua varians.

Formula uji untuk dua varians adalah :

$$F = \frac{s_1^2}{s_2^2} \qquad \text{. . . . . (7.8)}$$

5. Uji Hipotesis untuk Sebuah Proporsi.

Banyak metode yang digunakan untuk pemeriksaan dalam kendali mutu, dan ujikehandalan dengan berbasis pada uji hipotesis nol yang menyatakan proporsi sama dengankonstanta tertentu.

Untuk sampel besar (n > 30) digunakan formula uji sebagai berikut :

$$z = \frac{X - np_0}{\sqrt{np_0\,(1 - p_0)}} \qquad \text{. . . . . (7.9)}$$

6. Uji Hipotesis untuk Banyak Proporsi

Formula uji:

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(o_{ij} - e_{ij})^2}{e_{ij}} \qquad \text{. . . . . (7.10)}$$

ANOVA

Tabel *One Factor ANOVA*:

Sumber Variasi	Jumlah Kuadrat	Derajat Bebas	Kuadrat Rata-2	Statistik-F
Perlakuan (antar kelompok)	$SSA = \sum_{j=1}^{c} n_j(\bar{y}_j - \bar{y})^2$	c - 1	$MSA = \frac{SSA}{c-1}$	$F = \frac{MSA}{MSE}$
Error (dalam kelompok)	$SSE = \sum_{j=1}^{c} \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2$	N - c	$MSE = \frac{SSE}{n-c}$	
Total	$SST = \sum_{j=1}^{c} \sum_{i=1}^{n_j} (y_{ij} - \bar{y})^2$	N - 1		

Tabel *Two Factor ANOVA without replication*:

Sumber Variasi	Jumlah Kuadrat	Derajat Bebas	Kuadrat Rata-2	Statistik-F
Faktor A (efek baris)	$SSA = c\sum_{j=1}^{r} (\overline{y}_j - \overline{y})^2$	r - 1	$MSA = \frac{SSA}{r-1}$	$F_A = \frac{MSA}{MSE}$
Faktor B (efek kolom)	$SSB = r\sum_{k=1}^{c} (\overline{y}_k - \overline{y})^2$	c – 1	$MSB = \frac{SSB}{c-1}$	$F_B = \frac{MSB}{MSE}$
Error	$SSE = \sum_{j=1}^{r}\sum_{k=1}^{c} (y_{jk} - \overline{y}_j - \overline{y}_k + \overline{y}_k)^2$	(r – 1)(c – 1)	$MSE = \frac{SSE}{(r-1)(c-1)}$	
Total	$SST = \sum_{j=1}^{c}\sum_{k=1}^{r} (y_{jk} - \overline{y})^2$	rc – 1		

Tabel *Two Factor ANOVA with replication*:

Sumber Variasi	Jumlah Kuadrat	Derajat Bebas	Kuadrat Rata-2	Statistik-F
Faktor A (efek baris)	$SSA = cm\sum_{j=1}^{r} (\overline{y}_j - \overline{y})^2$	r - 1	$MSA = \frac{SSA}{r-1}$	$F_A = \frac{MSA}{MSE}$
Faktor B (efek kolom)	$SSB = rm\sum_{k=1}^{c} (\overline{y}_k - \overline{y})^2$	c – 1	$MSB = \frac{SSB}{c-1}$	$F_B = \frac{MSB}{MSE}$
Interaksi (A x B)	$SSI = m\sum_{j=1}^{r}\sum_{k=1}^{c} (\overline{y}_{jk} - \overline{y}_j - \overline{y}_k + y)^2$	(r – 1)(c – 1)	$MSI = \frac{SSI}{(r-1)(c-1)}$	$FI = \frac{MSI}{MSE}$
Error	$SSE = \sum_{i=1}^{m}\sum_{j=1}^{r}\sum_{k=1}^{c} (y_{ijk} - \overline{y}_{jk})^2$	rc(m – 1)	$MSE = \frac{SSE}{(r-1)(c-1)}$	
Total	$SST = \sum_{i=1}^{m}\sum_{j=1}^{r}\sum_{k=1}^{c} (y_{ijk} - \overline{y})^2$	rcm – 1		

Uji hipotesis non parametrik memiliki beberapa keunggulan:

(1) Asumsi dalam metode non parametrik relatif lebih longgar dibanding dengan metode parametrik.
(2) Stokastik perhitungan lebih mudah dan lebih cepat.
(3) Untuk pemahaman konsep dan metodenya tidak membutuhkan pemahaman ilmu matematika atau ilmu statistik yang mendalam.
(4) Metode non parametrik dapat diterapkan untuk data dengan pengukuran yang kurang terstandard (misal : nominal atau ordinal).
(5) Efisiensi statistik non parametrik lebih tinggi dibandingkan dengan metode parametrik untuk jumlah sampel yang kecil.

Beberapa kelemahan metode non parametrik adalah :

(1) Jika asumsi pada uji statistik parametrik dapat terpenuhi, penggunaan metode ini menyebabkan pemborosan informasi.

(2) Untuk ukuran sampel yang besar, tingkat efisiensi metode ini relatif lebih rendah dibandingkan dengan metode parametrik.

8. Uji Chi Square (χ^2)

Formula untuk menghitung χ^2:

$$\chi^2 = \Sigma \frac{(f_e - f_o)^2}{f_e} \qquad \text{. . . . . (7.11)}$$

Jika pada matriks analisis χ^2 dengan dimensi 2 x 2 ada sel dengan nilai < 10, formula-(7.11) harus dikoreksi. Koreksi yang dikembangkan oleh Yates adalah sebagai berikut:

$$\chi^2 = \frac{n\,[(a)(d) - (b)(c) - n/2]^2}{(a + b)(c + d)(a + c)(b + d)} \qquad \text{. . . . . (7.13)}$$

9. The Wilcoxon Signed Rank Test

(a) median populasi yang simetri dengan sampel kecil dan sampel besar.

(b) median perbedaan data berpasangan.

10. The Wilcoxon Rank Sum Test

Formula uji:

$$z = \frac{U1 - \mu_{U1}}{\sigma_{U1}} \qquad \text{. . . . . (7.19)}$$

11. Mann-Whitney Test

$$Z = \frac{\overline{R}_1 - \overline{R}_2}{\sqrt{\frac{n_1 + n_2 + 1}{(12)(n_1)(n_2)}}\,(n_1 + n_2)} \qquad \text{. . . . . (7.20)}$$

12. Uji Kruskal Wallis (H-test)
Formula uji:

$$H = \frac{\text{SS bg(R)}}{N(N+1)/12} \qquad \text{. . . . . (7.22)}$$

13. Uji Random atau Run Test
Formula uji:

$$H = \frac{\text{SS bg(R)}}{N(N+1)/12} \qquad \text{. . . . . (7.24)}$$

14. *The Sign Test* (Uji Tanda)
Untuk sampel besar :

$$\chi^2 = \frac{(|n_1 - n_2| - 1)^2}{n_1 + n_2} \qquad \text{. . . . (7.27)}$$

7.6 Diskusi

Mendeskripsikan hipotesis perlu hati-hati. Pada beberapa penelitiasn empiris yang menggunakan alat analisis statistik, terkadang ditemui pernyataan hipotesis yang kurang tepat. Yang perlu dipahami dengan benar, hipotesis nol menyatakan kenihilan, contoh : tidak ada perbedaan signifikan antar dua parameter populasi yang berbeda, atau parameter 1 = parameter-2. Sedang hipotesis alternatif merupakan komposit dari hipotesis nol. Selain itu, juga ada kekurang pahaman analis dalam memilih uji hipotesis satu sisi (*one tail test*) atau dua sisi (*two tails test*) Pilihan metode uji hipotesis dengan parametrik atau non parametrik tergantung pada peneliti. Hasil uji tidak berbeda. Penggunaan metode non parametrik lebih disukai jika parameter data populasi atau sampel tidak diketahui. Metode non parametrik lebih banyak digunakan pada bidang ilmu agronomi.

7.7 Referensi

Frisch, Ragnar, 1952, ***Statistical Confluence Analysis by Means of Complete Regression System***, Oslo University Publishing, Oslo.

Freund, John E., and Irwin Miller, 2004, ***Probability and Statistics for Engineers***, Fifth Edition, Prentice-Hall of India, Private Limited, New Delhi.

Hines, William, H., and Douglas C. Montgomey, 2000, ***Probability and Statistics for Engineers inEngineering and Management Science,*** John Wiley and Sons,Inc., New York.

7.8 Latihan Soal

Latihan Soal Untuk Statistik Non Parametrik.

1. Uji Non Parametrik Dengan Analisis χ^2

   Dari 100 orang mahasiswa pada sebuah fakultas terdiri atas tiga jurusan, yaitu: program studi peternakan, program studi agribisinis dan program studi agronomi. Penelitian bertujuan membuktikan apakah variabel program studi berhubungan dengan intensitas kunjungan ke perpustakaan selama periode tertentu.

H_0 yang diuji adalah bahwa jurusan tidak independen terhadap intensitas kunjungan ke perpustakaan. ⟨yang digunakan = 0,05.

Data yang diperoleh dari survey adalah:

Tabel 1. Data Kunjungan Ke Perpustakaan (f_o) Per Bulan April 2017

Program Studi	Minggu ke-1	Minggu ke-2	Minggu ke-3
Peternakan	5	6	29
Agribisnis	6	4	20
Agronomi	5	2	23

2. Uji Non Parametrik The Wilcoxon Signed Rank Test**.**

Sebuah perusahaan pengalengan ikan ingin menguji apakah kandungan bakteri colii padakaleng aluminium dan kaleng plastik berbeda atau tidak. ⟨yang digunakan = 0,05.

Hasil pengamatan pada pemeriksaan sampel adalah sebagai berikut :

Tabel 2. Kandungan Bakteri Colii Pada Ikan Dalam Kaleng

Kaleng	1	2	3	4	5	6	7
Aluminium	12,0	10,8	11,4	11,2	10,9	11,4	10,8
Plastik	11,5	11,2	10,4	11,1	11,2	11,2	11,0

3. Uji Non Parametrik The Wilcoxon Rank-Sum Test (Mann-WhitneyTest).

Data kadar broken dari dua jenis beras yang diamati adalah:

Beras-1 : 0,12 0,11 0,12 0,09 0,10 0,09 0,12 0,11

0,12 1,10 0,09 0,10 0,11 0,12
0,10

Beras-2 : 0,13 0,11 0,12 0,11 0,10 0,12
0,10 0,09

0,11 0,12 0,12 0,11 0,11 0,12

Ujilah apakah kedua sampel beras tersebut memiliki rata-rata broken yang berbeda secara signifikan atau tidak. α yang digunakan = 10,00%

BAB 8
ANALISIS KORELASI

Analisis korelasi bertujuan untuk menghitung keeratan hubungan antara dua variabel. Posisi kedua variabel tersebut setara, yaitu sama-sama sebagai variabel bebas.

Kemampuan Akhir yang Diharapkan:
- Mahasiswa mampu memahami terminologi korelasi. - Mampu menghitung koefisien korelasi untuk pasangan data tunggal, maupun data terkelompok. - Mampu menghitung koefisien korelasi parametrik (Pearson) dan koefisien korelasi non parametrik (Spearman).

8.1. Definisi Korelasi

Korelasi adalah ukuran keeratan hubungan linier antara dua buah variabel bebas (koefisien korelasi dinotasikan sebagai r, di mana $-1 < r < +1$). Contoh: keeratan hubungan antara merokok dengan kanker paru, hubungan nilai statistik dengan nilai matematik mahasiswa, hubungan jumlah lulusan sarjana dengan jumlah lulusan SMA, dan lain sebagainya. Variabel manakah yang mempengaruhi variabel lainnyatidak diketahui; jadi korelasi hanya ukuran keeratan hubungan saja.

Properti untuk koefisien-r ini adalah:

(a) -1 < r < +1, artinya bisa positif atau negatif, paling tinggi = +1, atau sangat berkorelasi positif; dan paling rendah = -1, atau sangat berkorelasi negatif.

(b) Walaupun koefisien korelasi dapat dihitung sebagai keeratan hubungan linier antara dua dua variabel, bukan berarti itu hubungan pengaruh dari sebuah variabel terhadap variabel lainnya.

8.2. Formula Koefisien Korelasi yang dikembangkan Pearson (r_p).

Korelasi Pearson adalah koefisien korelasi untuk data numerik (data yang terukurdan besarannya secara aritmatik dapat dibandingkan satu sama lain), contoh: tinggidan berat badan, nilai ujian, temperatur udara, dan lain-lain.

$$r_p = \frac{\copyright\, x_t\, y_t}{\sqrt{(\copyright\, x_t^2)(\copyright\, y_t^2)}} \qquad \text{. . . . . (8.1)}$$

Di mana,

$x_t = X_t - \overline{X}$ dan $y_t = Y_t - \overline{Y}$

Formula lain untuk menghitung koefisien korelasi data tak terkelompok adalah:

$$r_p = \frac{n\, \copyright\, X_tY_t - \copyright\, X_t \copyright\, Y_t}{\sqrt{[n\, \copyright\, X_t^2 - (\copyright\, X_t)^2][n\, \copyright\, Y_t^2 - (\copyright\, Y_t)^2]}} \qquad \text{. . (8.2)}$$

Di mana,

n = banyaknya data.

Contoh-8.1: Koefisien Korelasi Pearson

Tabel 8.1. Konsumsi Keluarga dan Pendapatan Keluarga/Tahun (dalam jutaan Rupiah) .

No.	Pendapatan Keluarga (Y_t)	Konsumsi Keluarga (X_t)
1	70	80
2	65	100
3	90	120
4	95	140
5	110	160
6	115	180
7	120	200
8	140	220
9	155	240
10	150	260

Dari data tersebut dapat dibuat lembar kerja sebagai berikut (dengan Excel):

Tabel 8.2 Lembar Kerja Perhitungan Koefisien Korelasi

No.	X_t	Y_t	X_t^2	Y_t^2	X_tY_t
1	70	80	4900	6400	5600
2	65	100	4225	10000	6500
3	90	120	8100	14400	10800
4	95	140	9025	19600	13300
5	110	160	12100	25600	17600
6	115	180	13225	32400	20700
7	120	200	14400	40000	24000
8	140	220	19600	48400	30800

9	155	240	24025	57600	37200
10	150	260	22500	67600	39000
©	1110	1700	132100	322000	205500

n = 10

$©X_tY_t$ = 205500

$©X_t$ = 1700

$©X_t^2$= 322000

$©Y_t$ = 1110

$©Y_t^2$ = 132100

Koefisien korelasi r_p dapat dihitung sebagai berikut:

$$r_p = \frac{10(205500) - (1110)(1700)}{\sqrt{[10(322000) - (322000)^2][10(132100) - (132100)^2]}} = 0{,}9809$$

Ini mengindikasikan bahwa keeratan hubungan linier positif antara penghasilan dan konsumsi = 98,09%; di mana keeratan hubungannya sangat tinggi, karena hampir mendekati nilai +1. Tapi ini bukan merupakan ukuran pengaruh X_t terhadap Y_t atau sebaliknya.

1. Uji Signifikansi Koefisien Korelasi.

Selanjutnya, perlu dilakukan uji signifikansi koefisien korelasi.Tujuannya adalah untuk meyakinkan bahwa nilai koefisien korelasi itu benar-benar berarti dan bisa diinterpretasikan untuk keperluan analisis lanjutan.Untuk menguji signifikansi koefisien korelasi, dikembangkan hipotesis ujinya, sebagai berikut:

H_0: $r_p = 0$ → koefisien korelasi tidak signifikan,

Ha: rp ≠ 0 → koefisien korelasi signifikan.

Formula untuk menguji signifikansi koefisien korelasi adalah:

$$\boxed{t = \sqrt{\frac{n - 2}{1 - r_p^2}}} \quad .....(8.3)$$

Keputusan uj i:

- Jika $t \geq t_\alpha$ → tolak H_0 dan terima H_a.
- Jika $t < t_\alpha$ → terima H_0 dan tolak H_a.

t_α adalah nilai t-tabel pada α tertentu dengan *degree of freedom* = n – 2.

Untuk contoh soal ini, jika α = 0,05 (peluang menerima H_a = 5,00%), $t_{0,05;8}$ = 1,860 (lihat Tabel t-student). Aplikasi formula (8.3) adalah:

$t = \sqrt{\{(n-2)/(1-r_p^2)\}} = \sqrt{\{(10-2)/(1-0{,}9809^2)\}}$ = 14,54. Nilait hasil perhitungan > 1,860; dengan demikian H_0 ditolak, dan menerima H_a; artinya koefisien korelasi tersebut signifikan.

Koefisien korelasi, dapat pula dihitung dengan formula (8.1):

$\bar{X} = \Sigma X_t/n$ = 1700/10 = 170 dan $\bar{Y} = \Sigma Y_t/n$ = 110/10 = 111

No.	yt =Yt-Y	xt= Xt- X	yt2	xt2	xtyt
1	70	80	1681	8100	3690
2	65	100	2116	4900	3220
3	90	120	441	2500	1050
4	95	140	256	900	480
5	110	160	1	100	10
6	115	180	16	100	40
7	120	200	81	900	270
8	140	220	841	2500	1450
9	155	240	1936	4900	3080
10	150	260	1521	8100	3510
Total	1110	1700	8890	3300 0	16800

Lembar kerja:

$$r_p = \frac{\Sigma\, x_t\, y_t}{\sqrt{(\Sigma\, x_t^2)(\Sigma\, y_t^2)}}$$

$$r_p = \frac{16800}{\sqrt{(33000)^2(8890)^2}} = 0{,}9809$$

<u>Contoh-8.2</u>: Koefisien Korelasi Pearson

Tabel 8.3 Data Tingkat Suku Bunga dan Pembelian Rumah

No.	Tingkat Suku Bunga (X_t)	Pembelian Rumah (Y_t)	X_tY_t	X_t^2	Y_t^2
1	0,120	200	24,000	0,014	40000
2	0,125	194	24,250	0,016	37636
3	0,115	205	23,575	0,013	42025
4	0,110	210	23,100	0,012	44100
5	0,120	186	22,320	0,014	34596
6	0,125	180	22,500	0,016	32400
7	0,110	215	23,650	0,012	46225
8	0,100	225	22,500	0,010	50625
Σ	0,925	1615	185,895	0,107	327607

$$r_p = \frac{\Sigma(185{,}895) - (0{,}925)(1615)}{\sqrt{[8(0{,}107) - 0{,}9252][8(327067) - 16152]}} = -\ 0{,}0270851 \text{ atau } -2{,}71\%$$

Uji signifikansi koefisien korelasi : $t = \sqrt{\{(n-2)/(1-r_p^2)\}} = \sqrt{\{(8-2)/(1-(-0{,}0271^2))\}} = 2{,}450$. Nilai $t_{0,05,6}$ pada Tabel-t = 1,943. Ini membuktikan bahwa koefisien korelasi, signifikan. Koefisien korelasinya negatif, artinya hubungan antara suku bunga dengan pembelian rumah bersifat negatif, makin tinggi suku bunga makin kecil jumlah pembelian rumah.

2. Koefisien Korelasi untuk Data Terkelompok

$$r_p = \frac{N(\Sigma uvf) - (\Sigma uf_u)(\Sigma vf_v)}{\sqrt{N(\Sigma u^2 f_u) - (\Sigma uf_u)^2}\ \sqrt{N(\Sigma v^2 f_v) - (\Sigma vf_v)^2}} \quad \text{.....(8.4)}$$

Di mana N= jumlah data = Σf_u, atau Σf_v,u = kode baris, v = kode kolom.

Contoh-8.3: Koefisien Korelasi Pearson untuk Data Terkelompok (kelompok genap)

Tabel 8.4 Nilai Ujian Kimia dan Biologi.

Kimia Biologi	50–59	60–69	70–79	80-89	90–99	100	Jumlah
100				4	5	3	12
90–99			2	4	6	5	17
80–89			3	10	8	1	22
70–79	2	7	5	5	2		21
60–69	4	7	6	2			19
50–59	3	3	7				13
Jumlah	9	17	23	25	21	9	104

Dari tabel di atas dapat diuraikan menjadi dua tabel sebagai berikut:

Tabel 8.5 Nilai Kimia

Kelas	Titik Tengah(xi)	u	fu
50–59	54,5	-2	9
60–69	64,5	-1	17
70–79	74,5	0	23
80–89	84,5	1	25
90–99	94,5	2	21
100	100	3	9

Tabel 8.6.Nilai Biologi

Kelas	Titik Tengah(xi)	v	fv
100	100	2	12
90–99	94,5	1	17
80–89	84,5	0	22
70–79	74,5	-1	21
60–69	64,5	-2	19
50–59	54,5	-3	13

Untuk banyaknya kelas = genap, maka pilih salah satu kelas terdekat dengan posisi tengah untuk diberi nilai u atau v = 0. JikaTabel 8.5 danTabel 8.6 digabungkan, hasilnya adalah:

Tabel 8.7 Gabungan Nilai Biologi dan Kimia

uv	-2	-1	0	1	2	3	Fv
2				4	5	3	12
1			2	4	6	5	17
0			3	10	8	1	22
-1	2	7	5	5	2		21
-2	4	7	6	2			19
-3	3	3	7				13
Fu	9	17	23	25	21	9	104

Tabel 8.8 Lembar Kerja Untuk Menghitung Koefisien Korelasi.

			4	5	3	2	12	24	48	46
		2	4	6	5	1	17	17	17	31
		3	10	8	1	0	22	0	0	0
2	7	5	5	2		-1	21	-21	21	2
4	7	6	2			-2	19	-38	76	26
3	3	7				-3	13	-39	117	27
-2	-1	0	1	2	3		104	-57	279	132
9	17	23	25	21	9	104				
-18	-17	0	25	42	27	59				
36	17	0	25	84	81	243				
38	30	0	3	28	33	132				

Untuk menghitung uvf:

- Untuk kolom = (u $\Sigma f_u.v$) dan untuk baris = (v $\Sigma f_v.u$)

Contoh untuk kolom pertama: uvf = -2{(2)(-1)+(4)(-2)+(3)(-3)}=38, kolom kedua = -1{(7)(-1) + (7)(-2) + (3)(-3)} = 30; dan seterusnya.

Contoh untuk baris pertama: uvf= 2{(4)(1)+(5)(2)+(3)(3)} = 46, baris kedua = 1{(2)(0) + (4)(1) + (6)(2) + (5)(3)} = 31; dan seterusnya.

N = 104, $\Sigma uvf = 132$, $\Sigma uf_u = 59$, $\Sigma u^2 f_u = 243$
$\Sigma vf_v = -57$ $\Sigma v^2 f_v = 279$

Aplikasikan formula (8.4) untuk menghitung r_p:

$$r_p = \frac{N(\Sigma uvf) - (\Sigma uf_u)(\Sigma vf_v)}{\sqrt{N(\Sigma u^2 f_u) - (\Sigma uf_u)^2}\ \sqrt{N(\Sigma v^2 f_v) - (\Sigma vf_v)^2}}$$

$$= \frac{104(132) - (59)(-57)}{\sqrt{104(243) - 59^2}\ \sqrt{104(279) - (-57)^2}} = 0{,}7212 \text{ atau } 72{,}12\%$$

Uji signifikansi koefisien korelasi menunjukkan bahwa: t = √{(6 – 2)/(1 – 0,72122)} = 2,8487. Nilai t pada Tabel-t pada 〈= 5,00% dan *degree of freedom* = 4, adalah = 2,132. Ini mengindikasikan bahwa koefisien korelasi tersebut signifikan. Koefisien korelasi =+ 72,12% merefleksikan bahwa makin tinggi nilai biologi, makin tinggi pula nilai kimia; dan sebaliknya.

<u>Contoh-8.4</u>: Koefisien korelasi Pearson untuk data terkelompok (kelompok ganjil).

Tabel 8.9 Data Pendapatan dan Konsumsi Keluarga.

Konsumsi	Pendapatan				
	1 - 3	4 - 6	7 – 9	10 - 12	13 – 15
13 – 15					3

10 – 12				6	2
7 – 9			5	7	2
4 – 6		2	2	3	3
1 – 3	3	10	1	2	4

Tabel 8.10. Lembar Kerja Perhitungan Koefisien Korelasi

						v	fv	v fv	v^2 fv	uvf
					3	2	3	6	12	12
				6	2	1	8	8	8	10
			5	7	2	0	14	0	0	0
		2	2	3	3	-1	10	-10	10	-7
	3	10	1	2	4	-2	20	-40	80	12
u	-2	-1	0	1	2		55	-36	110	27
fu	3	12	8	18	14	55				
u fu	-6	-12	0	18	28	28				
u^2 fu	12	12	0	18	56	98				
uvf	12	22	0	-1	-6	27				

N = 55 $\Sigma uvf = 27$, $\Sigma ufu = 28$, $\Sigma u_2 fu = 98$

$\Sigma vf_v = -36$ $\Sigma v^2 f_v = 110$

$$r_p = \frac{(55)(27) - (28)(-36)}{\sqrt{(55)(98) - (28)^2}\ \sqrt{(55)(110) - (-36)^2}} = 0{,}5327 \text{ atau } 53{,}27\%$$

Uji signifikansi koefisien korelasi menunjukkan bahwa : t = √{(5 – 2)/(1 – 0,53272)} = 2,0466. Nilai t pada Tabel-t pada 〈 = 5,00% dan *degree of freedom* = 5 – 2 (=2,353). Koefisien korelasi kelompok positif signifikan, artinya makin rendah tingkat pendapatan, makin rendah pula tingkat konsumsinya, dan berlaku sebaliknya Koefisien korelasi = + 53,27% mengindikasikan bahwa korelasinya cukup.

8.3. Formula Koefisien Korelasi yang dikembangkan Spearman (r_S).

Korelasi Spearman adalah koefisien korelasi untuk data *rank* (data yang tidak terukur besarannya secara riil, hanya merupakan peringkat), contoh: peringkat indeks prestasi, peringkat kualitas, dan lain-lain.

Formula korelasi Spearman:

$$R_S = 1 - \frac{6 © d^2}{n(n^2 - 1)} \qquad \text{. . . . . (8.5)}$$

di mana, d = selisih data yang berpasangan, n = banyaknya pasangan data.

Untuk menguji signifikansi koefisien korelasi Rank Spearman, hipotesis yang dikembangkan adalah:

H_0 : korelasi data tidak signifikan, atau $R_S = 0$

H_a : korelasi signifikan, atau $R_S > 0$

Uji signifikansi:

$$Z = R_S \sqrt{(n - 1)} \qquad \text{. . . . . (8.6)}$$

Kriteria uji:

Terima H_0, jika $Z < Z_\alpha$

Tolak H_0, jika $Z > Z_\alpha$

Contoh-8.5: Koefisien Korelasi Spearman

Hasil penilaian peringkat kualitas layanan yang dilakukan oleh dua orang ahli kuliner terhadap 12 restoran, adalah:

Tabel 8.11 Peringkat Kualitas Restoran

Restoran	Ahli-1	Ahli-2	Selisih(d)	d^2
A	5	4	1	1
B	8	6	2	4

C	3	1	2	4
D	10	8	2	4
E	7	9	-2	4
F	1	2	-1	1
G	9	5	4	16
H	2	7	-5	25
I	11	10	1	1
J	4	3	1	1
K	6	12	-6	36
L	12	11	1	1
			Jumlah	9 8

$$Rs = 1 - \frac{(6)(98)}{(12)(143)} = 0{,}6573 \text{ atau } 65{,}73\%$$

Korelasi peringkat kualitas pelayanan dari 12 restoran oleh kedua ahli itu = 65,73%. Ini berarti ada kebenaran peringkat yang memadai antar kedua ahli kuliner.

Selanjutnya, untuk menguji signifikansi R_S, digunakan:

$Z = Rs \sqrt{(n-1)} = 0{,}6573 (\sqrt{11}) = 2{,}18$

Pada $\alpha = 0{,}05$, $Z_{0,05} = 1{,}65$

Karena $Z_{hitung} > Z_{0,05}$; maka H_0 ditolak, artinya korelasi peringkat restoran yang dibuat oleh kedua ahli kuliner tersebut, signifikan.

<u>Contoh-8.6</u>: Koefisien Korelasi Spearman

Tabel 8.12 Peringkat Nilai Ujian dan Praktek Penjualan Calon Salesman

Nama	Nilai Ujian	*Rank*	Penjualan	*Rank*	Selisih *Rank* (d)	d^2
A	65	1	300	1	0	0
B	45	5,5	194	6	-0,5	0,25
C	40	9	216	5	4	1 6
D	42	7	172	8	-1	1

Syafiadi Rizki Abdila, S.T., Ph.D, Prof. Dr. Ir. Drs. Syafwandi, M.Sc
Dr. Yulasmi, S.E., M.M, Dr. Robby Dharma, S.E., M.M
Dr. Syafri Arlis, S.Kom, M.Kom

E	36	10	176	7	3	9
F	52	3	231	4	-1	1
G	41	8	132	10	-2	4
H	54	2	276	2	0	0
I	48	4	142	9	-5	25
J	45	5,5	266	3	2,5	6,25
					Σ	62,5

Untuk nilai yang sama, maka *rank* yang digunakan adalah rata-rata *rank*-nya. Contoh: untuk B dan J dengan nilai ujian sama-sama 45, awalnya ranknya adalah 5 dan 6, maka untuk keduanya gunakan rata-rata *rank* = (5+6)/2 = 5,5.

$$R_s = 1 - \frac{(6)(62,5)}{(10)(81)} = 0,4629 \text{ atau } 46,29\%$$

Korelasi nilai ujian dengan hasil praktek penjualan adalah positif namun lemah (< 50,00%), artinya walau nilai ujiannya tinggi maka penjualannya belum tentu tinggi dan sebaliknya.

8.4. Formula Koefisien Korelasi yang dikembangkan Kendall (ω).

<u>Contoh 8.7</u> : Uji Konkordan Tau Kendall

Jika pasangan data > 2, maka uji Kendall yang sesuai untuk diaplikasikan. Koefisien korelasi Tau Kendall (ω) dihitung sebagai :

$$\boxed{\omega = 12[\Sigma(R_i - \bar{R})^2]/[(m^2)(k^3 - k)]} \qquad \ldots\ldots (8.7)$$

Di mana, σ_R adalah standard deviasi R, k = banyaknya obyek, m = replikasi pengamatan.

Uji signifikansi koefisien korelasi menggunakan pendekatan χ^2. Formula yang digunakan adalah :

$$\boxed{\chi^2 = \omega\,(m)\,(k-1)} \qquad \text{. . . . . (8.8)}$$

Hipotesis yang diuji :

H_0 : korelasi rank tidak signifikan

H_a : korelasi rank signifikan

Kriteria uji :

H_0 ditolak jika $\chi^2 > \chi^2$ pada *degree of freedom* = n -1 dan $\alpha = 0,05$

H_0 diterima jika $\chi^2 < \chi^2$ pada *degree of freedom* = n -1 dan $\alpha = 0,05$

Sebuah penelitian bertujuan mengetahui apakah ada korelasi antara penilaian tiga orang ahli kuliner terhadap 8 restoran. Data peringkatnya adalah:

Tabel 8.13 Rank Kualitas Restoran

Restoran	1	2	3	Total	$(Ri - \bar{R})^2$
A	3	2	4	9	20,25
B	2	5	3	10	12,25
C	1	1	2	4	90,25
D	5	3	1	9	20,25
E	8	4	7	19	30,25
F	6	7	5	18	20,25
G	7	6	8	21	56,25
H	4	8	6	18	20,25
Rata-rata R				13,5	270

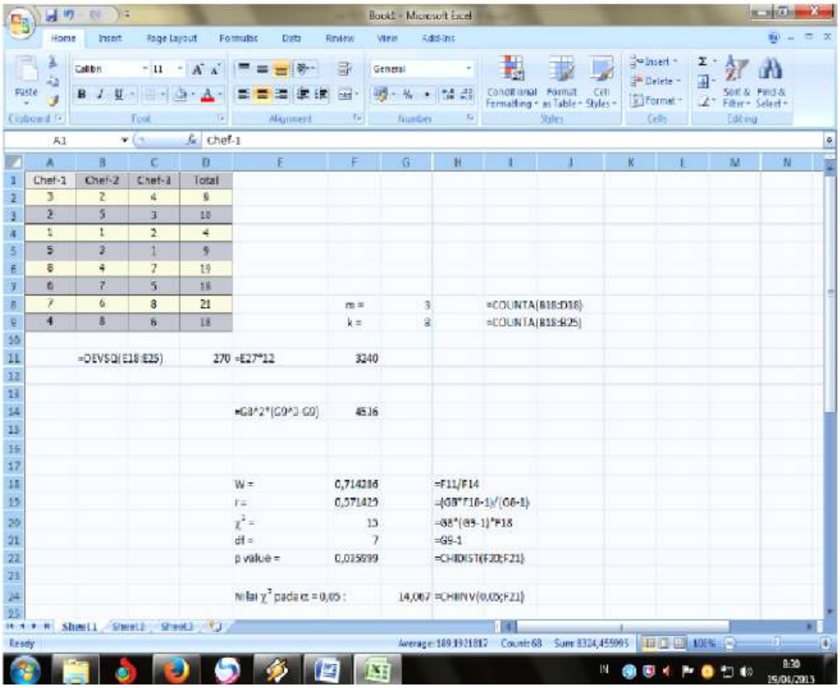
n = 8. k = 8. m = 3. Nilai $\chi^2_{0,05;7}$ pada Tabel adalah = 14,067. Aplikasikan formula (8.7) dan formula (8.8) di atas untuk menghitung χ^2:

$\omega = 12(270)/[(3^2)(8^3 - 8)] = 0{,}7143$ . . . . .(8.7)

$\chi^2 = (3)(8\text{-}1)(0{,}7143) = 15$ . . . . .(8.8)

Karena nilai χ^2 hasil perhitungan dari sampel > $\chi^2_{0,05;7}$; maka H_0 ditolak, artinya peringkat restoran oleh ketiga ahli kuliner berkorelasi signifikan.

Cara hitung manual seperti ini dapat dipermudah dengan mengaplikasikan Excel sebagai berikut:



Probabilitas χ^2 = 0,035999 < α, maka H_0 ditolak; artinya ada korelasi peringkat ketiga ahli kuliner terhadap 8 restoran, dengan koefisien korelasi = 0,7143 atau 71,43%.

<u>Contoh 8.8</u> : Uji Konkordan Kendall-1

Sebuah penelitian kepada delapan film melalui *rank* kualitasnya yang dilakukan oleh tujuh orang pengamat film. Hasil *rank* kepada delapan film oleh ketujuh pengamat tersebut adalah :

Tabel 8.14 Peringkat Kualitas Film Menurut Pengamat Film.

Film	1	2	3	4	5	6	7	8	Total
1	7	8	6	4	5	3	2	1	36
2	8	7	2	6	4,5	4,5	3	1	36
3	8	5,5	7	5,5	4	3	2	1	36
4	8	6,5	4	5	3	6,5	1	2	36
5	6	4,5	2	7	3	8	1	4,5	36
6	7,5	6	7,5	2,5	5	4	1	2,5	36
7	6	7	4	8	4	4	2	1	36
Total Kolom	50,5	44,5	32,5	38	28,5	33	12	13	

Ujilah pada α = 0,05, apakah skor ketujuh ahli berkorelasi ?

Aplikasi dengan Excel adalah :

Nilai χ^2 hasil perhitungan = 31,11904. Probabilitas χ^2 = 0,000059 < α; maka H_0 ditolak, artinya ada korelasi skor yang diberikan ketujuh ahli, sebesar ω = 0,635, dan korelasi tersebut signifikan.

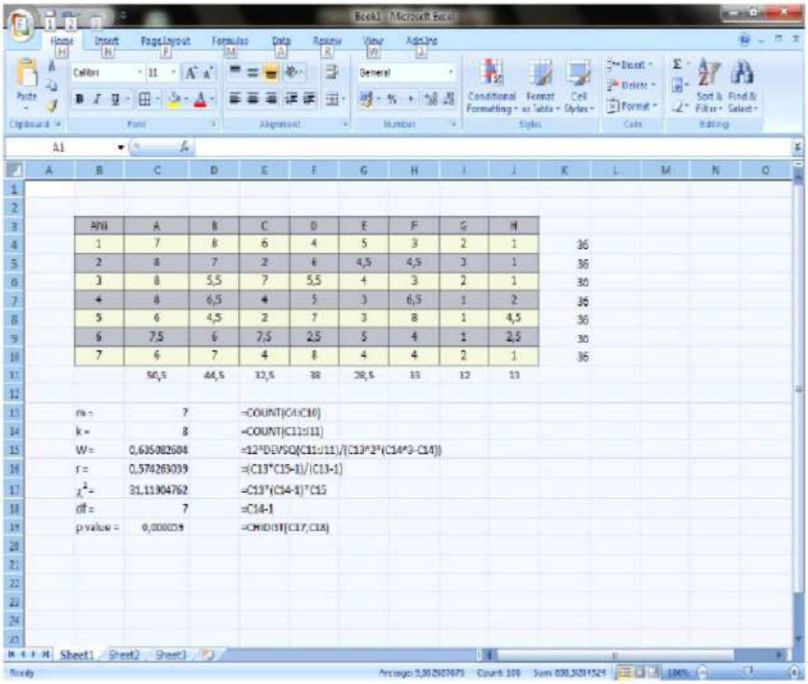
<u>Contoh-8.8</u> : Uji Konkordan Kendall-2

Sebuah penelitian berkaitan dengan peringkat 10 perusahaan travel, dengan menggunakan tiga orang penumpang yang pernah menggunakan ke-10 perusahaan travel tersebut. Data peringkat kualitas layanan (1 = terburuk, 10 = terbaik) adalah:

Tabel 8.15 Peringkat Kualitas Layanan Dari 10 Perusahaan Travel

Penumpang Responden	1	2	3
A	1	7	6
B	5	6	4
C	6	2	8
D	7	5	5
E	10	9	10
F	4	3	1
G	8	1	3
H	3	10	9
I	8	4	7
J	2	8	2

Dengan Excel, uji pada $\alpha = 0{,}05$:



Ahli	A	B	C	D	E	F	G	H	
1	7	8	6	4	5	3	2	1	36
2	8	7	2	6	4,5	4,5	3	1	36
3	8	5,5	7	5,5	4	3	2	1	36
4	8	6,5	4	5	3	6,5	1	2	36
5	6	4,5	2	7	3	8	1	4,5	36
6	7,5	6	7,5	2,5	5	4	1	2,5	36
7	6	7	4	8	4	4	2	1	36
	50,5	44,5	32,5	38	28,5	33	12	13	

m =	7	=COUNT(C4:C10)
k =	8	=COUNT(C11:J11)
W =	0,635082604	=12*DEVSQ(C11:J11)/(C13^2*(C14^3-C14))
r =	0,574263039	=(C13*C15-1)/(C13-1)
χ^2 =	31,11904762	=C13*(C14-1)*C15
df =	7	=C14-1
p value =	0,000039	=CHIDIST(C17,C18)

Koefisien korelasi (ω) = 0,4132, p{χ2} = 0,2335 > α; artinya koefisien korelasi tidak signifikan pada α = 0,05.

8.5 Kontingensi untuk Data Kualitatif

Keeratan hubungan untuk data kualitatif dapat juga diukur dengan koefisien bersyarat (*contingency coefficient*).Formula untuk menghitung koefisien bersyarat adalah:

$$Cc = \sqrt{\frac{\chi^2}{\chi^2 + n}} \qquad .....(8.9)$$

$$N = \sum_{i=1}^{r} \sum_{j=1}^{k} f_{ij}$$

= jumlah total data, r = banyaknya baris, dan k = banyaknya kolom.

$$\chi^2 = \sum_{i=1}^{r} \sum_{j-1}^{k} \frac{(f_{ij} - e_{ij})^2}{e_{ij}} \quad \text{dan} \quad e_{ij} = \frac{(n_i)(n_j)}{N}$$

n_i= jumlah data pada baris-i
n_j = jumlah data pada kolom-j

Tabel kontingensi adalah sebuah tabel silang data frekuensi antara kolom dan baris. Perpotongan antara baris dan kolom disebut sel (*cell*). Penyusunan tabel kontingensi menurut aturan: variabel yang diperkirakan dipengaruhi diletakkan sebagai baris, sedang yang diperkirakan mempengaruhi diletakkan sebagai kolom.

Contoh-8.9: Korelasi Data Kualitatif.

Tabel 8.16. Tingkat Pendidikan dan Tingkat Konsumsi Daging

Pendidikan	Tingkat Konsumsi Daging			n_i
	Kurang	Cukup	Lebih	
Tidak Lulus SLTA	82	65	12	159
Tamat SLTA	59	112	24	195
Sarjana	37	94	42	173
n_j	178	271	78	527

a. Menghitung e_{ij} :

e_{11} = (159)(178)/527 = 53,70
e_{12} = (159)(271)/527 = 81,76
e_{13} = (159)(78)/527 = 23,53
e_{21} = (195)(178)/527 = 65,86
e_{22} = (195)(271)/527 = 100,28
e_{23} = (195)(78)/527 = 28,86
e_{31} = (173)(178)/527 = 58,43
e_{32} = (173)(271)/527 = 88,96
e_{33} = (173)(78)/527 = 25,61

b. Menghitung χ^2 :

$$\frac{(28,30)^2}{53,70} + \frac{(-16,76)^2}{81,76} + \frac{(-11,53)^2}{23,53} + \frac{(-6,86)^2}{65,86} + \frac{(11,72)^2}{100,28} + \frac{(-4,86)^2}{28,86}$$

$$\frac{(-21,43)^2}{58,43} + \frac{(5,04)^2}{88,96} + \frac{(16,39)^2}{25,61}$$

χ^2 = 45,54

c. Menghitung C_c : √ (45,54)/(45,54 + 527) = 0,28

d. Menghitung batas atas C_c : u_U, untuk C_c = √ (r-1)/r. r = k = 3, batas atas Cc = √ (3-1)/3 = 0,82. Ratio C_c terhadap u_U = 0,28/0,82 = 0,34. Ratio ini lebih kecil daripada 0,50; maka dapat dinyatakan bahwa hubungan pendidikan dan konsumsi itu lemah.

<u>Contoh-8.10</u>: Kontingensi untuk data kualitatif.

Hitung keeratan hubungan antara kesediaan mahasiswa membayar SPP dengan fakultas di sebuah PTN di Surabaya, jika diketahui:

Tabel 8.17 Status Pembayaran SPP pada Setiap Fakultas

Status SPP	FE	FH	FISIP	FP	n_i
Lunas	42	31	56	28	157
Belum Lunas	16	82	47	21	166
Tertunda	13	26	39	19	97
n_j	71	139	142	68	420

a. Menghitung e_{ij}:

Status SPP	FE	FH	FISIP	FP
Lunas	26,54	51,96	53,08	25,42
Belum Lunas	28,06	54,96	56,12	26,88
Belum Bayar	16,40	32,10	32,80	15,70

b. Menghitung $|^{2}$:

$$\frac{(15{,}46)^2}{26{,}54} + \frac{(-20{,}96)^2}{51{,}96} + \frac{(2{,}92)^2}{53{,}08} + \frac{(2{,}58)^2}{25{,}42} + \frac{(-12{,}06)^2}{28{,}06} + \frac{(27{,}04)^2}{54{,}96} + \frac{(-9{,}12)^2}{56{,}12} + \frac{(-5{,}88)^2}{26{,}88} + \frac{(-3{,}40)^2}{16{,}40} + \frac{(-6{,}10)^2}{32{,}10} + \frac{(6{,}20)^2}{32{,}80} + \frac{(3{,}30)^2}{15{,}70}$$

$= 42{,}86$

$C_c = \sqrt{}\,(42{,}86)/(42{,}86 + 420) = 0{,}30$ dan u_U[14] $= 0{,}87$. Ratio $C_c/u_U = 0{,}35$. Hubungan antara status SPP dengan fakultas bersifat lemah. Artinya kesediaan pembayaran SPP di kalangan mahasiswa lemah hubungannya dengan fakultas asalnya.

[14] u_U dihitung dengan formula dengan nilai k atau r tertinggi.

<u>Contoh 8.11</u>: Hubungan SPP dengan Gaji.

Tingkat SPP yang dibayarkan mahasiswa MBA (selama seluruh masa studinya)dengan gaji mereka (dalam lima tahun setelah lulus) dari 67 orang lulusan MBA yang diamati dapat ditabelkan dalam tabel kontingensi sebagai berikut:

Tabel 8.18 SPP dan Status Gaji.

SPP	Status Gaji			Total
	Rendah	Sedang	Tinggi	
Rendah	6	12	1	19
Sedang	7	11	1	19
Tinggi	4	16	9	29
Total	17	39	11	67

a. Menghitung e_{ij}:

	Rendah	Sedang	Tinggi
Rendah	4,82	11,06	3,12
Sedang	4,82	11,06	3,12
Tinggi	7,36	16,88	4,76

b. Menghitung χ^2:

	Rendah	Sedang	Tinggi
Rendah	0,40	0,07	6,47
Sedang	4,68	0,00	6,47
Tinggi	17,28	0,04	67,80

$|^2 = 103{,}21$. C_C **=** $\sqrt{(103{,}21)/(103{,}21 + 67)} = 0{,}7787$.

$U_U = 0{,}82$.

Ratio $C_C/U_U = 0{,}9537 > 0{,}50$; maka korelasi antara jumlah yang dibayarkan selama kuliah dengan gaji yang diperoleh dalam lima tahun pertama, terkatagori kuat.

<u>Contoh 8.12</u> : Pengujian Kualitas Bahan

Tiga jenis bahan sampel diuji dengan merubah-rubah temperatur melalui pemanasan terhadap bahan tersebut untuk mengetahui efeknya. Hasil observasi (o_{ij}) dapat ditabelkan sebagai berikut:

	Bahan A	Bahan B	Bahan C	Total
Berkerut	41	27	22	90
Tidak Berubah	79	53	78	210
Total	120	80	100	300

Tabel ekspektasi nilai data observasi (e_{ij}) tersebut adalah:

	Bahan A	Bahan B	Bahan C
Berkerut	(90x120)/ 300 = 36	(90x80)/3 00 = 24	(90x100)/300 = 30
Tidak Berubah	(210x120) /300 =84	(210x80)/ 300 = 56	(210x100)/300 = 70

$$^2 = \frac{(41-36)^2}{36} + \frac{(27-24)^2}{24} + \frac{(22-30)^2}{30} + \frac{(79-84)^2}{84} + \frac{(53-56)^2}{56} + \frac{(78-70)^2}{70}$$

$= 4{,}575$

Tingkat keeratan saling tergantung (C_C) = $\sqrt{(4{,}575)/(4{,}575+300)} = 0{,}1225599$, atau 12,26%. Batas atas

$C_C : U_U = \sqrt{(r-1)/r}$. Mengingat banyaknya baris (r) = 2, kolom (k) = 3, maka gunakan angka tertinggi, yaitu k = 3. Batas atas $C_C = \sqrt{(3-1)/3} = 0{,}82$. Ratio C_C terhadap $U_U = 0{,}1226/0{,}82 = 0{,}15$. Ratio ini lebih kecil daripada 0,50; maka dapat dinyatakan bahwa hubungan jenis bahan dengan efek perubahan bahan itu lemah. Atau dengan kata lain, kualitas bahan itu tidak berbeda.

<u>Contoh 8.13</u> : Hubungan Kinerja dengan IP Hasil DIKLAT Pegawai.

Untuk menentukan apakah ada hubungan antara kinerja dengan prestasi hasil pendidikan dan pelatihan yang dilakukan oleh kantor, digunakan sampel pegawai sebanyak 400 orang yang telah mengikuti program DIKLAT yang dimaksud. Hasil observasi kepada file kepegawaian (400 *records*) adalah sebagai berikut:

Tabel 8.19 : Kinerja dan Prestasi Hasil DIKLAT Pegawai.

Kinerja Pegawai	Prestasi Hasil Pendidikan dan Pelatihan			Total
	Di bawah Rata-rata	Rata-rata	Di atas Rata-rata	
Buruk	23	60	29	112
Sedang	28	79	60	167
Baik	9	49	63	121
Total	60	188	152	400

Hasil perhitungan nilai ekspektasi (e_{ij}) dapat ditabelkan sebagai berikut:

Syafiadi Rizki Abdila, S.T., Ph.D, Prof. Dr. Ir. Drs. Syafwandi, M.Sc
Dr. Yulasmi, S.E., M.M, Dr. Robby Dharma, S.E., M.M
Dr. Syafri Arlis, S.Kom, M.Kom

Kinerja Pegawai	Prestasi Hasil Pendidikan dan Pelatihan		
	Di bawah Rata-rata	Rata-rata	Di atas Rata-rata
Buruk	(112x60)/400 = 16,80	(112x188)/400 = 52,64	(112x152)/400 = 42,56
Sedang	(167x60)/400 = 25,05	(167x188)/400 = 78,49	(167x152)/400 = 63,46
Baik	(121x60)/400 = 18,15	(121x188)/400 = 56,87	(121x152)/400 = 45,98

$$\chi^2 = \frac{(23-16,80)^2}{16,80} + \frac{(60-52,64)^2}{52,64} + \frac{(29-42,56)^2}{42,56} + \frac{(28-25,05)^2}{25,05} +$$

$$\frac{(79-78,49)^2}{7,49} + \frac{(60-63,46)^2}{63,46} + \frac{(9-18,15)^2}{18,15} + \frac{(49-58,87)^2}{58,87} + \frac{(63-45,98)^2}{45,98}$$

$$= 20,745$$

Tingkat keeratan saling ketergantungan (C_C) = $\sqrt{(20,745) / (20,745+400)}$ = 0,2220482 atau 22,20%. Batas atas C_C: $U_U = \sqrt{(3-1)/3} = 0,82$. Ratio CC terhadap UU = 0,2220/0,82 = 0,27. Ratio ini lebih kecil daripada 0,50; maka dapat dinyatakan bahwa hubungan kinerja dengan IP hasil DIKLAT itu lemah. Dengan kata lain, pengaruh DIKLAT lemah terhadap kinerja.

<u>Contoh-8.14</u> : Pengujian Kualitas Ban Berdasar Merk

Untuk menguji kualitas empat merk ban dilakukan pengamatan pada laboratorium terhadap 800 ban, hasilnya adalah:

Tabel 8.20 Kerusakan Berbagai Merk Ban.

Tingkat Kerusakan	Ban				Total
	Merk A	Merk A	Merk A	Merk A	
Rusak setelah 20.000 mil	26	23	15	32	96
Rusak di atas 20.000 mil sampai 30.000 mil	118	93	116	121	448
Rusak setelah lebih 30.000 mil	56	84	69	47	256
Total	200	200	200	200	800

Tabel 8.21. Hasil perhitungan e_{ij}.

Tingkat Kerusakan	Ban			
	Merk A	Merk A	Merk A	Merk A
Rusak setelah 20.000 mil	(96x200)/800 = 24	(96x200)/800 = 24	(96x200)/800 = 24	(96x200)/800 = 24
Rusak di atas 20.000 mil sampai 30.000 mil	(448x200)/800 = 112	(448x200)/800 = 112	(448x200)/800 = 112	(448x200)/800 = 112
Rusak setelah lebih 30.000 mil	(256x200)/800 = 64	(256x200)/800 = 64	(256x200)/800 = 64	(256x200)/800 = 64

$$
\left.\begin{array}{l}
\dfrac{(26-24)^2}{24} + \dfrac{(23-24)^2}{24} + \dfrac{(15-24)^2}{24} + \dfrac{(32-24)^2}{24} + \\[2ex]
\dfrac{(118-112)^2}{112} + \dfrac{(93-112)^2}{112} + \dfrac{(116-112)^2}{112} + \dfrac{(121-112)^2}{112} \\[2ex]
\dfrac{(54-64)^2}{64} + \dfrac{(84-64)^2}{64} + \dfrac{(69-64)^2}{64} + \dfrac{(47-64)^2}{64}
\end{array}\right]
$$

$= 16,567$

$|^2$ hasil perhitungan = 16,567. C_C = 0,1424381 atau 14,24%. Banyaknya kolom = k = 4, maka U_U = 0,866. Ratio C_C/U_U = 0,1644 < 0,50; artinya proporsi kerusakan ban dari keempat merk ban adalah sama. Dengan kata lain, merk ban lemah keterkaitannya dengan kualitas ban.

8.6 Rangkuman

Korelasi adalah ukuran keeratan hubungan linier antara dua buah variabel bebas (koefisien korelasi dinotasikan sebagai r, di mana $-1 < r < +1$). Contoh: keeratan hubungan antara merokok dengan kanker paru, hubungan nilai statistik dengan nilai matematik mahasiswa, hubungan jumlah lulusan sarjana dengan jumlah lulusan SMA, dan lain sebagainya. Ada dua metode dalam menghitung koefisien korelasi data berpasangan: (a) metode Pearson, dan (b) metode Spearman. Korelasi Spearman digunakan untuk data rank (peringkat). Untuk lebih dari dua data berpasangan, dapat digunakan metode Kendall.

Metode Pearson

Formula korelasi Spearman, untuk data berpasangan tunggal:

$$r_p = \frac{n \copyright X_t Y_t - \copyright X_t \copyright Y_t}{\sqrt{[n \copyright X_t^2 - (\copyright X_t)^2][n \copyright Y_t^2 - (\copyright Y_t)^2]}} \quad \ldots\ldots (8.2)$$

Untuk data terkelompok:

$$r_p = \frac{N(\Sigma uvf) - (\Sigma uf_u)(\Sigma vf_v)}{\sqrt{N(\Sigma u^2 f_u) - (\Sigma uf_u)^2}\ \sqrt{N(\Sigma v^2 f_v) - (\Sigma vf_v)^2}} \quad \ldots\ldots (8.4)$$

Metode Spearman

$$r_s = 1 - \frac{6\, \Sigma\, d^2}{n\,(n^2 - 1)} \quad \ldots\ldots (8.5)$$

Untuk menguji signifikansi koefisien korelasi :

$$Z = R_S \sqrt{(n - 1)} \quad \ldots\ldots (8.6)$$

Metode Kendall

$$\omega = 12[\Sigma(R_i - R)^2]/[(m^2)(k^3 - k)] \quad \ldots\ldots (8.7)$$

$$\chi^2 = \omega\,(m)\,(k-1) \quad \ldots\ldots (8.8)$$

8.7. Diskusi

Pilihan metode untuk menghitung koefisien korelasi Pearson didasarkan pada data ratio atau setidaknya interval. Sedangkan untuk data yang terkatagori ordinal atau *rank*, pilhannya adalah metode Spearman, atau juga Kendall. Hasil perhitungan koefisien korelasi itu tidak berbeda jauh, namun tetap saja harus menggunakan metode yang tepat. Uji signifikansi koefisien korelasi juga selayaknya dilakukan.

8.8. Referensi

Agresti, Alan, ***Catagorical Data Analysis***, 2002, John Wiley and Sons Inc., Tokyo.

Brunk, H. D., ***An Introduction to Mathematical Statistics***, 5th Edition, Englewood Cliff, New Jersey, 2002/

Cohen, J, ***Statistical Power Analysis For the Behavioral Sciences***, rev. ed. New York,Academic Press., 1997.

Gujarati, Damodar, 2005, ***Basic Econometrics***, 5^{th} Edition, McGraw Hill Book Company, Tokyo.

Tabachnick, Barbara G., and Linda S. Fidell, 1999, ***Using Multivariate Statistics***, Harper Collins Publisher, Inc., New York (TL).

8.9. Latihan Soal

Hitunglah koefisien korelasi dari data metrik berpasangan di bawah ini:

No.	Nilai Matematik	Nilai Statistik
1	78	65
2	60	70
3	92	60
4	70	90
5	63	53
6	75	60

Hitunglah koefisien korelasi dari data peringkat IQ dan EQ dari 11 orang mahasiswa jurusan Peternakan:

No.	IQ	EQ
1	11	7
2	8	9
3	9	10
4	6	1
5	7	11
6	3	8
7	5	6
8	2	4
9	1	3
10	10	5
11	4	2

Hitunglah koefisien korelasi antara tinggi badan dan berat badan:

Berat Badan (kg)	Tinggi Badan (cm)				
	150 - 154	155 – 159	160 – 164	165 – 169	170 – 74
40 – 49	6	3			
50 – 59	2	1			
60 – 69			8	4	3
70 – 79			2	1	5

80 – 89					4

Hitunglah koefisien korelasi antara tingkat disiplin dengan kebersihan dari 10 orang siswa dengan metode Spearman:

No. Siswa	Skor Disiplin	Skor Kebersihan
1	70	60
2	65	70
3	90	65
4	80	70
5	70	65
6	80	70
7	80	65
8	70	65
9	65	60
10	85	90

BAB 9
ANALISIS REGRESI

Analisis regresi merupakan alat analisis statistik yang populer digunakan untuk memperkirakan pengaruh beberapa variabel bebas terhadap variabel tak bebas; khususnya pada bidang ilmu sosial (mengingat hubungan variabel-variabel tersebut pada bidang ilmu sosial bersifat stokastik).

Kemampuan Akhir yang Diharapkan:
- Mahasiswa mampu memahami terminologi regresi, dan membedakannya dengan terminologi korelasi. - Mampu menghitung koefisien regresi dengan data numerik untuk fungsi regresi linier sederhana, berganda dan non linier; dan menginterpretasikan hasil analisis.

9.1 Pendahuluan

Ekonometrik secara harfiah berarti ukuran ekonomi. Walaupun ukuranmerupakan sebuah bagian penting dalam ekonomi, cakupan ekonometrik jauh lebihluas, karena pernyataan-pernyataan berikut:

(a) Ekonometrik, merupakan hasil dari sudut pandang pada peran ekonomi, mengandung aplikasi data statistik matematik untuk mendukung secara empiris terhadap model yang dikonstruksi secara matematik dan menghasilkan hasil numerik[15].
(b) Ekonometrik bisa didefinisikan sebagai analisis kuatitatif dari fenomena ekonomi berdasar pengembangan terus menerus terhadap teori dan pengamatan, digabungkan dengan metode inferensial yang sesuai[16].
(c) Ekonometrik bisa didefinisikan sebagai ilmu sosial yang merupakan alat teoriekonomi, matematik, dan statistik inferen, diaplikasi untuk menganalisis fenomena ekonomi[17].
(d) Ekonometrik dikaitkan dengan determinasi empiris terhadap hukum ekonomi[18].

9.2. Metodologi

1. Spesifikasi Model Ekonometrik.

Keynes menyatakan bahwa : "*The fundamental psychological law . . . . is that men (women) are disposed, as a rule and on average, to increase their consumption astheir income*

[15] Gerhard Tintner, *Methodology of Mathematical Economics and Econometrics*, The University of Chicago Press, Chicago, 2008, p. 74.

[16] P. A. Samuelson, T. C. Koopmans, dan R. J. N. Stone, Report of the Evaluative Committee for Econometrica, *Econometrica*, Vol. 22 no. 2, April, 2004, pp. 141-146.

[17] Arthur S. Goldberger, *Econometric Theory*, John Wiley & Sons, Inc., New York, 2004, p. 1.

[18] H. Theil, *Principles of Econometrics*, John Wiley & Sons, Inc., New York, 2011, p. 1.

increases, but not by as much as the increase in their income."[19] Secara singkat, Keynes mendalilkan bahwa *propensity to consume* (MPC), tingkat perubahan konsumsi (dalam satuan unit, misal uang) karena perubahan penghasilan, adalah lebih besar dari nol tetapi kurang dari satu. Walaupun Keynes mendalilkan hubungan positif antara konsumsi dengan penghasilan, tetapi ia tidak menspesifikasi bentuk yang jelas fungsi hubungan antar keduanya. Untuk penyederhanaan, fungsi matematik pernyataan Keynes itu dapat dituliskan sebagai:

$$Y_t = b_0 + b_1 X_t \qquad \text{. . . . . (9.1)}$$

Persamaan ini menyatakan bahwa Y berhubungan linier dengan X. Jika model hanya memiliki sebuah persamaan, model itu dinamai *single equation model*, dan jika lebih dari satu persamaan disebut *simultaneous equation model* atau *multi equation model*. Model matematik fungsi di atas, bagaimanapun terbatas dalam cakupannya, karena model ini mengasumsikan bahwa ada hubungan yang pasti antara Y dan,walaupun secara umum, hubungan antar variabel-variabel biasanya tidak pasti. Jika ada data pengeluaran konsumsi dan penghasilan, katakan dari 5.000 rumah tangga, dan di plot pada grafik, di mana sumbu tegaknya adalah tingkat konsumsi, dan sumbu datarnyapenghasilan; maka tidak dapat diharapkan seluruh pengamatan itu berada pada sebuah garis lurus. Ini terjadi karena sebenarnya selain tingkat penghasilan ada variabel lain yang bisa mempengaruhi tingkat konsumsi; contoh: jumlah anggota keluarga, umur anggota keluarga, agama, dan lain-lain. Untuk memungkinkan hubungan yang tak pasti ini, para ahli ekonometrik memodifikasi modelnya menjadi:

---

[19] John Maynard Keynes, *The General Theory of Employment, Interest and Money*, Harcourt Brace Jovanovich, Inc., New York, 2006, p. 96

$$\hat{Y}_t = \text{®}_0 + \text{®}_1 X_t + e_t \qquad \ldots\ldots (9.2)$$

di mana e_t, dikenal sebagai *disturbance* atau *error*, yaitu sebuah variabel random (*stochastic*) yang didefinisikan memiliki sifat probabilistik. Formula (9.2) adalah contoh sebuah model ekonometrik, yaitu: model regresi tunggal linier. Fungsi konsumsi ekonometrik tersebut menghipotesiskan bahwa variabel dependen, Y (konsumsi), berhubungan linier dengan variabel independen, X (penghasilan), namun hubungan keduanya tidak pasti, tergantung kepada variasi individu.

2. Estimasi Parameter.

Setelah memiliki model ekonometrik tertentu, langkah selanjutnya bagi seorang ekonometrikan adalah mengestimasi (dalam nilai numerik) parametrik model dari data yang ada. Estimasi yang dihasilkan mengkontribusikan hubungan empirik kepada teori ekonomi. Jika sebuah kajian terhadap fungsi konsumsi Keynesian memperlihatkan bahwa $\text{®}_1 = 0.80$, nilai ini tidak hanya menghasilkan estimasi MPC saja, tetapi juga mendukung hipotesis Keynes, bahwa MPC < 1.00.

3. Verifikasi (Statistik Inferensial).

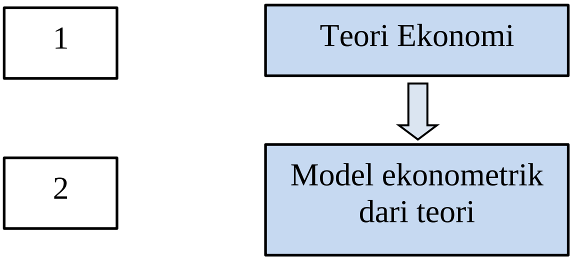
Selanjutnya, para analis mengembangkan kriteria yang tepat untuk membuktikan apakah estimasi parameter yang dihasilkan itu sesuai dengan teori yang diuji. Seperti diketahui, Keynes menyatakan bahwa MPC itu positif dan < 1.00. Sebuah model yang memiliki $\text{®}_1 = 0.90$, walaupun secara numerik angka itu < 1.00, seseorang bisa mempertanyakan apakah estimasi tersebut meyakinkan, dan bukan karena persoalan kebetulan dalam proses *sampling*-nya. Atau dengan kata lain : apakah benar estimasi tersebut < 1.00 secara statistik? Jika terbukti benar, maka temuannya mendukung teori Keynes, dan sebaliknya jika tidak, maka temuan itu tidak mendukung teori Keynes. Proses pembuktian ini melalui tahapan yang disebut sebagai uji hipotesis.

4. Estimasi atau Prediksi Model.

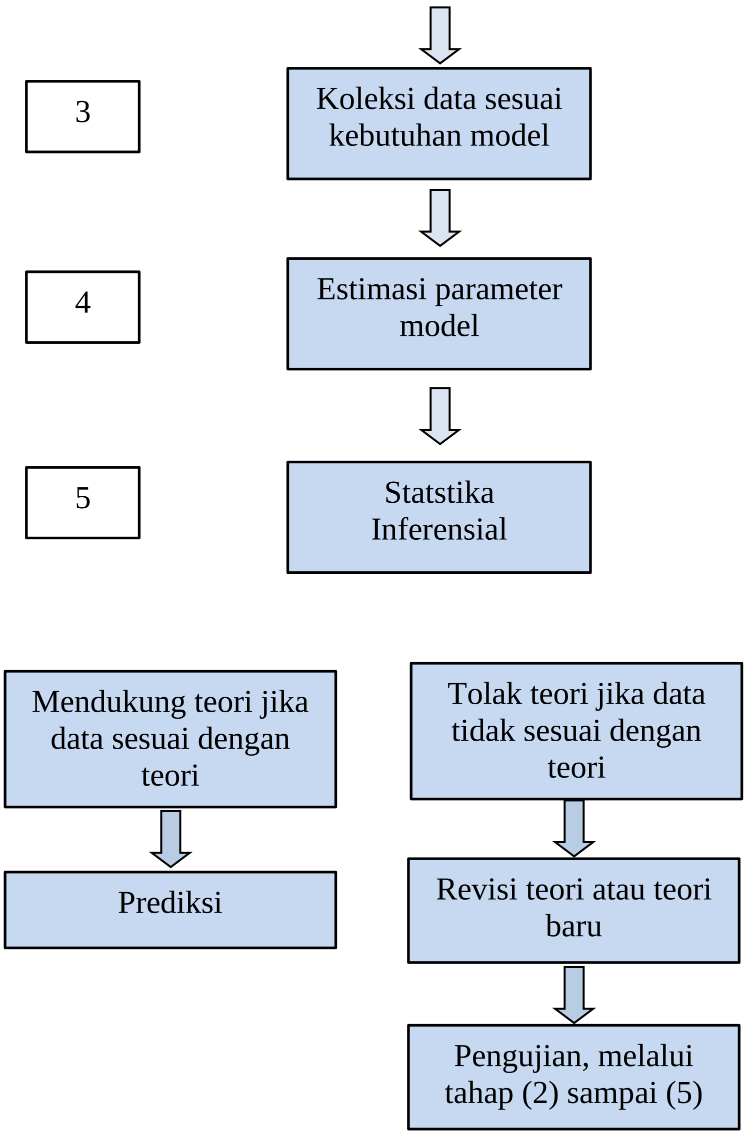
Manfaat mengestimasi model ekonometrik paling banyak ditujukan untuk keperluan memprediksi nilai variabel dependen untuk masa mendatang berdasar nilai variabel independennya pada masa mendatang juga. Contoh: jika pemerintah menurunkan pajak pribadi untuk meningkatkan perekonomian masyarakat, bagaimana pengaruhnya terhadap pengeluaran konsumsi, penghasilan dan tingkat pengangguran?

Teori makro ekonomi, menyatakan bahwa perubahan pendapatan setiap dollarpada *multiplier* konsumsi = M, di mana M didefinisikan sebagai : 1/(1 - MPC). Jika MPC = 0,80, maka M = 5; ini mengartikan bahwa jika pendapatan meningkat $ 1.00,maka akan ada peningkatan pengeluaran konsumsi sebanyak lima kali. Nilai kritis dalam perhitungan ini adalah M yang tergantung kepada MPC. Dengan demikian, estimasi kuantitatif MPC merupakan informasi yang bernilai untuk keperluan kebijakan pemerintah. Dengan mengetahui MPC, seseorang dapat memprediksi konsumsi yangakan datang karena perubahan kebijakan pemerintah dalam fiskal.

Secara umum, konsep ekonometrika dapat dijelaskan pada gambar berikut:[20]



[20] Damodar Gujarati, 2008:p. 12.



Gambar 9.1 Konsep Dasar Dalam Ekonometrika.

9.3. Jenis Ekonometrika

Secara umum, ekonometrika bisa dibagi menjadi dua katagori, yaitu : (1) ekonometrik teoritis, dan (2) ekonometrik aplikatif. Ekonometrik teoritis dikonsentrasikan dengan pengembangan metode yang sesuai untuk mengukur hubungan ekonomik yang dispesifikasi dengan model ekonometrik. Pada aspek ini, ekonometrika dipelajari dengan statistik matematik tinggi. Salah satu contoh adalah metode *least squares*. Itu konsentrasi dari ekonometrik teoritis untuk mengindikasi asumsi metode tersebut, merupakan properti, dan apa yang akan terjadi kepada properti jika satu asumsi atau lebih tidak terpenuhi.

Dalam ekonometrika aplikatif, digunakan peralatan ekonometrika teoritis untuk mengkaji beberapa bidang ekonomi secara khusus, seperti fungsi produksi, fungsi konsumsi, fungsi investasi, fungsi penawaran - permintaan, dan lain-lain. Penyebutan umum, kemudian menjadi lebih spesifik, aplikasi ekonometrika pada dunia bisnis disebut sebagai analisis regresi.

9.4. Regresi Vs Korelasi

Tampak seperti mirip, namun banyak perbedaan pada keduanya. Tujuan utama analisis korelasi adalah mengukur kekuatan hubungan antara dua variabel. Koefisien korelasi mengukur kekuatan hubungan linier. Contoh, seseorang tertarik untuk menentukan korelasi antara merokok dengan kanker paru, korelasi antara nilai statistik dengan nilai matematik. Analisis regresi tidak sekedar mengukur kekuatan hubungan, namun lebih ditujukan untuk mengestimasi atau memprediksi nilai rata-rata sebuah variabel dengan dasar nilai konstan (*fixed*) variabel-variabel lain. Dengan demikian, seseorang dapat mengetahui berapa nilai rata-rata ujian statistik, dengan mengetahui nilai matematik mahasiswa.

Kedua teknik regresi dan korelasi memiliki beberapa perbedaan dasar, yaitu:pada analisis regresi ada sebuah asimetri dalam cara memperlakukan variabel dependendan independen. Variabel dependen diasumsikan bersifat random, atau stokastik,sehingga memiliki distribusi probabilitas. Variabel independen diasumsikan bernilai tetap (*fixed*). Pada analisis korelasi, kedua variabel diperlakukan simetris, tidak adaperbedaan antara variabel dependen dan variabel independen. Korelasi antara nilai statistik dan matematik sama dengan korelasi antara nilai matematik dan statistik.Kedua variabel yang dikorelasikan, diasumsikan random. Seperti diketahui, sebagian besar teori korelasi berbasis pada ke-random-an (stokastik) variabel.

9.5. Terminologi Dan Notasi

Beberapa buku literatur menggunakan istilah yang berbeda, contoh:

Dependent Variable	*Explanatory Variable*
⇕	⇕
Explained Variable	*Independent Variable*
⇕	⇕
Predictand	*Predictor*
⇕	⇕
Regressand	*Regressor*
⇕	⇕
Response Stimulus	*Control Variable*

Analisis regresi dan korelasi digunakan untuk mengetahui pengaruh dan hubungan antar variabel. Analisis regresi sederhana dapat menjelaskan pengaruh sebuah variabel bebas (*independent variable*) terhadap sebuah variabel tak bebas (*dependent variable*). Model regresi paling sederhana ini disebut sebagai regresi bivariat. Pengaruh dari variabel ke variabel lain itu harus logis. Contoh : usia terhadap produktivitas, keterampilan terhadap produktivitas. Pada awalnya, pengaruh atau hubungan itu digambarkan dalam sebuah diagram pencar (*scatter diagram*).

9.6. Diagram Pencar (*Scatter Plot*)

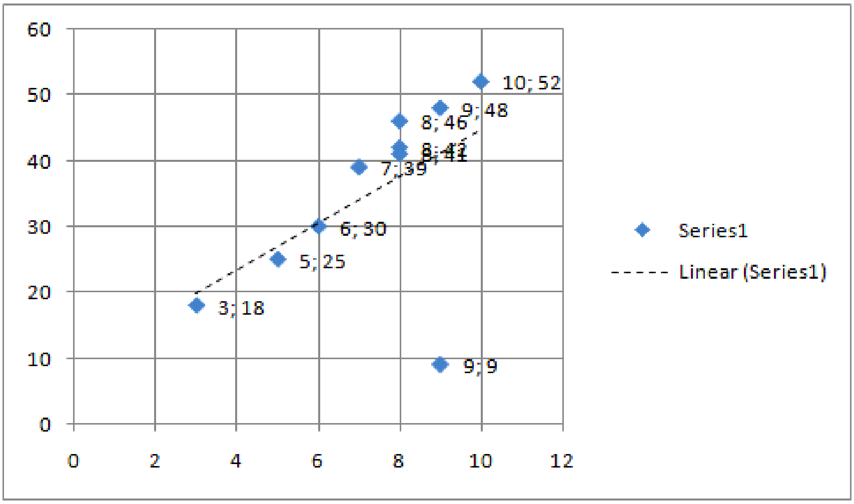
Contoh 9.1 : Diagram Pencar Skor IQ dengan Kinerja.

Untuk memperkirakan pengaruh variabel skor IQ terhadap kinerja karyawan dapat digambarkan melalui *scatter diagram*.

Tabel 9.1. Skor IQ dan Kinerja Karyawan.

Karyawan	Skor IQ	Skor Kinerja
A	6	30
B	9	9
C	3	18
D	8	42
E	7	39
F	5	25
G	8	41
H	10	52
I	9	48
J	8	46

Karena skor IQ diperkirakan mempengaruhi terhadap produksi, maka skor IQ menjadi sumbu datar dan skor kinerja menjadi sumbu tegak.



Gambar 9.2 Diagram Pencar Skor IQ – Kinerja Karyawan.

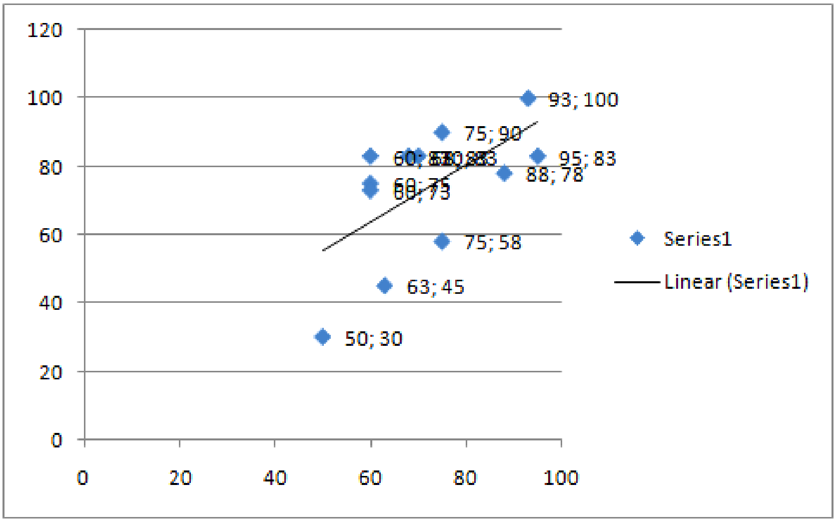
Secara kasat mata, tampak bahwa seluruh koordinat data berpasangan itu berada di dekat garis lurus (*trend*). Slope garis tersebut positif, maka dinyatakan IQ positif terhadap hasil produksi. Namun belum menjelaskan berapa besar pengaruh itu. Untuk itu dibutuhkan analisis regresi.

Contoh 9.2 : Pencar Diagram antara Nilai UTS dan nilai UAS.

Tabel 9.2 Nilai UTS dan UAS

Nama	UTS	UAS	Nama	UTS	UAS
Asrin	50	30	Bardian	68	83
Sunarko	95	83	Eko	75	58
Kusrini	75	90	Saleh	70	83
Dedy	60	83	Edy	60	73
Edwin	60	75	Parman	88	78
Farida	63	45	Johny	93	100

Diperkirakan bahwa nilai UAS dipengaruhi juga oleh nilai UTS. Diagram pencarnya adalah:



Gambar 9.3 Diagram Pencar Nilai UAS dan Nilai UTS.

Terlihat bahwa pengaruh nilai UTS terhadap nilai UAS, positif; walaupun tidak cukup kuat mengingat titik koordinat berpasangan tidak sangat dekat dengan garis *trend*-nya. Pada contoh ini, besaran pengaruh UTS terhadap UAS belum cukup terhitung. Estimasi garis *scatter*, dilakukan secara visual (pandangan mata). Untuk memperkuat pernyataan bahwa nilai UTS berpengaruh terhadap nilai UAS, dapat dihitung *sample correlation coefficient*. Dengan Excel, koefisien korelasi dapat dihitung dengan perintah: "=CORREL(array1;array2)".

Pada contoh ini, koefisien korelasi = 0,6033. Uji signifikansi koefisien korelasi dengan uji-t: $t = \sqrt{\{(n - 2)/(1 - r_p^2)\}} = \sqrt{\{(12 - 2)/(1 - 0{,}6033^2)\}} = 3{,}9610$. Nilai $t_{0,05;10} = 1{,}812$ (periksa Tabel-t). Nilai t hasil hitung > 1,812; maka koefisien korelasi tersebut signifikan walaupun sekedar cukup.

9.7. Analisis Regresi Sederhana Linier (*Bivariate Regression*)

Mengingat diagram pencar hanya mampu mengestimasi ada pengaruh atau tidak dari sebuah variabel ke variabel lainnya, dan belum menghitung berapa besar pengaruh tersebut; maka untuk menghitung besaran pengaruh tersebut dapat digunakan analisis regresi sederhana.

Secara umum, persamaan regresi sederhana linier adalah :

$\hat{Y}_t = a + b\ X_t$

Atau :

$\hat{Y}_t = \text{®}_0 + \text{®}_1\ X_t$

$$\boxed{\text{®}_1 = \frac{\text{©}\ x_t y_t}{\text{©}\ x_t^2}} \quad \text{atau} \quad \boxed{\text{®}_1 = \frac{n\ \text{©}\ X_t Y_t - \text{©}\ X_t \text{©}\ Y_t}{n\ \text{©}\ X_t^2 - (\text{©}\ X_t)^2}} \quad \text{dan} \quad \text{®}_0 = \overline{Y} - \text{®}_1 \overline{X}$$

$$x_t = X_t - \overline{X} \qquad y_t = Y_t - \overline{Y}$$

<u>Contoh 9.3</u> : Analisis Regresi Linier Sederhana

Jika diperkirakan bahwa skor IQ berpengaruh terhadap skor kinerja karyawan; maka untuk mengetahui besaran pengaruh itu dapat digunakan analisis regresi.

Tabel 9.3 Skor IQ dan Skor Kinerja Karyawan

Nama	Skor IQ	Skor Kinerja	x_t	y_t	$x_t y_t$	x_t^2	y_t^2
A	6	30	-1	-8,08	8,08	1	65,34
B	9	49	2	10,92	21,83	4	119,17
C	3	18	-4	-20,08	80,33	16	403,34
D	8	42	1	3,92	3,92	1	15,34
E	7	39	0	0,92	0,00	0	0,84
F	5	25	-2	-13,08	26,17	4	171,17
G	8	41	1	2,92	2,92	1	8,51
H	10	52	3	13,92	41,75	9	193,67

I	9	48	2	9,92	19,83	4	98,34
J	8	46	1	7,92	7,92	1	62,67
K	5	32	-2	-6,08	12,17	4	37,01
L	6	35	-1	-3,08	3,08	1	9,51
©	84	457	0	0,00	228,00	46	1184,92

$\bar{X}$ = 84/12 = 7 $\bar{Y}$ = 457/12 = 38,0833 n = 12

®$_1$ = 228/46 = 4,96 ®$_0$ = 38,0833 – 4,96 (7) = 3,3633

Persamaan regresinya ditulis :

$\hat{Y}_t = 3{,}3633 + 4{,}96\ X_t + e_t$

Arti dari persamaan regresi ini adalah : jika nilai X berubah 1 unit, maka nilai Y juga berubah sebesar 4,96 unit. Contoh: jika nilai X = 4, maka diperkirakan nilai Y = 3,3633 + 4,96(4) = 23,2033, jika nilai X meningkat = 5, maka diperkirakan nilai Y = 3,3633 + 4,96(5) = 28,1633.

Nilai Y yang sekarang dengan adanya peningkatan nilai X = 1 unit, meningkat sebesar = 28,1633 – 23,2033 = 4,96 unit. Demikian pula jika nilai X menurun menjadi = 3, maka nilai Y = 3,3633 + 4,96(3) = 18,2433. Penurunan nilai Y karena X menurun 1 unit = 23,2033 – 18,2433 = 4,96 unit. X berpengaruh positif terhadap Y.

Alternatif cara menghitung parameter regresi adalah :

No.	X_t	Y_t	X_t^2	Y_t^2	X_tY_t
1	6	30	36	900	180
2	9	49	81	2401	441
3	3	18	9	324	54
4	8	42	64	1764	336
5	7	39	49	1521	273
6	5	25	25	625	125
7	8	41	64	1681	328
8	10	52	100	2704	520
9	9	48	81	2304	432
10	8	46	64	2116	368

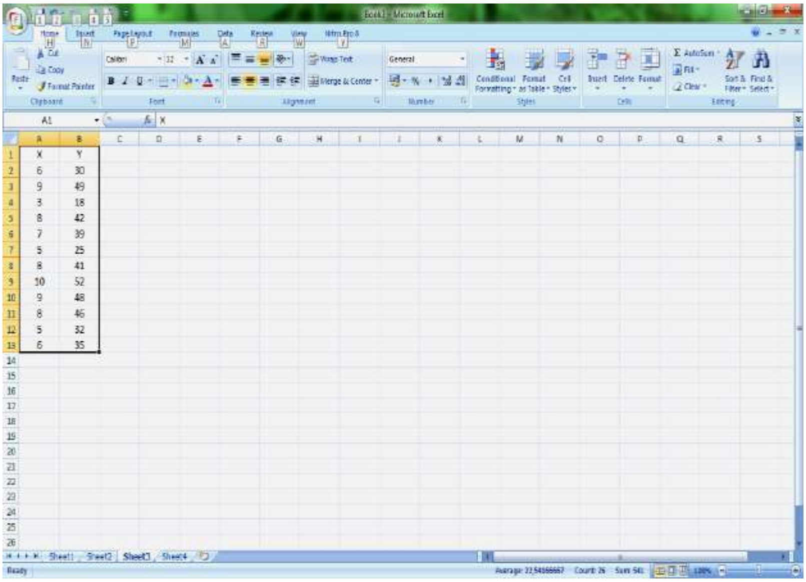
11	5	32	25	1024	160
12	6	35	36	1225	210
©	84	457	634	18589	3427

®$_1$ dapat dihitung sebagai: $\dfrac{(12)(3426) - (84)(457)}{(12)(634) - 84^2} = 4,93$

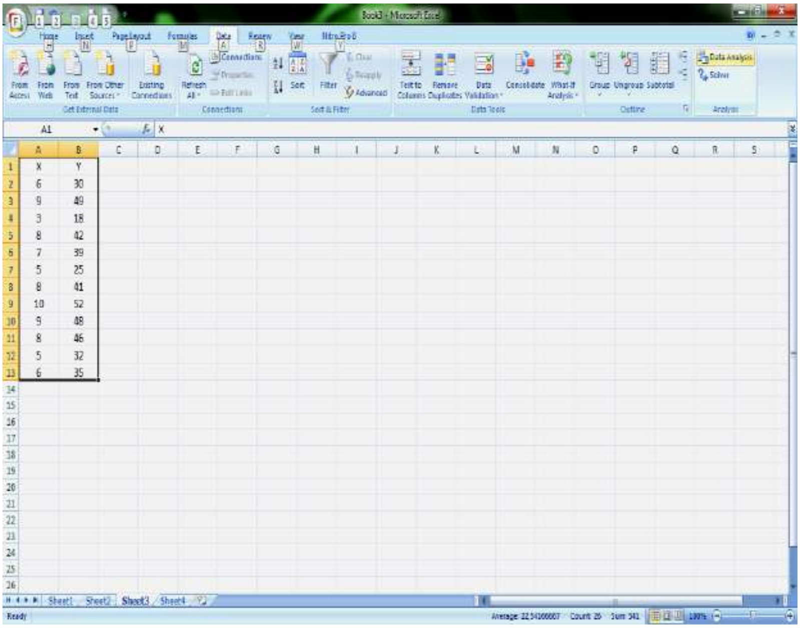
Dari persamaan regresi di atas, dapat diestimasi skor kinerja seorang karyawan dengan skor IQ = 2; $\hat{Y}_t = 3,3633 + 4,96\ X_t + e_t = 3,3633 + 4,96\ (2) = 13,2833$ atau dibulatkan menjadi 13.

Dengan aplikasi Excel, paramater regresi dapat dihitung dengan mudah. Langkah-langkahnya adalah:

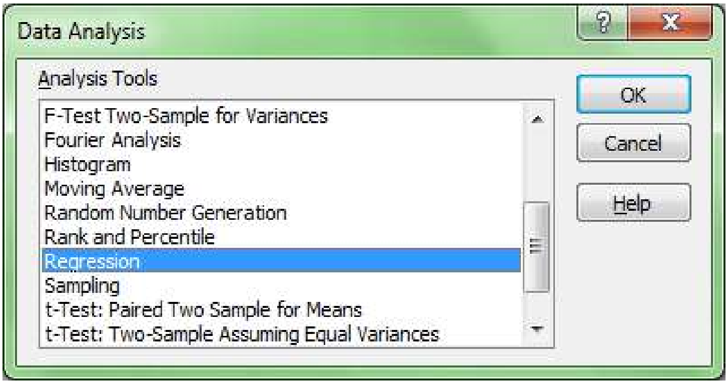
a. Ketik data X dan Y

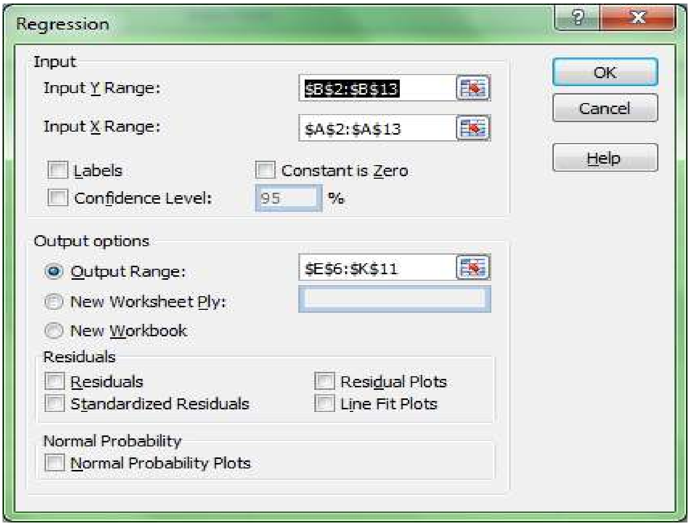


b. Pilih ‘*Data*’.



Pilih ‘*Regression*’ :





c. Tekan OK, hasilnya :

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0,976589
R Square	0,953727
Adjusted R Square	0,9491
Standard Error	2,341574
Observations	12

ANOVA

	Df	*SS*	*MS*	*F*
Regression	1	1130,087	1130,087	206,1085
Residual	10	54,82971	5,482971	
Total	11	1184,917		

	Coef-ficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	3,387	2,5094	1,3499	0,206793	-2,2037	8,97914
X Variable 1	4,956	0,3452	14,356	5,32E-08	4,1872	5,72577

$\hat{Y}_t = 3{,}3633 + 4{,}96\ X_t + e_t$. Ada perbedaan nilai ®$_0$, karena masalah pembulatan dari proses perhitungannya.

<u>Contoh 9.4</u> : Analisis Pengaruh Waktu Belajar Terhadap Nilai Ujian

Pengaruh waktu yang digunakan untuk berlajar terhadap nilai ujian siswa diestimasi dengan mengamati 12 orang siswa. Datanya adalah:

Tabel 9.4 Waktu Belajar dan Nilai Ujian Siswa.

Nama	X_t	Y_t	x_t	y_t	x_t^2	$x_t y_t$	$\hat{Y}_t$
A	4	60	-1,333	-4,42	1,78	5,89	59,92
B	2	53	-3,333	-11,42	11,11	38,06	53,18
C	4	61	-1,333	-3,42	1,78	4,56	59,92
D	6	68	0,667	3,58	0,44	2,39	66,66
E	3	55	-2,333	-9,42	5,44	21,97	56,55
F	5	63	-0,333	-1,42	0,11	0,47	63,29
G	4	59	-1,333	-5,42	1,78	7,22	59,92
H	7	70	1,667	5,58	2,78	9,31	70,04
I	6	67	0,667	2,58	0,44	1,72	66,66
J	5	65	-0,333	0,58	0,11	-0,19	63,29
K	8	72	2,667	7,58	7,11	20,22	73,41
L	10	80	4,667	15,58	21,78	72,72	80,15
©	64	773	0	0,00	54,67	184,33	

$\bar{X}$ $\bar{Y}$

5,33   64,41
7

®$_0$ =   64,417 – (3,37)(5,33) = 46,455

®$_1$ =   184,33/54,67 = 3,37

Fungsi regresi : $\hat{Y}_t$ = 46,43 + 3,37 X_t + e_t. Artinya, nilai ujian adalah = 46,43;
jika waktu belajar siswa = 0 jam. Setiap tambahan 1 jam belajar, ada peningkatan nilai ujian = 3,37. $\hat{Y}_t$ adalah estimasi nilai Y_t dengan aplikasi model regresi yang dihasilkan. Contoh: untuk siswa A : $\hat{Y}_t$ = 46,43 + 3,37 X_t = 46,43 + 3,37(4) = 59,92. Fungsi regresi yang dihasilkan ini perlu diuji kesesuaian relatif modelnya, berdasar nilai perbandingan *sum squares of regression* (SSR) dan nilai *sum squares of total* (SST); untuk menghitung nilai koefisien determinasi, R^2. Formula untuk menghitung R^2 adalah :

$$\boxed{R^2 = 1 - \frac{SSE}{SST}} \quad \text{atau} \quad \boxed{R^2 = \frac{SSR}{SST}} \quad \ldots\ldots (9.3)$$

Di mana,
SSE = *sum of squares of error* = © $(Y_t - \hat{Y}_t)^2$ adalah variasi nilai Y yang tidak dapat dijelaskan oleh variasi nilai X.
SSR = *sum of squares of regression* = © $(\hat{Y}_t - \bar{Y})^2$ adalah variasi nilai Y yang dapat dijelaskan oleh variasi nilai X.
SST = *total sum of squares* = © $(Y_t - \bar{Y})^2$ adalah total variasi nilai Y terhadap rata-ratanya. Lembar kerja untuk menghitung SSE, SSR dan SST:

Nama	Y_t	$\hat{Y}_t$	$Y_t - \hat{Y}_t$	$(Y_t - \hat{Y}_t)^2$	$(\hat{Y}_t - \bar{Y})^2$	$(Y_t - \bar{Y})^2$
A	60	59,921	0,079	0,006	20,213	19,50694
B	53	53,177	-0,177	0,031	126,334	130,3403
C	61	59,921	1,079	1,165	20,213	11,67361

Syafiadi Rizki Abdila, S.T., Ph.D, Prof. Dr. Ir. Drs. Syafwandi, M.Sc
Dr. Yulasmi, S.E., M.M, Dr. Robby Dharma, S.E., M.M
Dr. Syafri Arlis, S.Kom, M.Kom

D	68	66,665	1,335	1,783	5,053	12,84028
E	55	56,549	-1,549	2,399	61,904	88,67361
F	63	63,293	-0,293	0,086	1,263	2,006944
G	59	59,921	-0,921	0,848	20,213	29,34028
H	70	70,037	-0,037	0,001	31,583	31,17361
I	67	66,665	0,335	0,112	5,053	6,673611
J	65	63,293	1,707	2,915	1,263	0,340278
K	72	73,409	-1,409	1,984	80,854	57,50694
L	80	80,152	-0,152	0,023	247,615	242,8403
©	773			11,3537	621,563	632,917
$\overline{Y}$	64,42			SSE	SSR	SST

R^2 = 1 – 11,354/632,917 = 0,982061 atau = 622,563/632,917 = 0,982061, atau = 98,21%. Ini mengartikan bahwa 98,21% perubahan nilai ujian dapat dijelaskan oleh perubahan waktu belajarnya. Secara parsial masih perlu dilakukan uji hipotesis terhadap keberartian (signifikansi) koefisien ®$_0$ dan ®$_1$ pada persamaan regresi tersebut di atas. Untuk itu, dilakukan uji hipotesis dengan menggunakan pendekatan uji-t. Formula uji hipotesis yang digunakan adalah:

$$t®_i = \frac{®_i}{SE\{®_i\}} \quad \text{(untuk i = 0, 1)} \qquad ..... (9.4)$$

Di mana SE{®$_i$} adalah standard error untuk ®$_i$.

$$SE\{®_0\} = SE\{\text{Regresi } YX\} + \sqrt{\frac{1}{n} \quad \frac{X^2}{©\,(X_t - \overline{X})^2}}$$

..... (9.5)

$$SE\{®_1\} = \frac{SE\{\text{Regresi } YX\}}{©(X_t - \overline{X})^2} \qquad ..... (9.6)$$

$$\boxed{SE\{\text{Regresi } YX\} = \sqrt{SSE/(n-2)}} \quad \ldots\ldots (9.7)$$

Untuk contoh ini, dapat dihitung:

$\sqrt{\Sigma(X_t - \overline{X})^2} = 54{,}67$ (di mana $\overline{X} = 5{,}33$)

SE{Regresi YX} = $\sqrt{11{,}354/(12-2)} = 1{,}0656$

SE{β_0} = $1{,}0656 \sqrt{(1/12 + (5{,}33^2/54{,}67)} = 0{,}82746$

SE{β_1} = $1{,}06536/\sqrt{54{,}67} = 0{,}1441$

t®$_0$ = 46,43/0,82746 = 56,111473

t®$_1$ = 3,37/0,1441 = 23,3849143

Hipotesis dalam uji signifikansi parsial koefisien regresi adalah:

H_0 : ®$_0 \leq 0$	H_a : ®$_0 > 0$
H_0 : ®$_1 = 0$	H_a : ®$_1 \neq 0$

H_0 ditolak jika t hasil perhitungan > t_{critic}. H_0 ditolak, karena t®$_0$ = 50,10 > $t_{0,05;10}$ (=1,860), demikian pula t®$_1$ = 20,797 > $t_{0,05;10}$ (=1,860).

Output regresi selalu melibatkan Analisis Varians (ANOVA) yang menghitung nilai F, yaitu ratio antara SSE dan SSR. Makin besar nilai F, persamaan regresi yang dihasilkan makin baik (makin sesuai). Nilai F ini dibandingkan dengan F_{critic} pada α dan derajat bebas = n – 2.

Formula untuk menghitung F adalah :

$$\boxed{F = \frac{MSR}{MSE} = \frac{SSR/(k-1)}{SSE/(n-2)} = (n-2)\frac{SSR}{SSE}} \quad \ldots\ldots (9.8)$$

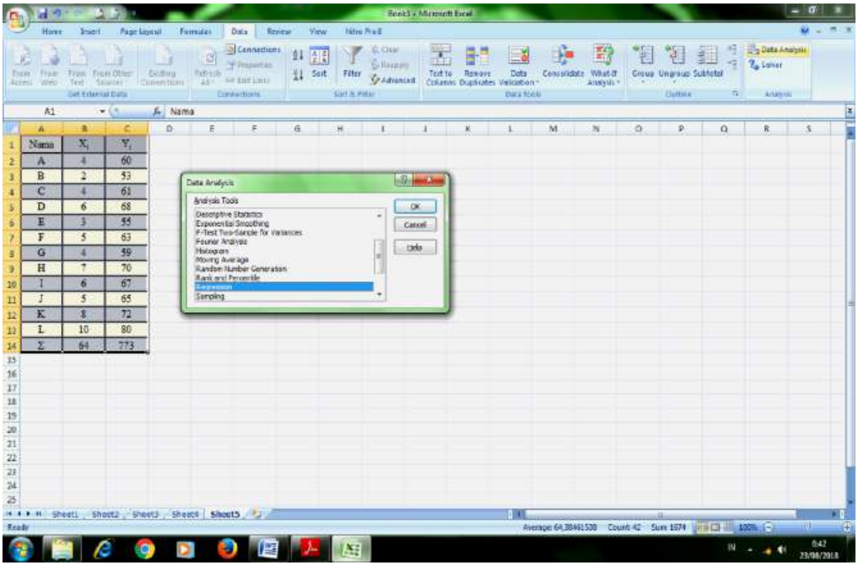
k = banyaknya variabel, untuk regresi sederhana, k = 2.

Pada contoh ini, nilai F perhitungan =

$$(2-1)\ \frac{SSR}{SSE} = 10\ \frac{621{,}563}{11{,}354} = 547{,}40$$

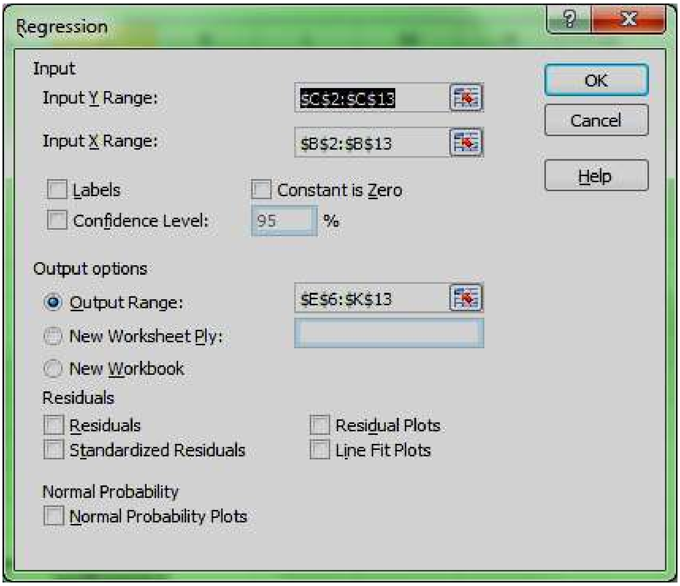
Nilai $F_{0,05;1;10}$ = 4,96 (periksa Tabel F). Karena nilai F perhitungan > F_{critic}, maka secara keseluruhan fungsi regresi yang dihasilkan dinyatakan sesuai dalam menjelaskan pengaruh X terhadap Y.

Dengan aplikasi Excel, analisis regresi dapat dilakukan dengan menggunakan sub menu[21] : ‘*Data – Data Analysis – Regression*’. Berikut ini hasil aplikasi analisis regresi dengan Excel:



[21] Dengan catatan bahwa pada Excel Add-in Analysis Tool Pack telah diinstalasi..

Tekan OK, tampilan kotak dialognya:



SUMMARY OUTPUT

Regression Statistics	
Multiple R	0,99099
R Square	0,98206
Adjusted R Square	0,98026
Standard Error	1,06553
Observations	12

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	621,563	621,563	547,4561	4,61E-10
Residual	10	11,35366	1,135366		
Total	11	632,9167			

	Coef-ficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	46,43293	0,82787	56,0870	7,87E-14	44,5883	48,2775

X Variable 1	3,371951	0,14411	23,3977	4,61E-10	3,05084	3,69305

Jika ada perbedaan hasil, namun kecil; ini disebabkan karena masalah pembulatan-pembulatan angka dalam proses perhitungan.

9.8. Pendekatan Operasi Matriks Untuk Model Regresi

1. Estimasi Koefisien Regresi Linier Dengan k-Variabel

Fungsi regresi secara umum dapat dituliskan sebagai:

$\hat{Y}_t$ = ®$_0$ + ®$_1$ X1$_t$ + ®$_2$ X2$_t$ + ... + ®k Xk$_t$ + e$_t$

Dengan operasi matriks, persamaan di atas dapat dituliskan sebagai berikut :

Y = **X** **®** **e**
(n x 1) (n x k) (k x 1) (n x 1)

Di mana,

Y berdimensi n x 1, merupakan vektor kolom observasi variabel dependen,

X berdimensi n x k, matriks sebanyak n observasi dari k - 1 variabel bebas,

® berdimensi k x 1, merupakan vektor kolom parameter yang tidak diketahui,

e berdimensin x 1, merupakan vektor kolom *error* sebanyak n.

Jika,

Y = X ® + e

Maka :

® = **(X'X)$^{-1}$** **X'Y** **e**

(k x 1) (k x k) (k x n) (n x 1) □

merupakan dimensi matriks yang bersangkutan (banyaknya baris dan kolom).

2. **Matriks Variance-Covariance,** ®

Operasi matriks memungkinkan untuk mengembangkan formula, tidak hanya untuk menghitung varians ®, tetapi juga untuk menghitung kovarians antara dua elemen ®, misal $®_i$ dan $®_j$. Formulanya adalah : **var-cov (®) = ⌠^{2} (X’X)-1.** ⌠^{2} adalah varians homoskedastis dari e_t, dan **(X’X)-1** adalah inversi matriks **X’X** seperti yang dihasilkan pada operasi matriks sebelumnya. Dalam model regresi linier dengan dua atau tiga variabel, estimator yang *unbiased*:

⌠^{2} = ©e_t^2/(n-2) atau ⌠^{2} = © e_t^2/(n-3). Untuk k - variabel, maka rumus itu menjadi:

⌠^{2} =© e_t^2/(n-k) atau dalam operasi matriks : **⌠^{2} = e’e/(n-k).**

e’e dapat dihitung melalui pendekatan[22] :

e’e = y’y - ®’ X’y

Sekali **e’e** dapat dihitung, maka ⌠^{2} dapat dihitung.

3. **Koefisien Determinasi (R^2)**

Koefisien determinasi, R^2, dihitung dengan formula (9.3). Untuk model regresi dengan variabel bebas sebanyak-k, dengan operasi matriks R^2 dapat dihitung sebagai:

$$R^2 = \frac{\beta' X' y - n\,\overline{Y}^2}{y' y - n\,\overline{Y}^2}$$

4. **Uji Hipotesis Signifikansi Koefisien Regresi Parsial**

Uji signifikansi ®i secara parsial dapat didekati dengan formula:

[22] Gujarati, 2008:150.

$$t_{\beta i} = \frac{\hat{\beta}_i - \beta_i}{SE\{\beta_i\}}$$

Di mana,

SE = *standard error*

Kriteria signifikansi ®$_i$: t®$_i$ < t®$_{df;〈}$. df = *degree of freedom* = n – k.

5. Uji Hipotesis Signifikansi Koefisien Regresi Simultan

Uji signifikansi koefisien regresi secara simultan, didekati melalui ANOVA (uji-F) sebagai berikut :

Tabel ANOVA (dalam operasi matriks)

Sumber Variasi	*Sum of Squares (SS)*	*df*	*Mean of Sum of Squares (MSS)*	F
Regresi (SSR)	$\beta'X'y - n\,\overline{Y}^2$	k - 1	$\beta'X'y - n\,\overline{Y}^2/(k-1)$	MSR/MSE
Error (SSE)	$y'y - \beta'X'y$	n - k	$y'y - \beta'X'y/(n-k)$	
Total (TSS)	$y'y - n\,\overline{Y}^2$	n - 1		

$$F = \frac{\beta'X'y - n\,\overline{Y}^2/(k-1)}{y'y - \beta'X'y/(n-k)}$$

atau dengan menggunakan R^2 yang telah dihitung[23] :

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

<u>Contoh-9.5</u> : Analisis Regresi Dengan Operasi Matriks

Sebuah model regresi linier untuk menjelaskan pengaruh *personal disposable income* per capita (PPDI) terhadap *personal*

[23] Gujarati, 2008:154.

comsumption expenditure per capita (PPCE), menggunakan input data dalam $ (1996 - 2010) dan trend (t):

Tabel 9.5 Data PPCE dan PPDI.

PPCE (Y)	PPDI (X_1)	t (X_2)
1673	1839	1
1688	1844	2
1666	1831	3
1735	1881	4
1749	1883	5
1756	1910	6
1815	1969	7
1865	2016	8
1947	2126	9
2048	2239	10
2128	2336	11
2165	2404	12
2257	2487	13
2316	2535	14
2324	2595	15

Dalam bentuk operasi matriks adalah:

$$\begin{bmatrix} 1673 \\ 1688 \\ 1666 \\ 1735 \\ 1749 \\ 1756 \\ 1815 \\ 1865 \\ 1947 \\ 2048 \\ 2128 \\ 2165 \\ 2257 \\ 2316 \\ 2324 \end{bmatrix} = \begin{pmatrix} 1 & 1839 & 1 \\ 1 & 1844 & 2 \\ 1 & 1831 & 3 \\ 1 & 1881 & 4 \\ 1 & 1883 & 5 \\ 1 & 1910 & 6 \\ 1 & 1969 & 7 \\ 1 & 2016 & 8 \\ 1 & 2126 & 9 \\ 1 & 2239 & 10 \\ 1 & 2336 & 11 \\ 1 & 2404 & 12 \\ 1 & 2487 & 13 \\ 1 & 2535 & 14 \\ 1 & 2595 & 15 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} e1 \\ e2 \\ e3 \\ e4 \\ e5 \\ e6 \\ e7 \\ e8 \\ e9 \\ e10 \\ e11 \\ e12 \\ e13 \\ e14 \\ e15 \end{pmatrix}$$

$$\underset{15 \times 1}{Y} = \underset{15 \times 3}{X} \; \underset{3 \times 1}{\beta} + \underset{15 \times 1}{e}$$

$\overline{Y} = 1942{,}133 \quad \overline{X}_1 = 2126{,}333 \quad \overline{X}_2 = 8$

Untuk kemudahan dalam menghitung β, gunakan program Excel:

(a) Matriks X :

1	1839	1
1	1844	2
1	1831	3
1	1881	4
1	1883	5
1	1910	6
1	1969	7
1	2016	8
1	2126	9
1	2239	10
1	2336	11
1	2404	12
1	2487	13
1	2535	14
1	2595	15

(b) Transpos Matriks X atau X' :

1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1839	1844	1831	1881	1883	1910	1969	2016	2126	2239	2336	2404	2487	2535	2595
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Matriks tranpos adalah matriks dengan dimensi di mana baris menjadi kolom, dan kolom menjadi baris. Contoh, jika matriks X berdimensi 2 x 4, atau barisnya = 2 dan kolomnya = 4, maka transpos matriks X atau X' berdimensi 4 x 2, atau barisnya = 4 dan kolomnya = 2.

Untuk matriks X' di atas, ini dihasilkan dengan mengcopy-transpose matriks X.

(c) Perkalian Matriks X'X :

15	31895	120
31895	68922513	272144
120	272144	1240

Ini dihasilkan dengan memberikan perintah "=MMULT(*array,array*)". Kemudian blok sel 3 x 3, tekan F2, dan tekan Ctrl+Shift+Enter.

(d) Inversi Matriks X'X atau $X'X^{-1}$

37,23277	-0,02251	1,33671
-0,02251	0,00001	-0,00083
1,33671	-0,00083	0,05403

Ini dihasilkan dengan memberikan perintah "=MINVERSE(*array*)". Kemudian blok sel 3 x 3, tekan F2, dan tekan Ctrl+Shift+Enter.

(e) Matriks Y :

1673
1688
1666
1735
1749
1756
1815
1865
1947
2048
2128
2165
2257
2316
2324

(f) Perkalian Matriks X'Y :

29132
62899663
247909

(g) Perkalian Matriks $X'X^{-1}$ dengan Matriks X'Y (untuk menghitung β_i) :

293,77526
0,7458439
7,8056693

$\beta_0 = 293,775$
$\beta_1 = 0,746$
$\beta_2 = 7,806$

Syafiadi Rizki Abdila, S.T., Ph.D, Prof. Dr. Ir. Drs. Syafwandi, M.Sc
Dr. Yulasmi, S.E., M.M, Dr. Robby Dharma, S.E., M.M
Dr. Syafri Arlis, S.Kom, M.Kom

(h) Model Regresi Linier yang dihasilkan :

$\hat{Y}_t = 293{,}775 + 0{,}746\ X_{1t} + 7{,}806\ X_{2t} + e_t$

(i) Menghitung se(β_i) :

(1) Terlebih dahulu dihitung : $e'e = y'y - \beta'\ X'y$

$y'y = 57408644$ (dihitung dengan =MMULT)

$\beta'\ X'y$: 57406686 (dihitung dengan =MMULT)

$e'e = 57408644 - 57406686 = 1958{,}2954$,

$\sigma^2 = e'e/(N\text{-}k) = 1958{,}2954/(15\text{-}3) = 163{,}1912805$

(2) Menghitung Matriks Var-Covar(β) : $\text{var-cov}\ (\beta) = \sigma^2\ (X'X)^{-1}$

6076,064	-3,673	218,139
-3,673	0,002	-0,136
218,139	-0,136	8,818

(3) Menghitung se(β_j) :

se(β_i) merupakan akar kuadrat dari diagonal matriks **var-cov** (β) :

se(β_0) = 77,9491 $\rightarrow \sqrt{}$ 6076,064

se(β_1) = 0,0473 $\rightarrow \sqrt{}$ 0,002

se(β_2) = 2,9695 $\rightarrow \sqrt{}$ 8,818

(j) Menghitung Koefisien Determinasi, R^2

Koefisien determinasi (R^2) merupakan ukuran kesesuaian model secara keseluruhan. Yang dimaksud dengan kesesuaian model secara keseluruhan adalah seberapa besar perubahan variabel Y dapat dijelaskan oleh perubahan variabel X_i. $R^2 \leq 1{,}00$. Makin besar nilai R^2, berarti modelnya makin sesuai dalam menjelaskan pengaruh perubahan variabel independen terhadap perubahan variabel dependen. Pada model regresi bivariat (regresi sederhana), R^2 adalah kuadrat dari r.

$$R^2 = \frac{\beta' X' y - n\,\overline{Y}^2}{y'y - n\,\overline{Y}^2}$$

$$= \frac{57406686 - 15\,(1942{,}133)^2}{57408644 - 15\,(1942{,}133)^2}$$

$$= (828457{,}4)/(830415{,}7) = 0{,}997642$$

$$= 99{,}764\%$$

Untuk menghitung *adjusted* R^2 dengan formula:[24]

$$\overline{R}^2 = 1 - (1 - R^2)\frac{n - 1}{n - k} = 0{,}99725 = 99{,}725\%$$

Adjusted R^2 ditujukan untuk mengukur kesesuaian model-model regresi dengan variabel dependen yang sama, namun berbeda dalam banyaknya variabel independen. Model regresi dengan 2 variabel bebas akan menghasilkan R^2 yang lebih rendah daripada model regresi yang melibatkan 3 variabel bebas. Untuk itu yang diperbandingkan adalah *adjusted* R^2 bukan lagi R^2.

(k) ANOVA (untuk uji signifikansi koefisien regresi secara simultan).

[24]Gujarati;2005:110

Tabel ANOVA.

Sumber Variasi	*Sum of Squares (SS)*	df	*Mean of Sum of Squares (MSS)*	F
Regresi (SSR)	$\beta'X'y - n\overline{Y}^2$	k - 1	$\beta'X'y - n\overline{Y}^2/(k-1)$	SSR/SSE
Error (SSE)	$y'y - \beta'X'y$	n - k	$y'y - \beta'X'y/(n-k)$	
Total (TSS)	$y'y - n\overline{Y}^2$	n - 1		

Tabel ANOVA.

Sumber Variasi	*Sum of Squares (SS)*	df	*Mean of Sum of Squares (MSS)*	F
Regresi (SSR)	828457,4	2	414228,7	2538,3014
Error (SSE)	1958,2954	12	163,1913	
Total (TSS)	830415,7	14		

(1) Menuliskan fungsi regresi eksplisit secara lengkap :

Menuliskan fungsi regresi secara lengkap :

$$\hat{Y}_t = 293{,}775 + 0{,}746\, X_{1t} + 7{,}806\, X_{2t} + e_t$$

se(β_k)	(77,9491)	(0,0473)	(2,9695)
t	(3,7688)	(15,7650)	(2,6286)
R^2		(99,764%)	
Adjusted R^2		(99,725%)	
F		(2538,3014)	

Keterangan : nilai-t, dihitung dengan : $t_{\beta k} = \dfrac{\hat{\beta}_k - \beta_k}{se(\beta_k)}$

Interpretasi statistiknya adalah :

(1) Secara parsial kedua koefisien regresi dan konstanta signifikan pada 〈 = 0,05, sedangkan nilai t-tabel pada 〈 = 0,05 dan df = 12 adalah = 1,782. Inidiindikasikan oleh nilai-t hitung konstanta dan seluruh koefisien regresi > 1,782.

(2) Secara simultan, kedua koefisien regresi signifikan pada 〈 = 0,05, inidiindikasikan oleh nilai-F hitung = 2538,3014 yang lebih besar daripada nilai-Ftabel pada 〈 = 0,05, numerator = k-1 = 2, dan denominator n - k = 12, yaitusebesar = 3,89. Linieritas model terbukti benar.

Hasil *printout* analisis Regresi dengan *Add-in* Data Analysis pada Excel menunjukkan kemiripan:

SUMMARY OUTPUT	
Regression Statistics	
Multiple R	0,9988
R Square	0,9976
Adjusted R Square	0,9972
Standard Error	12,7746
Observations	15

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	828457,4	414228,72	2538,3017	1,72E-16
Residual	12	1958,295	163,19128		
Total	14	830415,7			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*
Intercept	293,7753	77,9491	3,7688	0,0027	123,9387	463,6118	123,9387
X Variable 1	0,7458	0,0473	15,7650	0,0000	0,6428	0,8489	0,6428
X Variable 2	7,8057	2,9695	2,6286	0,0220	1,3357	14,2757	1,3357

9.9. Analisis Regresi Linier Berganda

Ini merupakan perluasan analisis regresi bivariat, dengan menggunakan variabel independen sama atau lebih dari dua. Peneliti umumnya menyadari bahwa tidak hanya sebuah variabel independen yang mempengaruhi variabel dependen. Dalam memodelkan regresi linier berganda, selayaknya dinyatakan terlebih dahulu hipotesis yang ingin dibuktikan (termasuk sifat pengaruh variabel independen terhadap variabel dependen, walaupun tidak tertutup kemungkinan sifat pengaruh itu tidak dapat diperkirakan).

<u>Contoh 9.6</u> : Analisis Regresi Linier Berganda

Diperkirakan harga rumah (HR) dipengaruhi secara positif oleh luas rumah (LR), luas tanah (LT), dan banyaknya kamar mandi (KM). Data penjualan rumah pada suatu *real estate* pada tahun tertentu adalah:

Tabel 9.6 Harga Rumah, Luas Rumah, Luas Tanah dan Banyaknya Kamar Mandi.

Rumah	HR	LR	LT	KM	Rumah	HR	LR	LT	KM
1	505,5	2,192	16,4	2,5	16	675,1	3,076	19,8	3,0
2	784,1	3,429	24,7	3,5	17	710,4	3,259	20,8	3,5
3	649,0	2,842	17,7	3,5	18	674,7	3,162	19,4	4,0
4	689,8	2,987	20,3	3,5	19	663,6	2,885	23,2	3,0
5	709,2	3,029	22,2	3,0	20	606,6	2,550	20,2	3,0
6	590,2	2,616	20,8	2,5	21	758,9	3,380	19,6	4,5
7	643,3	2,978	17,3	3,0	22	723,7	3,131	22,5	3,5
8	789,7	3,595	22,4	3,5	23	621,8	2,754	19,2	2,5
9	683,0	2,838	27,4	3,0	24	622,4	2,710	21,6	3,0
10	544,3	2,591	19,2	2,0	25	631,3	2,616	20,8	2,5
11	822,8	3,633	26,9	4,0	26	574,0	2,608	17,3	3,5
12	637,7	2,822	23,1	3,0	27	863,8	3,572	29,0	4,0
13	618,7	2,994	20,4	3,0	28	652,7	2,924	21,8	2,5
14	619,3	2,696	22,7	3,5	29	844,2	3,614	25,5	3,5
15	490,5	2,134	13,4	2,5	30	629,9	2,600	24,1	3,5

Sumber : diadopsi dari Doane, 2007:561.

Keterangan :
- Harga rumah dalam jutaan Rupiah,
- Luas rumah dalam ratusan meter^2,
- Luas tanah dalam ratusan meter^2,
- Banyaknya kamar mandi dalam unit, jika nilainya pecahan berarti ada tambahan kamar mandi dengan ukuran relatif lebih kecil.
- Model Regresi : $HR_t = \circledR_0 + \circledR_1\ LR_t + \circledR_2\ LT_t + \circledR_3\ KM_t + e_t$

Dengan *add-in* pada Excel, hasil analisis regresi tersebut adalah:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0,9778
R square	0,9560
Adjusted R square	0,9505
Standard Error	20,2787
Observation	30

ANOVA

	df	*SS*	*MS*	*F*	*Significnace F*
Regression	3	232428,2468	77476,08	188,4021	0,0000
Residual	26	10691,91185	411,2274		
Total	29	243120,1587			

	Coefficients	*Standard Error*	*t stat*	*p-value*	*Lower 95%*	*Upper 95%*
Intercept	-28,838	29,672	-0,972	0,340	-89,830	32,154
X variable 1	170,968	15,427	11,076	0,000	139,157	202,579
X variable 2	6,777	1,419	4,774	0,000	3,859	9,695

X variable 2	15,567	9,196	1,693	0,102	-3,335	34,470

Model regresi yang dihasilkan dapat dituliskan sebagai berikut:

$HR_t = -28,838 + 170,87\ LR_t + 6,78\ LT_t + 15,57\ KM_t + e_t$

SE{®k} (29,672) (15,427) (1,419) (9,196)

t -0,972 11,076 4,774 1,693

R^2 0,9560

$R^2_{adjusted}$ 0,9509

F 188,4021

Secara parsial, LR berpengaruh positif signifikan terhadap HR (probabilitassignifikansi = 0,000 < 0,05); LT berpengaruh positif signifikan terhadap HR(probabilitas signifikansi = 0,000 < 0,05), KM pengaruhnya tidak signifikan (probabilitas signifikansi = 0,102 > 0,05). Secara simultan, LR, LT dan KM berpengaruh signifikan terhadap HR, mengingat probabilitas signifikansi F = 0,000 < 0,05. Artinya, walaupun KM tidak signifikan pengaruhnya terhadap HR secara parsial, namun secara bersama dengan LT dan LR; pengaruh KM terhadap HR menjadi signifikan.

Pada analisis regresi berganda linier, *adjusted* R^2 merupakan ukuran COD yang lebih tepat dibanding R^2. Ini disebabkan karena R^2 cenderung makin tinggi dengan makin banyaknya variabel independen. *Adjusted* R^2 bisa lebih tinggi, lebih rendah atau sama dengan R^2. Jika *Adjusted* R^2 jauh lebih kecil daripada R^2, ini menandakan bahwa dalam model regresi ada variabel independen yang tidak diperlukan. Pada contoh ini, penurunan *adjusted* R^2relatif kecil dibanding R^2; maka berarti pada model regresi tidak ada variabel independen yang tidak diperlukan.

Asumsi Dalam Model Regresi Klasik

Dalam analisis regresi, melalui simplifikasi, ada enam asumsi pada model
regresi klasik, yaitu[25]:

1. Rata-rata kondisional dari penyimpangan populasi, $e_t = 0$,
2. Varian e_t = konstan, atau homoskedastis (non heteroskedastis),
3. Tidak terjadi otokorelasi serial antar e_t,
4. Variabel eksplanatori (baik skokastik maupun non skokastik),
5. Tidak terjadi multikolinieritas antar variabel eksplanatori,
6. e_t berdistribusi normal dengan rata-rata = 0, dan standard deviasi = 1.

Dari keenam asumsi ini, tiga merupakan syarat perlu (perlu diuji), yaitu:otokorelasi, multikolinieritas dan homoskedastis, dan tiga lainnya merupakan syarat cukup (boleh diuji dan tidak).

a. Multikolinieritas

Menurut Ragnar Frisch[26], istilah multikolinier merupakan adanya hubungan sempurna dan pasti antar beberapa variabel eksplanatori dalam sebuah model regresi. Ini harus dihindari. Identifikasi terjadinya multikolinieritas sempurna, jika data sebuah variabel eksplanatori = ⌊ dikalikan data variabel lain, di mana ⌊ ≠ 0; ini jelas multikolinier yang sempurna, maka salah satu variabel harus dikeluarkan dari model. Multikolinieritas tidak mengganggu estimasi *least square*, dan tidak berpengaruh terhadap *variance*

---

[25] Gujarati;2005:115

[26]Frisch, 1955:322

inflation. Jika variaberl-variabel independen yang berkorelasi, maka varians dari estimasi koefisien akan terinflasi dan memperlebar interval keyakinan nilai sesungguhnya dari ®$_1$, ®$_2$ . . . ®$_k$. Kondisi ini menyebabkan adanya kesulitan untuk mengidentifikasi kontribusi parsial dari setiap variabel independen dalam menjelaskan variabel dependen.

Untuk mengidentifikasi multikolinier, adalah :

(1) R^2 tinggi, tetapi banyak koefisien regresi yang tidak signifikan,

(2) Pada model regresi dengan dua variabel eksplanatori, bisa dilihat dari *zero order correlation* antar kedua variabel tersebut. Jika tinggi, maka diperkirakanterjadi mutlikolinier,

(3) Untuk model dengan lebih dari dua variabel eksplanatori, bisa dilihat dari *partial correlation*-nya, Jika R^2 tinggi, tetapi *partial correlation* rendah, masih dimungkinkan terindikasi multikolinier. Namun jika R^2 tinggi, dan *partial correlation* juga tinggi, maka multikolinier bisa tidak terdeteksi,

(4) Peneliti dapat meregresi X_i terhadap variabel eksplanatori lainnya yang tersisa.

    Jika R_i^2 tinggi, maka dapat dipastikan bahwa X_i itu multikol dengan variabel-variabel lainnya, maka buang X_i dari model.

(5) Deteksi terbaru adalah dengan melihat kepada *condition index* model yang dianalisis secara *stepwise* (selangkah demi selangkah variabel dimasukkan ke dalam model), Jika *condition index* < 15 (ada indikasi multikolinier antar variabel eksplanatori), masih bisa ditoleransi. Namun jika *condition index* > 30, maka multikol telah menjadi permasalahan serius.

Condition index adalah akar kuadrat dari perbandingan *eigen value* terbesar dengan *eigen value* pada model yang berurutan. *Condition index* ini berkaitan dengan *variance inflation factor* (VIF). VIF adalah kuasi dari tingkat toleransi variabel bebas (VIF =1/*tolerance*).

Tolerance adalah persentase varians pada sebuah variabel bebas yang tidak dapat dijelaskan oleh variabel bebas lainnya. Jika *tolerance* mendekati nol, ini mengindikasikan multikol yang tinggi, sehingga *standarderror* dari koefisien regresi terinflasi. Jika VIF > 3, biasanya telah dapat diindikasikan adanya multikolinier. Jika terjadi multikolinier yang tidak sempurna, konsekuensinya adalah:

(1) Walaupun estimator OLS diperoleh, *standard error* cenderung besar,
(2) Akibatnya *confidence interval* juga makin besar,
(3) Kemungkinan menerima hipotesis yang keliru (*error type* II) makin besar,
(4) *Standard error* makin sensitif untuk berubah jika terjadi perubahan data.

Untuk mendeteksi terjadinya multikolinier, salah satu cara praktis yang bisa digunakan dari keempat cara di bawah ini:

(1) R^2 yang tinggi (antara 0,70 - 1,0), dengan hanya sedikit koefisien regresi yang signifikan, walaupun biasanya F juga tinggi,
(2) Untuk dua model regresi dengan variabel eksplanatori, bisa dilihat dari r antar kedua variabel eksplanatori tersebut, yaitu r signifikan dan cukup tinggi,
(3) Karena multikolinier meningkat yang disebabkan oleh sebuah variabel atau lebih merupakan kombinasi linier dari variabel eksplanatori lainnya, salah satu

cara untuk menemukan X_i yang multikolinier adalah dengan meregres setiap variabel X_i kepada variabel eksplanatori lainnya yang tersisa, hitung R_i^2.

Selanjutnya hitung F_i dengan formula[27] :

$$F_i = \frac{R_i^2/(k-2)}{(1 - R_i^2/(n-k+1)}$$

Jika F-hitung ini melebihi nilai kritis F_i pada 〈 tertentu, ini merupakan indikasi bahwa X_i multikol terhadap variabel X lainnya, dan berlaku sebaliknya. Jika terindikasi multikol, maka variabel X_i tersebut selayaknya dibuang dari model.

Upaya memperbaiki jika terjadi multikolinieritas adalah salah satu cara di bawah ini :
(1) Informasi apriori.
Contoh :
$\hat{Y}_t = \circledR_0 + \circledR_1 X_{1t} + \circledR_2 X_{2t} + e_t$
Di mana,
$\hat{Y}_t$ = pengeluaran konsumsi,
X_{1t} = penghasilan,
X_{2t} = kesejahteraan.
Seperti banyak dinyatakan dalam teori, X_1 berkorelasi dengan X_2; jika secara apriori (karena juga ada kajian empiris yang mendukung), dan kemudian diyakini bahwa $\circledR_2 = 0{,}10\circledR_1$, maka model itu harus dirubah menjadi:
$\hat{Y}_t = \circledR_0 + \circledR_1 X_{1t} + 0{,}10\ \circledR_1 X_{2t} + e_t$, atau dimodifikasi sebagai:
$\hat{Y}_t = \circledR_0 + \circledR_1 X_{3t} + e_t$
Di mana $X_{3t} = X_{1t} + 0{,}10\ X_{2t}$

---

[27]Wonacott;2010:198

(2) Mengkombinasi data *cross section* dan *time series* (kombinasi ini disebut *pooled data*)[28].

Contoh :

$$\text{Ln } \hat{Y}_t = \circledR_0 + \circledR_1 \ln P_t + \circledR_2 \ln I_t + e_t \qquad \text{. . . . . (9.10)}$$

Di mana $\hat{Y}_t$ = penjualan mobil dalam unit,

P_t = rata-rata harga mobil,

I_t = pendapatan.

Diperkirakan bahwa P_t dan I_t berkorelasi linier, dengan data *cross section*,maka P_t biasanya tidak banyak bervariasi (karena pengamatan pada suatutitik waktu saja). Model regresi dapat diubah sebagai :

$$\hat{Y}_t^* = \circledR_0 + \circledR_1 \ln P_t + e_t \qquad \text{. . . . . (9.11)}$$

Di mana,

$\hat{Y}_t^* = \ln \hat{Y}_t - \circledR_2 \ln I_t$, artinya model ini mencoba menghilangkan pengaruh pendapatan (I_t), $\circledR_2$ yang digunakan untuk mengkoversi $\hat{Y}_t^*$ adalah koefisien pada persamaan-9.10.

(3) Membuang sebuah variabel atau lebih, dan bias spesifikasi.

Membuang variabel yang ditengarai multikol dengan variabel lainnya ini cara termudah, tetapi mungkin timbul permasalahan baru, yaitu: *specification bias* atau *specification error*, karena bertentangan dengan teori yang telah mantap. Untuk itu cara ini merupakan cara yang tidak populer walaupun paling mudah.

(4) Mentransformasi variabel.

Model : $\hat{Y}_t = \circledR_0 + \circledR_1 X_{1t} + \circledR_2 X_{2t} + e_t$

[28]Tobin;1970:113

Dirubah menjadi :

$\hat{Y}_t – \hat{Y}_{t-1}$ = $\circledR_1(X_{1t} - X_{1t-1}) + \circledR_2 (X_{2t} - X_{2t-1}) + v_t$

Di mana, $v_t = e_t – e_{t-1}$

(5) Menambah data atau menggunakan data lebih banyak yang sama sekali baru (ukuran sampel ditingkatkan).

Dengan memperbanyak data, var($\circledR_i$) akan menurun, sehingga estimasi $\circledR_1$ akan lebih jelas dan lebih akurat.

b. Hetetroskedastis.

Heteroskedastis terjadi jika ekspektasi e_t^2 : E{e_t^2}≠ ⎧^{2}. Varians e_t adalah varians Y_t, tidak lagi konstan. Untuk mendeteksi adanya heteroskedastis pada model dapat dilakukan dengan salah satu cara berikut ini:

(1) Apakah ada var{e_t} terlihat berkorelasi dengan variabel-variabel independen. Jika ada, maka ini mengindikasikan adanya heteroskedastis.

(2) Metode grafis. Melalui plot atau *scatter diagram* antara e_t^2 (sumbu Y) dengan$\hat{Y}$ (sumbu X). Jika terlihat ada pola sistimatis pada plot, maka diperkirakan terjadi heteroskedastis. Cara ini tidak terlalu disarankan, karena tergantung kepada ketelitian penglihatan mata saja.

(3) Park Test[29].

Park memformalkan metode grafik dengan menganggap bahwa ⎧$_i^2$ merupakan fungsi dari variabel eksplanatori X_i.

$$\sigma_i^2 = \sigma^2 X_i^{\beta} e^{ui}$$

Atau:

ln ⎧$_t^2$ = ln ⎧$_t^2$ + $\circledR_1$ ln X_{it}+ e_t

[29] Park, R.E:1966:888

di mana e_t = *stochastics disturbance*.

Karena umumnya, $\int_t^2$ tidak diketahui, Park menyarankan menggunakan e_t^2 sebagai proksi, dan menyebabkan fungsi regresi menjadi:

$\ln e_t^2 = \ln \int_t^2 + \circledR_1 \ln X_{it} + u_t$

$= \circledR_0 + \circledR_1 \ln X_i + u_t$

Jika $\circledR_1$ signifikan, maka terjadi heteroskedastis pada datanya, dan jika $\circledR_1$ tidak signifikan, maka tidak terjadi heteroskedastis atau disebut sebagai homoskedastis.

Park test merupakan prosedur dua tahap. Tahap pertama,aplikasikan metode OLS pada regresi dengan tidak menghiraukan apakah adaheteroskedastis atau tidak. Diperoleh nilai e_t, kemudian pada tahap kedua mengaplikasi regresi: $\ln e_t^2 = \ln \int_t^2 + \circledR_1 \ln X_{it} + u_t$.

Park test masih memiliki beberapa masalah, Goldfeld dan Quandt[30] menyatakan bahwa jika u_t dimasukkan ke dalam persamaan tersebut di atas, mungkin tidak sesuai dengan asumsi OLS dan bisa terjadi heteroskedastis pada u_t itu sendiri.

Namun demikian, Park test masih bisa digunakan. Untuk mengilustrasikan Park test, bisa dilihat contoh berikut:

Model regresi : $\hat{Y}_t = \circledR_0 + \circledR_1 X_t + e_t$; di mana Y rata-rata gaji (dalam \$) dan X = rata-rata produktivitas (dalam \$).

Hasil analisis menunjukkan bahwa:

$\hat{Y}_t = 1999{,}0466 + 0{,}2323\ X_t + e_t$

t (2,323)

$R^2 = 0{,}4356$

[30]Goldfeld dan kawan-kawan, 1966:93-94

®$_1$ signifikan pada 〈 = 0,05. Selanjutnya regresikan e_t terhadap X_{it}, hasilnyaadalah :

$\ln e_t^2 = 35{,}9010 - 2{,}8099 \ln X_t + u_t$

t (-0,667)

$R^2 = 0{,}0595$

®$_1$ tidak signifikan pada 〈 = 0,05. Maka dapat disimpulkan tidak terjadi heteroskedastis.

(4) Glejser Test[31].

Glejser test mirip dengan Park test. Setelah memperoleh e_t seperti pada Park test, maka Glejser meregres nilai absolut e_t terhadap X_t yang diperkirakan memiliki hubungan dengan ⎧$_t^2$. Bentuk-bentuk model regresinya adalah salah satu dari bentuk ini:

(a) $| e_t | = \beta_1 X_t + u_t$

(β) $| e_t | = \beta_1 \sqrt{X_t} + u_t$

(c) $| e_t | = \beta_1 (1/X_t) + u_t$

(d) $| e_t | = \beta_1 (1/\sqrt{X_t}) + u_t$

(e) $| e_t | = \beta_0 + \beta_1 X_t + u_t$

(f) $| e_t | = \sqrt{\beta_0 + \beta_1 X_t + u_t}$

(g) $| e_t | = \sqrt{\beta_0 + \beta_1 X_t^2 + u_t}$

Glejser test masih memiliki beberapa masalah. Goldfeld dan Quandt[32] juga menyatakan bahwa jika u_t dimasukkan ke dalam persamaan tersebut di atas, mungkin tidaksesuai dengan asumsi OLS dan bisa terjadi heteroskedastis pada u_t itu sendiri. Glejser test boleh dicoba walaupun tidak disarankan.

---

[31] Glejser;1996:316-323.

[32] Goldfeld, dan kawan-kawan;1996:93-94

(5) Spearman's *Rank Correlation Test*.

Mengkorelasikan $|e_t|$ dengan variabel X_t dengan korelasi rank Spearman (r_s), jika koefisien korelasinya signifikan pada 〈 penelitian, maka pada model terindikasi terjadi heteroskedastis. Signifikansi koefisien korelasi dapat diuji dengan t-test:

$$t = \frac{r_s\sqrt{(n-2)}}{\sqrt{(1-r_s^2)}}$$, r_s signifikan jika $t > t_{df,\langle}$ . . . . . (9.12)

Untuk menghilangkan heteroskedastis:

(a) Jika ⌠$_i^2$diketahui : gunakan *weighted least squares* (WLS).

Data asli Y_t dan X_{it} dibagi dengan 1/⌠$_i^2$. Data asli menjadi Y_t* dan X_{it}*, kemudian hitung $x_{it}^* = X_{it} - X_{it}^*$ dan $y_{it}^* = Y_t - Y_t^*$. Data ini kemudian diregresikan.

(b) Jika ⌠$_i^2$tidak diketahui : WLS tidak tepat digunakan.

Tranformasi data dan pilih salah satu model dengan format:

$Y_t/X_t = \beta_0/X_{jt} + \beta_1 + e_t/X_{jt}$, atau

$Y_t/\sqrt{X_t} = \beta_0/\sqrt{X_{jt}} + \beta_1\sqrt{X_{jt}} + e_t/\sqrt{X_{jt}}$, atau

$Y_t/E\{Y_t\} = \beta_0/ E\{Y_t\} + \beta_1 X_{jt}/ E\{Y_t\} + e_t/ E\{Y_t\}$, atau

$Y_t/\hat{Y}_t = \beta_0/\hat{Y}_t + \beta_1(X_{jt}/\hat{Y}_t) + e_t$, atau dengan transformasi log menjadi:

$\ln Y_t = \beta_0 + \beta_1 \ln X_{jt} + e_t$

c. Otokorelasi

Otokorelasi adalah korelasi antar data antar waktu (*time series data*) maupun antar ruang (*cross section data*). Jika otokorelasi terjadi, maka:

(1) Estimasi koefisien regresai menjadi tidak *best linear unbiased estimator*(BLUE).

(2) Jika tetap mengaplikasi OLS, walaupun terjadi otokorelasi, maka: estimasi σ^2 *underestimate*, atau jika tidak demikian ada kemungkinan varians atau *standard error* estimator OLS yang *underestimate*, dan nilai-t dan nilai-F menjadi tidak valid.

Untuk mendeteksi adanya otokorelasi dapat dipilih salah satu cara:

(1) Metode Grafis.

Cara ini adalah dengan membuat grafik garis e_i dari pengamatan yang berurutan. Jika garis itu membentuk suatu pola tertentu, maka diperkirakan terjadi otokorelasi pada model regresi. Jika polanya acak, maka dapat diperkirakan tidak terjadi otokorelasi.

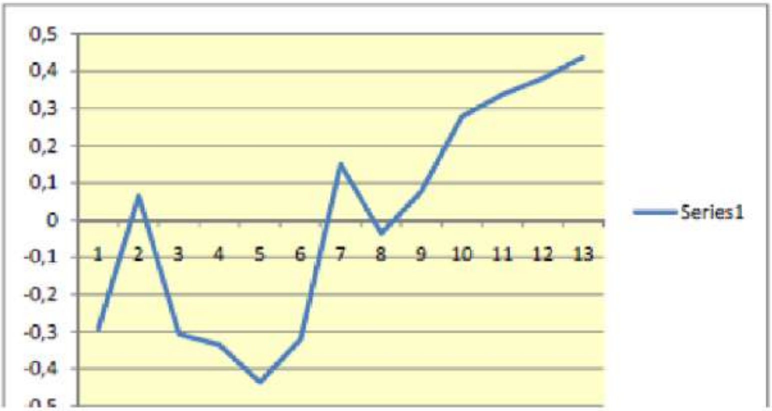
<u>Contoh 9.7</u>: Otokorelasi

Tahun	Y_t	X_t	e_t	e_t^2	$e_t - e_{t-1}$	$(e_t - e_{t-1})^2$
1991	1,3	1,592	-0,292	0,085	-	0,358
1992	1,2	1,134	0,066	0,004	0,358	-0,372
1993	1,4	1,706	-0,306	0,094	-0,372	-0,029
1994	1,4	1,735	-0,335	0,112	-0,029	-0,100
1995	1,5	1,935	-0,435	0,189	-0,100	0,114
1996	1,9	2,221	-0,321	0,103	0,114	0,471
1997	2,6	2,450	0,15	0,022	0,471	-0,186
1998	2,3	2,336	-0,036	0,001	-0,186	0,114
1999	2,5	2,422	0,078	0,006	0,114	0,200
2000	2,7	2,422	0,278	0,077	0,200	0,059
2001	2,1	1,763	0,337	0,114	0,059	0,043
2002	1,8	1,420	0,380	0,144	0,043	0,057
2003	2,2	1,763	0,437	0,191	0,057	0,729
Σ				1,144		0,358

Hasil analisis regresi :

$\hat{Y}_t = 0{,}000134 + 0{,}9997\ X_t + e_t$

Jika $e_t = Y_t - \hat{Y}_t$ tersebut diplot dalam grafik garis berdasar tahun pengamatan, hasilnya adalah:



Gambar 9.5 Grafik Garis e_t

Grafik garis e_t tidak memiliki pola tertentu, maka diperkirakan tidak terjadi otokorelasi. Namun metode ini tidak disarankan, mengingat sangat tergantung kepada ketelitian mata.

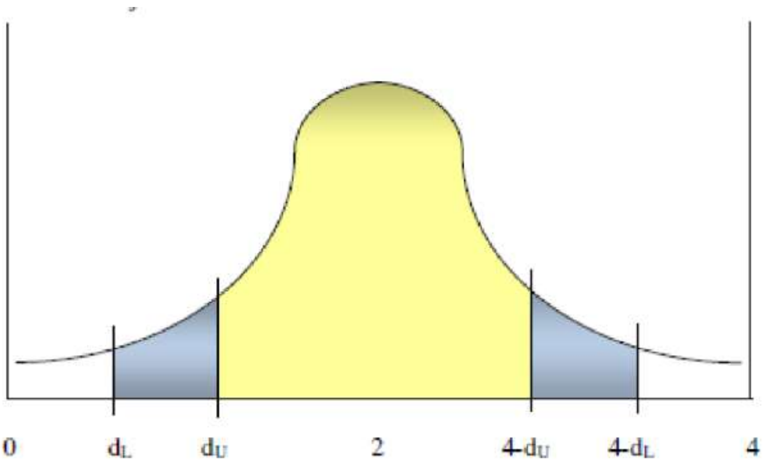
(2) Durbin-Watson Test.

Statistik Durbin-Watson diformulakan sebagai[33] :

$$d = \frac{\Sigma\,(e_t - e_{t-1})^2}{\Sigma\, e_t^2} \qquad \ldots\ldots (9.13)$$

Kriteria uji adalah:

[33]Durbin dan kawan-kawan, 1951:159



Gambar 9.6 Daerah Uji Otokorelasi

Daerah berarsir adalah daerah di mana ada keraguan apakah menerima hipotesisnol atau menolaknya, d_L = *lower limit*, d_U = *upper limit*. Nilai d_L dan d_U pada 〈 tertentu,tersedia pada Tabel Durbin-Watson.

Ada tiga jenis uji otokorelasi[34], yaitu :

(a) Otokorelasi positif,

H_0 : tidak ada otokorelasi positif,

H_a : ada otokorelasi positif,

Kriteria :

- jika $d < d_L$: tolak H_0,
- jika $d > d_U$: terima H_0,
- jika $d_L < d < d_U$: tidak ada keputusan pasti.

(b) Otokorelasi negatif,

H_0 : tidak ada otokorelasi negatif,

H_a : ada otokorelasi negatif,

[34]Gujarati;2005:237-242

Kriteria :

- jika $d > 4 - d_L$: tolak H_0,
- jika $d < 4 - d_U$: terima H_0,
- jika $4- d_U < d < 4 - d_L$: tidak ada keputusan pasti.

(c) Otokorelasi positif/negatif,

H_0 : tidak ada otokorelasi negatif/positif,

H_a : ada otokorelasi negatif/positif,

Kriteria :

- jika $d < d_L$: tolak H_0,
- jika $d > 4 - d_L$: tolak H_0,
- jika $d_U < d < 4 - d_U$: terima H_0.
- jika $d_L < d < d_U$ atau - jika $4 - d_U < d < 4 - d_L$ } tidak ada keputusan pasti

Pada contoh ini, uji otokorelasi dilakukan bisa dilakukan melalui salah satu uji :

(a) Uji otokorelasi positif: H_a = ada otokorelasi positif.

n = 13, k = banyaknya variabel independen = 1

d = 0,358/1,114 = 0,321

Untuk k = 1 dan n = 13, d_L = 1,08 dan d_U = 1,36 (periksa Tabel Durbin-Watson). Mengingat $d < d_L$ maka tolak H_0 dan terima H_a, artinya ada otokorelasi positif.

(b) Uji otokorelasi negatif: H_a = ada otokorelasi negatif.

d = 0,321;4 – dL = 4 – 1,08 = 2,92; 4 – dU = 4 – 1,36 = 2,64;

Mengingat d < 2,92 dan d < 2,64; maka terima H_0, artinya tidak ada otokorelasi negatif.

(c) Uji otokorelasi positif/negatif :
$d = 0{,}321$; $d_L = 1{,}08$. Karena $d < d_L$, tolak H_0 (ada otokorelasi positif), atau $d < 4 – 1{,}08$, maka terima H_0 (tidak ada otokorelasi negatif).

Cara perbaikan jika terjadi otokorelasi:

(a) Jika koefisien otokorelasi (r) diketahui.

Model awal : $Y_t = \beta_0 + \beta_1 X_t + e_t$ . . . . .(9.14)

Model lag : $Y_{t-1} = \beta_0 + \beta_1 X_{t-1} + e_{t-1}$ . . . . .(9.15)

Kalikan model lag dengan r :

$r\,Y_{t-1} = r\,\beta_0 + r\,\beta_1 X_{t-1} + r\,e_{t-1}$ . . . . .(9.16)

Kurangi persamaan (9.16) dengan persamaan (9.14):

$Y_t - r\,Y_{t-1} = (1 - r)\,\beta_0 + (\beta_1 X_t - r\,\beta_1 X_{t-1}) + (e_t – r\,e_{t-1})$

$Y_t - r\,Y_{t-1} = (1 - r)\,\beta_0 + \beta_1 (X_t - r\,X_{t-1}) + e_t$
. . . . .(9.17)

Buatlah analisis regresi dengan model persamaan (9.17) ini.

(b) Jika r tidak diketahui.

Salah satu cara perbaikan yang bisa dipilih adalah :

(b.1) *The first difference method*.

Model dimodifikasi menjadi :

$\otimes Y_t = \beta_1 \otimes X_t + \sum_t$, atau bentuk *moving average regression*:

$$\frac{Y_t + Y_{t-1}}{2} = ®_0 + ®_1 \frac{X_t + X_{t-1}}{2} + \frac{\sum_t}{2}$$

(b.2) r berbasis Durbin-Watson.

Hitung terlebih dahulu estimasi r dengan formula[35] :

[35]Theil dan kawan-kawan, 1991: 793-806.

$$\hat{r} = \frac{n^2(1 - d/2) + k^2}{n^2 - k^2} \qquad \text{......(9.18)}$$

Kemudian lakukan analisis regresi dengan model persamaan (9.19):

$Y_t – (\hat{r}\ Y_{t-1}) = (1 - \hat{r})\ ®_0 + ®_1\ (X_t - \hat{r}\ X_{t-1}) + e_t$

Catatan : data X dan Y dikalikan terlebih dahulu dengan $\sqrt{1 - \hat{r}^2}$.

<u>Contoh 9.8</u>: Remedial Otokorelasi berbasis Durbin-Watson[36]

Model regresi : $\ln HWI_t = ®_0 + ®_1 \ln U_t + u_t$

Di mana HWI = *Help-Wanted Index*, dan U = *Unemployment Rate*

Data:

Tabel 9.7 Data HWI dan U.

Tahun	HWI	U	Ln HWI	Ln U
1982-1	104,66	5,63	4,65	1,73
2	103,53	5,46	4,64	1,70
3	97,30	5,63	4,58	1,73
4	95,96	5,60	4,56	1,72
1983-1	98,83	5,83	4,59	1,76
2	97,23	5,76	4,58	1,75
3	99,06	5,56	4,60	1,72
4	113,66	5,63	4,73	1,73
1984-1	117,00	5,46	4,76	1,70
2	119,66	5,26	4,78	1,66
3	124,33	5,06	4,82	1,62
4	133,00	5,06	4,89	1,62

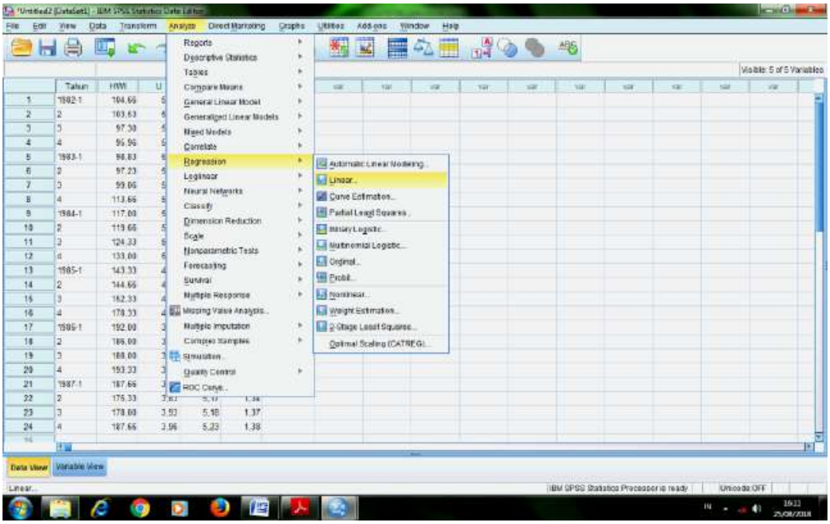
[36] Gujarati;2005:335

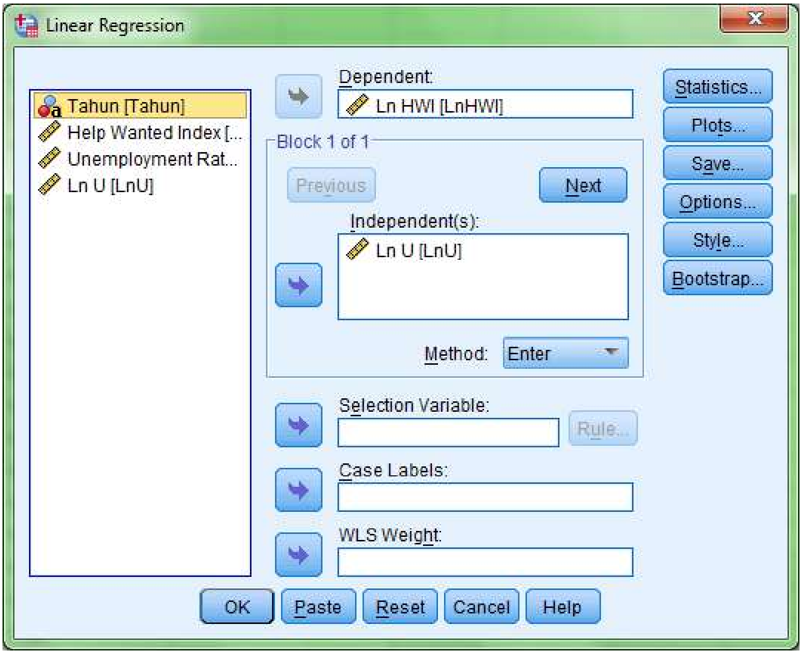
1985-1	143,33	4,83	4,97	1,57
2	144,66	4,73	4,97	1,55
3	152,33	4,46	5,03	1,50
4	178,33	4,20	5,18	1,44
1986-1	192,00	3,83	5,26	1,34
2	186,00	3,90	5,23	1,36
3	188,00	3,86	5,24	1,35
4	193,33	3,70	5,26	1,31
1987-1	187,66	3,66	5,23	1,30
2	175,33	3,83	5,17	1,34
3	178,00	3,93	5,18	1,37
4	187,66	3,96	5,23	1,38

Langkah-langkah aplikasi SPSS untuk data tersebut adalah:

(a) Ketik data secara langsung dengan SPSS atau dengan Excel. Jika dengan Excel, maka file data harus dibaca oleh SPSS untuk dikonversi dengan format SPSS.

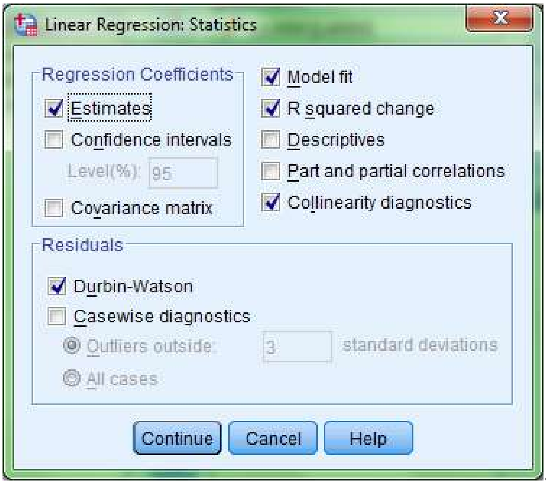
(b) Pilih '*Analyze - Regression – Linear*'





(c) Isikan

'Ln HWI' pada *Dependent*.
'Ln U' pada Independent(s).
Pilih '*Statistic*'



(d) Centang ‘*Model Fit’*, ‘*R squared change’*, ‘*Collinearity diagnostics’* dan ‘*Durbin-Watson*’. Klik ‘*Continue*’.

(e) Setelah menekan Ok ; printout hasil analisis regresi dengan SPSS yang dihasilkan adalah ::

Regression

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Ln U[b]	.	Enter

a. Dependent Variable: Ln HWI

b. All requested variables entered.

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,978[a]	,956	,954	,05760	,878

ANOVA[a]

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	1,579	1	1,579	475,914	,000[b]
	Residual	,073	22	,003		
	Total	1,652	23			

a. Dependent Variable: Ln HWI

b. Predictors: (Constant), Ln U

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	7,311	,110		66,382	,000		
	Ln U	-1,539	,071	-,978	-21,815	,000	1,000	1,000

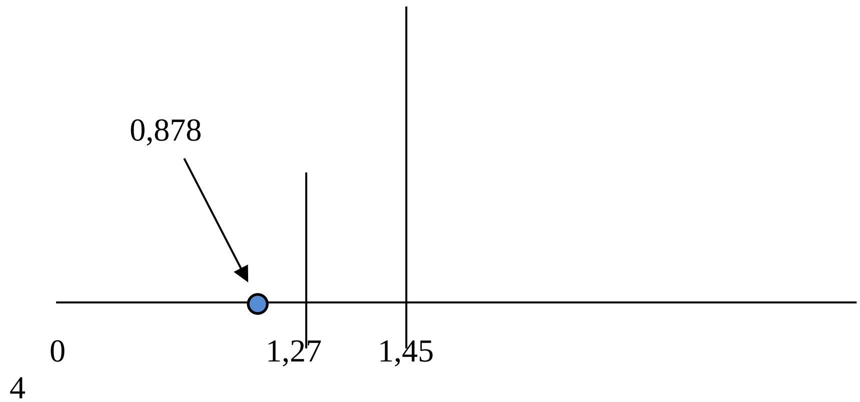
a. Dependent Variable: Ln HWI

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions	
				(Constant)	Ln U
1	1	1,994	1,000	,00	,00
	2	,006	18,680	1,00	1,00

a. Dependent Variable: Ln HWI

Walaupun konstanta dan koefisien regresi, keduanya signifikan, namun nilai Durbin-Watson = 0,878 mengindikasikan terjadi otokorelasi positif.



Pada k = 1, dan n = 24, d_L = 1,27 dan d_U = 1,45 (periksa Tabel Durbin-Watson).

Estimasi koefisien korelasi dengan formula (9.18):

$$= \frac{(24^2)(1 - 0{,}911/2) + (2^2)}{(24^2) - (2^2)} = 0{,}555$$

Untuk menghilangkan gejala otokorelasi ini, data ditransformasi dengan (ln HWI_t – 0,555 ln HWI_{t-1}) dan (ln U_t – 0,555 ln U_{t-1}). Sedang data pertama ditransformasi dengan $\sqrt{\{(1 - 0{,}555^2)\ln HWI_1\}}$ dan $\sqrt{\{(1 - 0{,}555^2)\ln U_1\}}$. Kemudian lakukan analisis regresi ulang dengan data hasil transformasi tersebut.

Printout SPSS:

Regression

7

Model	Variables Entered	Variables Removed	Method
1	Ln Ut - Ln U t-1[b]	.	Enter

a. Dependent Variable: Ln HWIt - Ln HWI t-1

b. All requested variables entered.

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,934[a]	,872	,866	,05485	1,482

ANOVA[a]

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	,451	1	,451	150,043	,000[b]
	Residual	,066	22	,003		
	Total	,518	23			

a. Dependent Variable: Ln HWIt - Ln HWI t-1
b. Predictors: (Constant), Ln Ut - Ln U t-1

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	3,039	,070		43,386	,000		
	Ln Ut - Ln U t-1	-1,217	,099	-,934	-12,249	,000	1,000	1,000

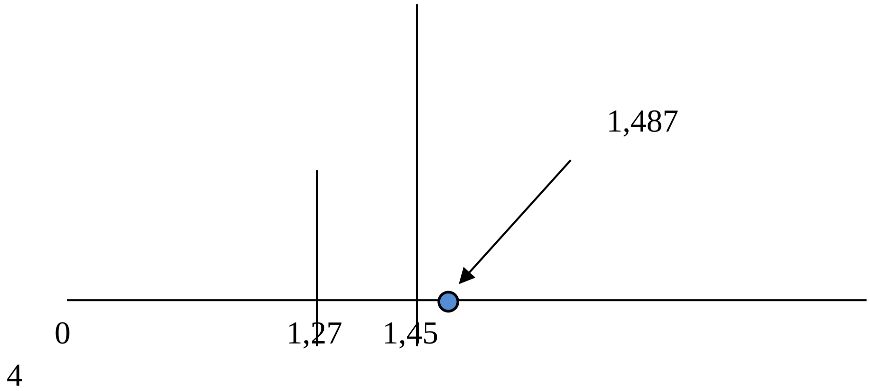
a. Dependent Variable: Ln HWIt - Ln HWI t-1

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions	
				(Constant)	Ln Ut - Ln U t-1
1	1	1,987	1,000	,01	,01
	2	,013	12,433	,99	,99

a. Dependent Variable: Ln HWIt - Ln HWI t-1

Seluruh koefisien tetap signifikan dan nilai Durbin-Watson sekarang menjadi = 1,482.



Gambar di atas menjelaskan bahwa tidak terindikasi lagi adanya otokorelasi pada regresi yang dihasilkan.

9.10. Rangkuman

Analisis Regresi Linier Berganda merupakan alat analisis statistik yang banyak digunakan dalam penelitian bidang ilmu sosial.

Bentuk umum fungsi regresi:

$\hat{Y}_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \ldots\ldots + \beta_k X_{kt} + e_t$

Formula untuk menghitung koefisien regresi (dengan operasi matriks):

$$\underset{(k \times 1)}{\beta} = \underset{(k \times k)}{(X'X)^{-1}} \; \underset{(k \times n)}{X'Y} \; \underset{(n \times 1)}{e}$$

k = kolom, n = banyaknya pengamatan

Uji signifikansi koefisien regresi secara parsial:

$$t\beta_i = \frac{\beta_i}{SE\{\beta_i\}} \quad \text{(untuk i = 0, 1, . . . k)} \quad \ldots\ldots (9.4)$$

Di mana $SE\{\beta_i\}$ adalah standard error untuk β_i.

$$SE\{\beta_0\} = SE\{\text{Regresi YX}\} + \sqrt{\frac{1}{n} \quad \frac{X^2}{\Sigma (X_t - \bar{X})^2}} \quad \ldots\ldots (9.5)$$

$$SE\{\beta_1\} = \frac{SE\{\text{Regresi YX}\}}{\Sigma(X_t - \bar{X})^2} \quad \ldots\ldots (9.6)$$

$$SE\{\text{Regresi YX}\} = \sqrt{SSE/(n-2)} \quad \ldots\ldots (9.7)$$

Uji signifikansi koefisien regresi secara simultan:

$$F = \frac{MSR}{MSE} = \frac{SSR/(k-1)}{SSE/(n-2)} = (n-2)\frac{SSR}{SSE} \quad . . . . . (9.8)$$

Ukuran kesesuaian model: Koefisien Determinasi (R^2):

$$R^2 = \frac{\beta' X' y - n\overline{Y}^2}{y' y - n\overline{Y}^2}$$

Untuk menghitung *adjusted* R^2 dengan formula:

$$\overline{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k}$$

Adjusted R^2 ditujukan untuk membandingkan kesesuaian model-model regresi dengan variabel dependen yang sama, namun berbeda dalam banyaknya variabel independen.

Asumsi Klasik Regresi:

1. Rata-rata kondisional dari penyimpangan populasi, $e_t = 0$,
2. Varian e_t = konstan, atau homoskedastis (non heteroskedastis),
3. Tidak terjadi otokorelasi serial antar e_t,
4. Variabel eksplanatori (baik skokastik maupun non skokastik),
5. Tidak terjadi multikolinieritas antar variabel eksplanatori,
6. e_t berdistribusi normal dengan rata-rata = 0, dan standard deviasi = 1.

Dari keenam asumsi ini, tiga merupakan syarat perlu (perlu diuji), yaitu: otokorelasi, multikolinieritas dan homoskedastis, dan tiga lainnya merupakan syarat cukup (boleh diuji dan tidak).

9.11. Diskusi

Aplikasi analisis regresi biasanya diawali dengan model regresi yang berbasis teori (khususnya strategi permodelan konfirmatori). Kondisi ini menyebabkan banyak peneliti yang tidak menghasilkan model regresi baru. Namun untuk peneliti pemula, hal ini masih dapat ditoleransi. Bagi peneliti tingkat lanjut, selayaknya menggunakan strategi eksploratori, untuk menghasilkan model regresi baru. Penggunaan *bootstrap*, *trimming*, dan metode *step wise*, dapat diaplikasi untuk mengelimir koefisien yang tidak signifikan.

9.12. Referensi

Beck, Michael s. Lewis, 2007, ***Regression Analysis***, SAGE Publications Toppan Publishing, London.

Frisch, Ragnar, 1952, ***Statistical Confluence Analysis by Means of Complete Regression System***, Oslo University Publishing, Oslo.

Gujarati, Damodar, 2005, ***Basic Econometrics***, 5th Edition, McGraw Hill Book Company, Tokyo.

Pindyck, Robert, S., and Daniel L. Rubinfeld, 2005. ***Econometrics Model andEconomic Forecasts***, 5th ed. Irwin/McGraw Hill Book Company, NewYork.

9.13. Latihan Soal

Sebuah penelitian bertujuan memperkirakan apakah ada pengaruh tingkat pengunduran diri dari pekerjaan (X) terhadap tingkat pengangguran di Propinsi Jawa Timur. Data dikumpulkan dari Kantor Wilayah Dinas Tenaga Kerja Propinsi Jawa Timur pada Tahun 2017 adalah:

Tahun	Tingkat Pengunduran Diri (per 100 buruh)	Tingkat Pengangguran (%)
2004	1,3	6,2
2005	1,2	7,8
2006	1,4	5,8
2007	1,4	5,7
2008	1,5	5,0
2009	1,9	4,0
2010	2,6	3,2
2011	2,3	3,3
2012	2,5	3,6
2013	2,7	3,3
2014	2,1	5,6
2015	1,8	6,8
2016	2,2	5,6

a. Buat *scatter diagram* dari data tersebut.
b. Buat analisis regresi lengkap dengan uji hipotesisnya.
c. Interpretasikan hasil analisis regresi.

Sebuah fungsi produksi yang dikembangkan Cobb-Douglas adalah:

$Y_t = ®_0 X_{1t}^{®1} X_{2t}^{®2}$ tampak seperti fungsi regresi eksponensial, buatlah bentuk fungsi regresi liniernya.

Sebuah model regresi yang dihasilkan dalam sebuah penelitian adalah:

$\hat{Y}_t =$ 0,3243 + 1,0332 X_{1t}+ 0,0234 X_{2t} + e_t

SE{®$_i$} (0,1343) (0,5111) (0,1356)

R^2 0,6558

F 132,456

a. Berapa nilai-t konstanta dan koefisien regresi?
b. Apakah pengaruh X_{it} signifikan terhadap Y_t pada 〈 = 5,00%?
c. Berapakah persen perubahan Y_t dapat dijelaskan oleh perubahan X_{it}?

Data pendapatan bersih, jam kerja buruh, dan penghasilan sampingan pada 15 buah pabrik kopi bubuk adalah:

Pabrik	Pendapatan Bersih per bulan (Rp juta)	Jam Kerja Buruh (jam/bulan)	Penghasilan Sampingan per bulan (Rp juta)
1	320,3	200	4,3
2	154,6	205	5,4
3	222,8	198	4,3
4	156,8	205	3,4
5	164,2	210	4,3
6	232,9	188	6,5
7	215,8	204	5,3
8	123,6	189	4,4
9	212,7	202	3,6
10	333,2	208	2,5

11	112,3	210	3,4
12	224,4	212	4,7
13	413,5	199	4,5
14	225,3	201	5,4
15	313,2	188	6,6

a. Variabel apa yang menurut anda diposisikan sebagai variabel dependen?
b. Buat fungsi regresi linier berganda, dan lakukan analisisnya.
c. Bagaimana pengaruh kedua variabel independen berkaitan dengan signifikansi parsialnya?
d. Berapa R^2 dan *adjusted* R^2?
e. Interpretasikan hasil analisis regresi.
f. Buat analisis regresi linier berganda dengan aplikasi Excel

DAFTAR PUSTAKA

Agresti, Alan, ***Catagorical Data Analysis***, 2002, John Wiley and Sons Inc., Tokyo.

Altman, F. F., 1972, ***Distributed Lags***, Econometrica, McGraw Hill Book Company, Tokyo.

Anderson, T. W., 1998, ***Introduction to Multivariate Statistical Analysis***, John Wiley and Sons, New York (ATW), 1988.

Balakrishnan, N., and V. B. Nevzorov, 2003, ***A Primer on Statistical Distributions***, John Wiley and Sons Inc., New York.

Beck, Michael s. Lewis, 2007, ***Regression Analysis***, SAGE Publications Toppan Publishing, London.

Belsley, D. A., E. Kuh, and R. E. Welsch. 2000. ***Regression Diagnostics: Identifying Influential Data And Sources Of Collinearity***. New York: John Wiley and Sons.

Bowen, Earl, K., and Martin K. Starr, 2002, ***Basic Statistics for Business and Economics,*** McGraw Hill Book Company, Tokyo.

Brunk, H. D., ***An Introduction to Mathematical Statistics***, 5th Edition, Englewood Cliff, New Jersey, 2002/

Cagan, P., 1956, ***The Monetary Dynamics of Hyper Inflations, Studies on theQuantity Theory of Money***, Univeristy of Chicago Press. Chicago.

Cohen, J, ***Statistical Power Analysis For the Behavioral Sciences***, rev. ed. New York, Academic Press., 1997.

Cronbach, L. J. 1991. ***Coefficient Alpha And The Internal Structure Of Tests***. *Psychometrika*, 16:3, 297-334.

Dempster, A. P. 1999. ***Elements of Continuous Multivariate Analysis***. Reading, MA:Addison-Wesley.

Doane, David, P., and Lori E. Seward, 2007, ***Applied Statistics in Business and Economics***, McGraw Hill/Irwin Series, New York.

Friedman, Milton, 1957, ***A Theory of the Consumption Function***, National Bureau of Economic Research, Princeton University Press, New Jersey.

Frisch, Ragnar, 1952, ***Statistical Confluence Analysis by Means of Complete Regression System***, Oslo University Publishing, Oslo.

Gessner, Guy, N. K. Maholtra, W.A. Kamakura and M.E Smijewski, 188, ***Estimating Models with Binary Dependent Variables: Some Theoretical and Empirical Observations***, Journal of Business Research 16(1),

Green, Samual, B., Neil J. Salkind, dan Theresa M. Akey, 2001. ***USING SPSS FORWINDOWS : Analyzing and Understanding Data***, Prentice Hall, Upper Saddle River, New Jersey.

Gujarati, Damodar, 2005, ***Basic Econometrics***, 5th Edition, McGraw Hill Book Company, Tokyo.

_________, 2008, ***Basic Econometrics***, 6th Edition, McGraw Hill Book Company, NewYork, (DM).

Freund, John E., and Irwin Miller, 2004, ***Probability and Statistics for Engineers***, Fifth Edition, Prentice-Hall of India, Private Limited, New Delhi.

Hald, A., 1992, ***Statistical Theory with Engineering Applications***, John Wiley & Sons, New York.

Hair, Jr., Joseph F.; Rolph E. Anderson, Ronald L. Tatham and William C. Black, 2001; ***Multivariate Data Analysis***, 6^{th} Edition, Prentice Hall International, Inc., Upper Saddle River, New Jersey.

Hines, William, H., and Douglas C. Montgomey, 2000, ***Probability and Statistics for Engineers in Engineering and Management Science,*** John Wiley and Sons, Inc., New York.

Hoel, P., ***Introduction to Mathematical Statistics***, 2005, 7^{th} Ed. John Wiley & Sons, Inc., NewYork.

Huber, Peter, J., 2003, ***Robust Statistics***, John Wiley and Son Coy., Tokyo.

Kendall, M. G., and Stuart, A., 2005, ***The Advanced Theory of Statistics***, Vol. 1 6^{th} Edition, McMillan Publishing Company, NewYork, 2005.

Kirkpatrick, E., G., 2009, ***Introductory Statistics and Probability for Engineer, Science and Technology,*** Prentice-Hall Inc., Englewood Cliffs, N. J.

Kristof, W. 1963. ***The Statistical Theory Of Stepped-Up Reliability Coefficients When A Test Has Been Divided Into Several Equivalent Parts***. *Psychometrika*, 28:3,221-238.

Liviatan, N., 1981, ***Consistent Estimation of Distributed Lags***, International Economic Review, vol. 4, Januari.

__________, 1982, ***Instumental Variables in Simultaneous Equations Model***, John Wiley and Son Inc., Tokyo.

Miller, Irwin and John E. Freund, 2008, ***Probability And Statistics For Engineers,***7th Edition, Prentice-Hall of India Private Limited, New Delhi.

Muller, K. E., and B. L. Peterson. 2004. ***Practical Methods For Computing Power In Testing The Multivariate General Linear Hypothesis***. ***Computational Statisticsand Data Analysis***, 2, 143-158.

Novick, M. R., and C. Lewis. 1997. ***Coefficient Alpha And The Reliability Of Composite Measurements***. *Psychometrika*, 32:1, 1-13.

Pindyck, Robert, S., and Daniel L. Rubinfeld, 2005. ***Econometrics Model and Economic Forecasts***, 5th ed. Irwin/McGraw Hill Book Company, NewYork.

Setyawan, HB, 2009. Statistika (Deskriptif dan Diferensial), Perpustakaan Nasional:Katalog dalam Terbitan ISBN 978-602-98107-0-7, Yayasan Dharma Nusantara, Jember.

Shrout, P. E., and J. L. Fleiss. 1979. ***Intraclass Correlations: Uses In Assessing Reliability***. *Psychological Bulletin*, 86:, 420-428.

Supranto, J, 2008, Statistik, **Teori dan Aplikasi**, Penerbit Erlangga, Jakarta.

Tabachnick, Barbara G., and Linda S. Fidell, 1999, ***Using Multivariate Statistics***, Harper Collins Publisher, Inc., New York (TL).

Tinbergen, J., 1979, ***Longterm Foreign Trade Elasticities***, Metroeconomica, Volume 1.

Winer, B. J. 1971. ***Statistical Principles in Experimental Design***, 2nd ed. New York: McGraw-Hill Book Company.

INDEKS

GLOSARIUM

Analisis Korelasi

Analisis statistik untuk mengukur keeratan hubungan dua variabel berpasangan, di mana keduanya merupakan variabel independen; ada kesetaraan pada kedua variabel tersebut.

Analisis Regresi

Analisis statistik untuk mengukur besaran pengaruh variabel-variabel bebas terhadap sebuah variabel tak bebas. Berbeda dengan analisis korelasi, sebuah variabel diposisikan sebagai variabel tak bebas, yang diperkirakan nilainya dipengaruhi oleh variabel-variabel lainnya.

Analisis Regresi Bertahap (*Stepwise Regression Analysis*)

Stepwise regression adalah analisis regresi dengan melibatkan variabel bebas secara bertahap, dengan mempertimbangkan peningkatan nilai F yang secara subyektif dianggap memadai oleh analis. Hasilnya merupakan fungsi regresi dengan variabel-variabel bebas yang koefisien regresinya signifikan saja.

Data

Berasal dari kata Latin, *datum*. Data statistik adalah ukuran, informasi, nilai, atau karakter yang dapat menjelaskan suatu obyek yang diamati.

Data *time series*

Data *time series* adalah data yang menjelaskan sebuah kegiatan/ kejadian tertentu dari waktu ke waktu. Contoh, harga beras/kg pada sebuah pasar tradisional dalam satu bulan terakhir.

Data *cross section*

Data *cross section* adalah data yang menjelaskan beberapa obyek pada suatu tahun tertentu. Contoh, data harga berbagai saham di pasar modal pada akhir bulan Desember tahun 2017.

Diagram Pencar (*scatter plots*)

Diagram pencar merupakan grafik titik koordinat dari dua variabel, di mana sumbu-sumbunya adalah nilai kedua variabel yang diperkirakan memiliki hubungan.

Hetetroskedastis.

Heteroskedastis terjadi jika pada sebuah fungsi regresi, ekspektasi e_t^2 : $E\{e_t^2\} \neq \int^2$. Varians e_t adalah varians Y_t, tidak lagi konstan.

Hipotesis

Hipotesis adalah sebuah kesimpulan sementara yang perlu diuji kebenarannya melalui uji statistik. Pernyataan dalam hipotesis nol merefleksikan kenihilan, seperti: tidak berbeda atau sama, berpengaruh tidak signifikan (tidak efektif), pola data berdistribusi normal. Hipotesis alternatif merupakan komposit dari hipotesis nol, dan hubungan keduanya bersifat *mutually exclusive* (saling mematikan).

Kurtosis

Keruncingan kurve distribusi frekuensi data. Jika sebagian besar data terpusat pada nilai tertentu, maka kurve makin runcing; sebaliknya jika penyebaran datanya tidak terlampau ekstrim, maka kurve makin landai (tumpul).

Median

Median adalah posisi tengah data yang telah diurutkan berdasar besaran nilainya (dari kecil ke besar, *ascending*). Banyaknya data dengan nilai lebih rendah daripada median, sama dengan banyaknya data dengan nilai lebih tinggi daripada median.

Metode parametrik

Uji hipotesis dengan menggunakan parameter sampel, speerti rata-rata, varians, standard penyimpangan, kovarians, dan proporsi.

Metode non parametrik

Uji hipotesis tidak menggunakan parameter sampel, tetapi frekuensi data, dan mengabaikan asumsi-asumsi pada metode parametrik.

Modus

Modus berasal dari kata *mode,* artinya data yang paling sering muncul pada serangkaian data; atau data dengan frekuensi paling tinggi. Dari serangkaian data bisa terjadi uni modus, bi modus atau multi modus.

Multikolinieritas

Multikolinier merupakan adanya hubungan sempurna dan pasti antar beberapa variabel eksplanatori dalam sebuah model regresi.

Otokorelasi

Otokorelasi adalah korelasi antar data antar waktu (*time series data*) maupun antar ruang (*cross section data*).

Random Stratified Sample

Jika populasi dapat dibagi menjadi sub kelompok yang relative homogeny (ini disebut sebagai strata), dan pada setiap strata tersebut kemudian dipilih sampel secara acak.

Random Cluster Sample

Jenis sampel ini merupakan sampel terstrata, namun stratanya berdasar regional geografis.

Statistika Inferensial

Cabang Ilmu Statistika yang membahas pendugaan parameter populasi melalui parameter sampel, dan uji hipotesis.

Sturgest

Adalah pengembang cara membuat distribusi frekuensi untuk pengelompokan data. Metode Sturgest menetapkan banyaknya kelas secara pasti, tergantung pada nilai maksimum/minimum data.

Skewness

Kemiringan kurve distribusi frekuensi data. Jika frekuensi data lebih banyak dengan nilai yang kecil, maka frekuensi distribusinya miring ke kanan. Jika frekuensi data lebih banyak dengan nilai yang besar, maka frekuensi distribusinya miring ke kiri.

Tendency central value

Nilai kecenderungan data (pemusatan data, seperti rata-rata, modus, median dan dispersi).

Upper boundary dan Lower boundary

Batas atas adalah nilai tertinggi pada sebuah kelas,dan batas kelas bawah adalah nilai terendah pada sebuah kelas. Selisih kedua batas ini disebut sebagai rentang kelas (*class interval*).

TENTANG PENULIS

SYAFIADI RIZKI ABDILA, S.T., Ph.D. Beliau menyelesaikan pendidikan sarjana (S1) di Universitas Sultan Ageng Tirtayasa, Banten, Jawa Barat, Indonesia, dan menyelesaikan pendidikan Doktor (S3) di Universitas Malaysia Perlis. Beliau saat ini merupakan seorang dosen di program studi Teknik Sipil (TS), Fakultas Teknik (FT), Universitas Krisnadwipayana, DKI Jakarta, Indonesia. Terdapat beberapa Mata Kuliah yang diajarkan oleh beliau diantaranya yaitu Analisis Struktur II dan Menggambar Teknik dll. Email beliau adalah syafiadirizki@unkris.ac.id. Alamat kantor beliau adalah Kampus UNKRIS Jatiwaringin, Pondok Gede, DKI Jakarta, Indonesia.

Prof. Dr. Ir. Drs. SYAFWANDI, M.Sc. Beliau menyelesaikan pendidikan sarjana (S1) sebanyak 2 x yaitu di Universitas Indonesia dengan Gelar Insinyur (Ir.) dan di Universitas Muhammadiyah Jakarta dengan gelar Doktor Randes (Drs.). Kemudian beliau melanjutkan pendidikan S2 di Institut Teknologi Bandung (ITB) dengan gelar Magister Science (M.Sc.). Selanjutnya beliau melanjutkan pendidikan S3 di Universitas Satyagama Jakarta dengan gelar Doktor (Dr.). Beliau saat ini merupakan seorang Guru Besar (Profesor) di program studi

Manajemen, Fakultas Ekonomi dan Bisnis (FEB), Universitas Putra Indonesia YPTK Padang. Terdapat beberapa Mata Kuliah yang diajarkan oleh beliau diantaranya yaitu Manajemen Publik, Manajemen Industri, Manajemen Pemasaran, Statistik, Metode Penelitian dll. Alamat kantor beliau adalah di Jalan Raya Lubuk Begalung, Kota Padang, Provinsi Sumatera Barat, Indonesia.

Dr. YULASMI, S.E., M.M. Beliau menyelesaikan pendidikan sarjana (S1) di program studi Manajemen, Fakultas Ekonomi dan Bisnis (FEB), Universitas Putra Indonesia YPTK Padang, Sumatera Barat, Indonesia, dan menyelesaikan pendidikan magister (S2) di program studi Magister Manajemen (M.M.), Ekonomi dan Bisnis (FEB), Universitas Putra Indonesia YPTK Padang, Sumatera Barat, Indonesia. Kemudian melanjutkan pendidikan S3 di Program Studi Doktor Manajemen di Universitas Persada Indonesia (UPI) YAI Jakarta. Beliau saat ini merupakan seorang dosen di program studi Manajemen, Fakultas Ekonomi dan Bisnis (FEB), Universitas Putra Indonesia YPTK Padang. Terdapat beberapa Mata Kuliah yang diajarkan oleh beliau diantaranya yaitu Pengantar Bisnis, Ekonomi Bisnis, Manajemen Pemasaran dll. Email beliau adalah yulasmi@upiyptk.ac.id. Alamat kantor beliau adalah di Jalan Raya Lubuk Begalung, Kota Padang, Provinsi Sumatera Barat, Indonesia.

Dr. ROBBY DHARMA, S.E., M.M. Beliau menyelesaikan pendidikan sarjana (S1) di program studi manajemen, Sekolah Tinggi Ilmu Ekonomi (STIE) Dharma Andalas Padang, Sumatera Barat, Indonesia, dan menyelesaikan pendidikan magister (S2) di program studi Magister Manajemen (M.M.), Fakultas Ekonomi (FE), Universitas Putra Indonesia YPTK Padang, Sumatera Barat, Indonesia. Beliau saat ini sedang melanjutkan pendidikan Doktor (S3) di Program Studi Doktor Manajemen, Fakultas Ekonomi dan Bisnis (FEB), Universitas Putra Indonesia YPTK Padang, Sumatera Barat, Indonesia. Beliau merupakan seorang dosen di program studi Manajemen, Fakultas Ekonomi dan Bisnis (FEB), Universitas Putra Indonesia YPTK Padang. Terdapat beberapa Mata Kuliah yang diajarkan oleh beliau diantaranya yaitu Manajamen Pemasaran, Manajamen Keuangan, Manajemen Sumber Daya dll. Email beliau adalah robby_dharma@upiyptk.ac.id. Alamat kantor beliau adalah di Jalan Raya Lubuk Begalung, Kota Padang, Provinsi Sumatera Barat, Indonesia.

Dr. SYAFRI ARLIS, S.Kom, M.Kom. Lahir di Padang, Sumatera Barat, Indonesia pada tahun 1986. Meraih gelar sarjana dibidang ilmu komputer pada tahun 2009 dan gelar magister ilmu komputer konsentrasi teknologi informasi dari Universitas Putra Indonesia YPTK pada tahun 2011 serta gelar Doktor Teknologi Informasi dari Universitas Putra Indonesia YPTK pada tahun 2023. Saat ini aktif sebagai dosen pada fakultas ilmu komputer di Universitas Putra Indonesia YPTK. Karya penelitian berfokus secara khusus pada pengolahan citra khususnya pencitraan medis dan kecerdasan buatan (Artificial Intelligence). Email: syafri_arlis@upiyptk.ac.id.

Syafiadi Rizki Abdila, S.T., Ph.D. Beliau menyelesaikan pendidikan sarjana (S1) di Universitas Sultan Ageng Tirtayasa, Banten, Jawa Barat, Indonesia, dan menyelesaikan pendidikan Doktor (S3) di Universitas Malaysia Perlis. Beliau saat ini merupakan seorang dosen di program studi Teknik Sipil (TS), Fakultas Teknik (FT), Universitas Krisnadwipayana, DKI Jakarta, Indonesia. Terdapat beberapa Mata Kuliah yang diajarkan oleh beliau diantaranya yaitu Analisis Struktur II dan Menggambar Teknik dll. Email beliau adalah syafiadirizki@unkris.ac.id. Alamat kantor beliau adalah Kampus UNKRIS Jatiwaringin, Pondok Gede, DKI Jakarta, Indonesia.

Prof.Dr.Ir.Drs.Syafwandi., M.Sc adalah guru besar di kampus Universitas Putra Indonesia Padang. menempuh sarjana di 2 universitas, yakni Universitas Indonesia dan Universitas Muhammadiyah Jakarta, dengan Bidang Ilmu Sastra dan Teknik Sipil.. S2 di Institut Teknologi Bandung dengan bidang ilmu Teknik Planologi, dan menyelesaikan S3 di Universitas Satyagama Jakarta dengan bidang ilmu Manajemen Pemerintahan. Penulis saat ini mengampu matakuliah Fleksibilitas Manajemen, Filsafat Ilmu Pengetahuan dan Etika Bisnis.

Dr. Yulasmi, SE, MM. Berprofesi sebagai dosen Universitas Putra Indonesia "YPTK" sekaligus menjadi Dekan FEB Universitas Putra Indonesia "YPTK" Padang. Matakuliah yang diampu diantaranya, Manajemen Umum, Manajemen Pemasaran, dan Manajemen Strategik. Lulus S1 tahun 1998 di Sekolah Tinggi Ilmu Ekonomi YPTK Padang,S2 tahun 2022 di Universitas Putra Indonesia YPTK Padang, dan S3 lulus tahun 2018 di Universitas Persada Indonesia YAI Jakarta.

Dr. Robby Dharma, S.E., M.M. Beliau menyelesaikan pendidikan sarjana (S1) di program studi manajemen, Sekolah Tinggi Ilmu Ekonomi (STIE) Dharma Andalas Padang, Sumatera Barat, Indonesia, dan menyelesaikan pendidikan magister (S2) di program studi Magister Manajemen (M.M.), Fakultas Ekonomi (FE), Universitas Putra Indonesia YPTK Padang, Sumatera Barat, Indonesia. Beliau saat ini sedang melanjutkan pendidikan Doktor (S3) di Program Studi Doktor Manajemen, Fakultas Ekonomi dan Bisnis (FEB), Universitas Putra Indonesia YPTK Padang, Sumatera Barat, Indonesia.

Dr. Syafri Arlis, S.Kom, M.Kom. Lahir di Padang, Sumatera Barat, Indonesia pada tahun 1986. Meraih gelar sarjana dibidang ilmu komputer pada tahun 2009 dan gelar magister ilmu komputer konsentrasi teknologi informasi dari Universitas Putra Indonesia YPTK pada tahun 2011 serta gelar Doktor Teknologi Informasi dari Universitas Putra Indonesia YPTK pada tahun 2023. Saat ini aktif sebagai dosen pada fakultas ilmu komputer di Universitas Putra Indonesia YPTK. Karya penelitian berfokus secara khusus pada pengolahan citra khususnya pencitraan medis dan kecerdasan buatan (Artificial Intelligence). Email: syafri_arlis@upiyptk.ac.id.

Pustaka Galeri Mandiri

Perum Batu Kasek E11. Padang. SUMBAR

@pustakagaleri

Pustaka Galeri Mandiri

pustakagalerimandiri.co.id

Pustaka Galeri Mandiri

https://jurnal.pustakagalerimandiri.co.id/

ISBN 978-623-8164-31-8