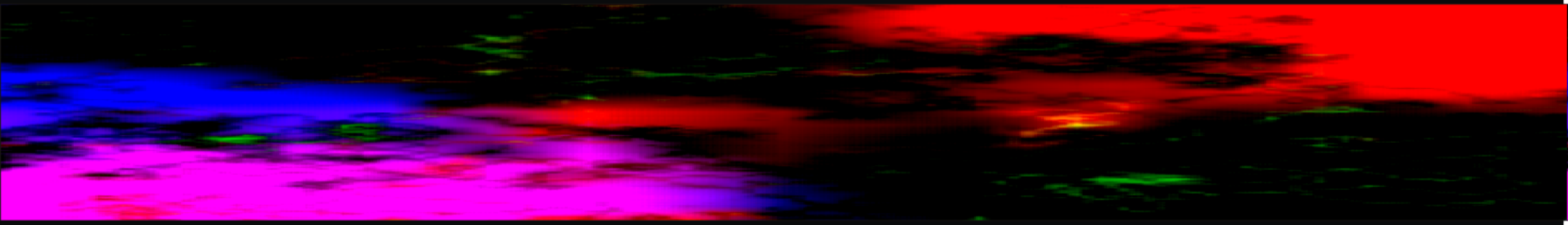
Modul 1

Statistika & Probabilitas

Modul ini adalah pedoman atau panduan dalam Proses Belajar dan Mengajar yang merupakan penunjang mata kuliah Probabilitas dan Statistika bagi mahasiswa Sekolah Tinggi Manajemen & Informatika (STMIK) Sumedang,baik jurusan Teknik Informatika, Manajemen Informatika maupun Sistem Informasi. Modul ini diharapkan dapat membantu mahasiswa dalam mempersiapkan dan mengikuti proses belajar dan mengajar dengan lebih baik, terarah dan terencana.

Statistika & Probabilitas



Reny Rian Marliana

Reny Rian Marliana

PROBABILITAS DAN STATISTIKA

Reny Rian Marliana, S.Si., M.Stat.

SEKOLAH TINGGI MANAJEMEN INFORMATIKA & KOMPUTER

SUMEDANG

2016

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kepada Allah SWT yang senantiasa memberi rahmat dan petunjuk-Nya dalam menyelesaikan modul mata kuliah Probabilitas dan Statistika untuk mahasiswa/mahasiswi Sekolah Tinggi Manajemen & Informatika (STMIK) Sumedang.

Adapun tujuan penulisan modul ini adalah sebagai pedoman atau panduan dalam Proses Belajar dan Mengajar yang merupakan penunjang mata kuliah Probabilitas dan Statistika bagi mahasiswa/mahasiswi Sekolah Tinggi Manajemen & Informatika (STMIK) Sumedang, baik jurusan Teknik Informatika, Manajemen Informatika maupun Sistem Informasi. Modul ini diharapakan dapat membantu mahasiswa/mahasiswi dalam mempersiapkan dan mengikuti proses belajar dan mengajar dengan lebih baik, terarah dan terencana.

Modul ini adalah hasil saduran dari beberapa penulis pada beberapa literatur. Penulis menyadari bahwa tanpa rahmat Allah SWT, serta bimbingan, bantuan, doa dan dorongan dari berbagai pihak, modul ini tidak mungkin dapat terselesaikan dengan baik pada waktunya. Untuk itu, penulis mengucapkan banyak terima kasih kepada semua pihak yang telah membantu baik secara langsung maupun tidak langsung, terutama kepada para penulis pada beberapa literatur yang menjadi saduran atau referensi dalam penulisan modul ini. Terima kasih telah memberikan kontribusi yang sangat besar kepada kemudahan penulisan modul ini.

Penulis menyadari bahwa dalam penulisan modul ini masih jauh dari kesempurnaan. Oleh karena itu, kritik dan saran yang sifatnya membangun sangat penulis harapkan. Kritik dan saran dapat disampaikan melalui email renyrian@stmik-sumedang.ac.id.

Akhir kata, semoga modul ini dapat bermanfaat bagi Proses Belajar Mengajar khususnya pada mata kuliah Probabilitas & Statistika.

Sumedang, Juni 2016

Penulis

DAFTAR ISI

Pengantar Dasar Statistika

1.1. Pengertian Statistika

Statistika berasal dari bahasa Italia, yaitu *statista* yang berarti Negarawan. Gottfried Achenwall (1719-1722) mengartikan Statistika sebagai keterangan-keterangan yang dibutuhkan oleh Negara. Sebagai contoh statistik penduduk, statistik kesehatan, statistik pembangunan. Sedangkan menurut Sudjana (2005) statistika adalah pengetahuan yang berhubungan dengan cara-cara pengumpulan data, pengolahan atau penganalisisannya dan penarikan kesimpulan berdasarkan kumpulan data dan penganalisisan yang dilakukan.

Menurut DR. Boediono (2014) statistika adalah pengetahuan yang berkaitan dengan metode, teknik atau cara untuk mengumpulkan data, mengolah data, menyajikan data, menganalisis data dan menarik kesimpulan atau menginterpretasikan data. Dengan demikian jelas bahwa statitika adalah sebuah ilmu pengetahuan mengenai pengumpulan, penyusunan, pengolahan dan analisis data yang diperlukan dalam menyusun sebuah keputusan.

Ruang lingkup statistika terbagi menjadi dua, yaitu statistika deskriptif dan statistika inferensial/induktif. Statistika yang berkenaan dengan metode atau cara mendeskripsikan, menggambarkan atau menguraikan data atau sifatnya eksploratif disebut sebagai statistika deskriptif. Dengan kata lain, statistika deskriptif lebih ditekankan pada penyajian data agar informasi yang terkandung dalam data mudah difahami oleh pengguna data. Data dapat disajikan secara *numerik* atau secara grafik. Penyajian data secara numerik akan melibatkan perhitungan ukuran-ukuran deskriptif seperti rata-rata, simpangan baku, varians, kuartil, persentil dan sebagainya. Sedangkan penyajian data secara grafik bisa melalui *bar chart, pie chart, box plot, scatter plot* atau *steam & leaf diagram*.

Statistika yang berkenaan dengan cara penarikan kesimpulan berdasarkan data yang diperoleh dari sampel untuk menggambarkan karakteristik atau ciri-ciri dari suatu populasi atau sifatnya adalah konfirmatif disebut sebagai statistika inferensial/induktif. Populasi adalah keseluruhan objek, baik itu hasil menghitung maupun mengukur yang dibatasi oleh kriteria tertentu. Besaran yang menjadi ciri populasi dinamakan sebagai parameter, misalnya rata-rata (μ) atau simpangan baku (σ). Sedangkan sampel adalah bagian dari populasi yang menjadi perhatian kita (DR. Boediono, 2014). Besaran yang menjadi ciri sampel dinamakan sebagai statistik, misalnya rata-rata ($\bar{x}$) atau simpangan baku (s).

Ruang lingkup statistika inferensial/induktif terbagi menjadi dua yaitu penaksiran parameter dan pengujian hipotesis. Penaksiran parameter adalah perhitungan nilai-nilai besaran yang menjadi ciri sebuah populasi yang diperoleh dari data sampel. Dengan kata lain, dalam penaksiran parameter, kita menghitung statistik sampel yang akan digunakan untuk menduga nilai dari parameter populasi. Sebagai contoh nilai $\bar{x}$ digunakan untuk menduga nilai μ, atau nilai s digunakan untuk menduga nilai σ. Penaksiran parameter ini dapat berupa satu nilai (titik) ataupun dalam bentuk interval.

Pengujian hipotesis adalah langkah-langkah yang dilakukan dengan tujuan untuk menguji kebenaran suatu hipotesis yang merupakan rumusan pernyataan ilmiah sebagai jawaban terhadap masalah serta masih memerlukan pengujian empiris. Misalkan terdapat pernyataan bahwa rata-rata masa pakai baterei laptop ASUS adalah 2 tahun. Untuk menguji kebenaran pernyataan tersebut diperlukan sebuah langkah empiris yang disebut sebagai pengujian hipotesis. Terdapat dua kemungkinan hasil pengujian hipotesis yaitu menolak atau menerima hipotesis. Menolak hipotesis artinya bahwa hipotesis tidak benar sedangkan menerima hipotesis artinya tidak cukup bukti untuk menolak hipotesis.

1.2. Sejarah Statistika

Perkembangan sejarah penggunaan statistika dapat dilihat pada Tabel berikut ini :

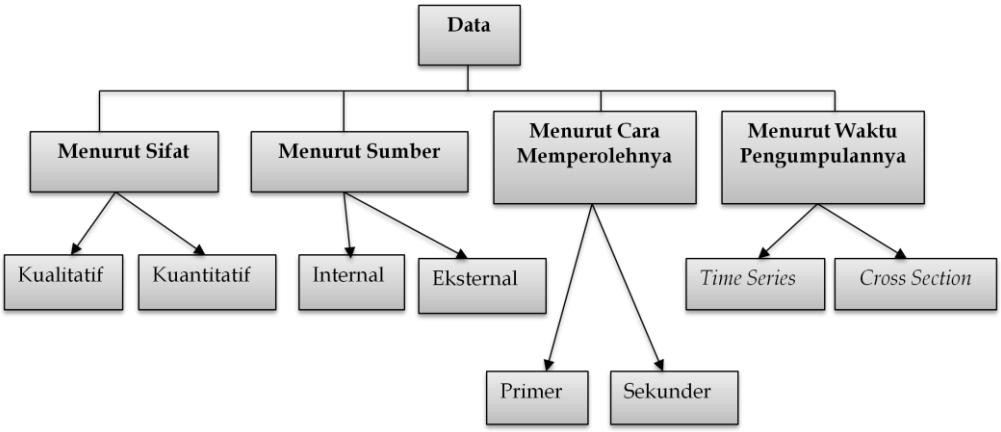
Tahun	Penggunaan Statistika
Abad 18	Negara-negara Babilon, Mesir, dan Roma mengeluarkan catatan tentang nama, usia, jenis kelamin, pekerjaan dan jumlah anggota
1632	Penyusunan buku Bill of Mortality pada tahun 1632 di Inggris yang berisi tentang catatan kelahiran dan kematian di Inggeris menurut jenis kelamin
1662	Pemerintahan Inggris mengembangkan catatan Mingguan tentang kematian menjadi catatan kelahiran dan kematian
1772-1791	G.Achenwall menggunakan istilah Statistika sebagai kumpulan data tentang negara
1791-1799	Dr.E.A.W. Zimmesman mengenalkan kata Statistika dalam bukunya Statistical Account of Scotland
1981-1935	R. Fisher mengenalkan analisa varians dalam literatur statistikanya
Abad 19-20	Teori statistika disumbangkan oleh Bayes, Ronald Fisher, W.S Gosset, Galton, dan Karl Pearson

1.3.Data Statistika

Seperti yang sudah diuraikan pada sub bab sebelumnya, statistika adalah ilmu yang berhubungan dengan bagaimana cara mengumpulkan data, menyusun, mengolah dan membuat kesimpulan dari data yang diperoleh. Data berarti sesuatu yang diketahui atau dianggap, artinya, data dapat memberikan gambaran tentang suatu keadaan (*Webster's New World Dictionary*, dalam Supranto, 2008). Sementara menurut Sudjana (2005) data statistik adalah semua keterangan atau ilustari mengenai suatu hal bisa berbentuk kategori, misalnya rusak, baik, senang, puas, berhasil, gagal dan sebagainya, juga bisa berbentuk bilangan. Rudolf J, Freud (2003) menyebutkan bahwa *A set of data is a collection of observed values representing one or more characteristics of some objects or units.* Dengan demikian, jelas bahwa data adalah kumpulan nilai hasil observasi bisa berupa kategori maupun bilangan.

Dasar penggunaan data adalah untuk membuat keputusan oleh para *decision maker* yang dapat digunakan sebagai dasar suatu perencaan, alat pengendalian dan dasar evaluasi. Sebagai contoh, Pemerintah, untuk dapat memutuskan tentang kestabilan keadaan sosial dan ekonomi masyarakat, maka diperlukan gambaran tentang keadaan sosial dan ekonomi negara yang dapat dilihat dari data mengenai kegiatan ekonomi seperti produksi, perdagangan, konsumsi, pendapatan, harga dan lain-lain. Contoh lainnya di sebuah perusahaan, untuk dapat mengetahui perkembangan usahanya suatu perusahaan harus memiliki data mengenai hasil produksi, data hasil penjualan, data personalia, data keuangan, data peralatan, data mengenai persentase pelanggan yang tidak puas dan lain-lain.

Data dikelompokkan menjadi beberapa jenis menurut sifatnya, menurut sumber, menurut cara memperolehnya dan menurut waktu pengumpulannya.



Data menurut sifatnya dikelompokkan menjadi dua yaitu data kualitatif dan data kuantitatif. Data kualitatif adalah data yang dinyatakan dalam bentuk kategori, misalnya sukses-gagal, setuju-tidak setuju, laki-laki-perempuan, dan sebagainya. Sedangkan data

kuantitatif adalah data yang dinyatakan dalam bentuk angka bisa berupa hasil perhitungan atau hasil pengukuran, misalnya berat badan, tinggi badan, temperatur udara, banyaknya produk yang terjual.

Data menurut sumbernya dibedakan menjadi dua yaitu data intern dan data eksternal. Data intern adalah data yang diperoleh atau bersumber dari dalam suatu instansi (lembaga, organisasi), sedangkan data ekstern adalah data yang diperoleh atau bersumber dari luar instansi (DR. Boediono, 2014). Pengusaha mencatat segala aktivitas perusahaannya sendiri, misalnya keadaan pegawai, pengeluaran, keadaan barang di gudang, hasil penjualan, keadaan produksi pabriknya dan lain-lain aktivitas yang terjadi di dalam perusahaan itu. Data yang diperoleh demikian merupakan data intern (Sudjana, 2005). Dalam berbagai situasi, untuk perbandingan, misalnya diperlukan data dari sumber lain di luar perusahaan tadi. Data demikian merupakan data ekstern (Sudjana, 2005).

Data menurut cara memperolehnya terbagi menjadi dua, yaitu data primer dan data sekunder. Pada dasarnya data primer dan data sekunder termasuk ke dalam data ekstern (DR. Boediono, 2014). Data primer adalah data yang langsung dikumpulkan oleh orang yang berkepentingan atau yang memakai data tersebut, data yang diperoleh melalui wawancara atau kuesioner merupakan contoh data primer. Sedangkan data sekunder adalah data yang tidak secara langsung dikumpulkan oleh orang yang berkepentingan dengan data tersebut. Data yang diperoleh dari laporan tahunan perusahaan untuk keperluan menulis skripsi merupakan contoh data sekunder.

Data menurut waktu pengumpulannya terbagi menjadi dua yaitu data *times series* dan data *cross section*. Data yang dikumpulkan dari waktu ke waktu untuk menggambarkan suatu perkembangan atau kecenderungan keadaan atau peristiwa atau kegiatan, dimana jarak atau interval dari waktu ke waktu sama disebut sebagai data *times series* atau data berkala (DR. Boediono, 2014). Data *time series* sering juga disebut sebagai data deret waktu yang merupakan sekumpulan hasil observasi yang diatur dan didapat menurut urutan kronologis, biasanya dalam interval waktu yang sama (Sudjana, 2005). Pengertian lain diungkapkan oleh J.Supranto (2008), data berkala adalah data yang dikumpulkan dari waktu ke waktu untuk menggambarkan perkembangan suatu kegiatan misalnya perkembangan produksi, harga, hasil penjualan, jumlah personil, penduduk, jumlah kecelakaan, jumlah kejahatan, jumlah peserta KB dan lain sebagainya. Data berkala juga merupakan suatu rangkaian atau seri dari nilai-nilai suatu variabel yang dicatat dalam jangka waktu yang berurutan (Atmaja, 2009). Contoh dari data *time series* adalah penjualan mingguan sebuah produk di toko, produksi

bulanan di sebuah perusahaan industri atau produksi tahunan bijih besi di Indonesia. Sementara, data *cross section* adalah data yang dikumpulkan dalam satu waktu atau dalam titik waktu tertentu. Contohnya adalah laporan keuangan per 31 Desember 2006 di sebuah perusahaan tertentu, data pelanggan Telkom pada bulan Mei 2015.

Agar hasil sebuah penelitian yang dilakukan dapat dipercaya maka data yang digunakan haruslah data yang baik. Sebuah data dikatakan data yang baik jika memenuhi kriteria sebagai berikut :

1. Objektif, artinya data yang digunakan harus sesuai dengan kondisi sebenarnya bukan merupakan hasil penilaian subjektif dari penelitinya.
2. Refresentatif, artinya data yang digunakan harus mewakili semua karakteristik dari elemen-elemen dalam populasi.
3. Kesalahan sampling kecil artinya tingkat ketelitian lebih besar.
4. Tepat waktu
5. Relevan, artinya data yang digunakan harus sesuai dengan tujuan dari penelitian yang dilakukan.

1.4. Metode Pengumpulan Data

Data dapat dikumpulkan dengan dua cara yaitu melalui sensus dan sampling. Apabila seluruh elemen populasi diselidiki satu-persatu, cara pengumpulan datanya disebut sebagai sensus. Data yang diperoleh disebut dengan data sebenarnya (*true value*). Ukuran-ukuran statistika (rata-rata, simpangan baku, dll) yang diperoleh disebut sebagai parameter.

Apabila yang diselidiki adalah elemen sampel atau hanya sebagian elemen dari populasi, cara pengumpulannya disebut sebagai sampling. Data yang diperoleh disebut data perkiraan (*estimated value*). Ukuran-ukuran statistika (rata-rata, simpangan baku, dll) yang diperoleh disebut sebagai statistik.

Terdapat beberapa alasan sehingga peneliti cenderung lebih memilih proses sampling daripada sensus, yaitu :

a) Mengurangi biaya, apabila kita melakukan penelitian terhadap sebagian dari anggota populasi maka berakibat pada penghematan biaya.
b) Masalah tenaga, jelas bahwa semakin banyak objek yang kita teliti, maka akan berakibat pada semakin banyaknya tenaga yang kita butuhkan baik itu tenaga pengumpul data, pencatat atau entri data maupun pengolah data. Apabila ada

keterbatasan untuk ketiga hal tersebut, maka sampling merupakan alternatif terbaik untuk dilakukan.

c) Efisiensi waktu, apabila diinginkan kesimpulan yang segera, maka sampling akan lebih tepat untuk digunakan. Hal ini dikarenakan dengan memperkecil banyaknya objek yang akan diteliti maka data akan lebih cepat diperoleh dan dianalisis.
d) Tingkat ketelitian lebih besar, dalam suatu proses penelitian dari mulai pengumpulan data, pencatatan, dan penganalisisan data harus dilakukan dengan benar dan tepat. Apabila kita telah memakai tenaga-tenaga yang berkualitas baik dan diberi latihan intensif serta pengawasan terhadap pekerjaan lapangan diperketat tetapi memberikan volume pekerjaan yang besar dan cenderung monoton, maka akan menimbulkan kebosanan baik itu dari pencacah maupun peneliti. Oleh karena itu, akan diperoleh data yang kurang dapat dipercaya kebenarannya.
e) Penelitian bersifat destruktif atau penelitian yang sifatnya merusak, sensus tidak mungkin dilakukan untuk objek yang sifatnya merusak. Misalnya dalam menguji golongan darah seseorang, maka tidak mungkin semua darah dikeluarkan untuk diperiksa. Jadi dalam hal ini sensus, tidak mungkin lagi untuk dilakukan.
f) Faktor ekonomis, yang dimaksud dengan faktor ekonomis adalah kesepadanan antara biaya, tenaga dan waktu yang dikeluarkan dengan informasi yang akan diperoleh. Apabila nilai dari informasi tersebut tidak sepadan dengan biaya, tenaga dan waktu, maka sensus menjadi tidak baik lagi untuk dilakukan.

Dari kedua cara pengumpulan data tersebut, jelas bahwa kita akan selalu bertemu dengan elemen populasi atau sampel. Rudolf J. Freud (2003) menyebutkan bahwa *a population is a data set representing the entire entity of interest.* Sama halnya dengan yang diungkapkan oleh Sudjana (2005) populasi adalah totalitas semua nilai yang mungkin, baik hasil menghitung maupun pengukuran, kuantitatif ataupun kualitatif, daripada karakteristik tertentu mengenai sekumpulan objek lengkap dan jelas. Sementara menurut DR.Boediono (2014) populasi adalah suatu keseluruhan pengamatan atau obyek yang menjadi perhatian kita. Berdasarkan ketiga definisi tersebut, dapat disimpulkan bahwa secara umum populasi merupakan keseluruhan objek, baik itu hasil menghitung maupun mengukur yang dibatasi kriteria tertentu.

Sudah disebutkan sebelumnya bahwa dalam pengumpulan data melalui sampling objek yang diselidiki (diobservsi atau diteliti) merupakan elemen sampel. Menurut Rudolf J. Freud

(2003) *a sampel is a data set consisting of a portion of a population. Normally a sampel is obtained in such a way as to be representative of the population.* Definisi lain disebutkan oleh Sudjana (2005), sampel adalah sebagian yang diambil dari populasi dengan menggunakan cara-cara tertentu. Sama halnya dengan yang diungkapkan oleh DR. Boediono (2014) sampel adalah bagian dari populasi yang menjadi perhatian kita.

Dalam pengumpulan data melalui sampling, yang menjadi persoalan adalah bagaimana cara kita memilih atau menentukan elemen populasi sebagai elemen sampel. Sebagai contoh, sebuah penelitian dilakukan untuk mengetahui rata-rata konsumsi BBM/minggu dari kendaraan bermotor roda empat di Kabupaten Sumedang. Kita akan mengalami kesulitan dalam pengumpulan data jika dilakukan secara sensus karena membutuhkan waktu, biaya dan tenaga yang cukup besar. Untuk mengurangi resiko tersebut maka sampling akan lebih baik digunakan untuk penelitian tersebut. Misalnya berdasarkan informasi dari Dinas Perhubungan Kabupaten Sumedang, diketahui bahwa banyaknya kendaraan bermotor roda 4 di Kabupaten Sumedang adalah 1500 kendaraan, kemudian berdasarkan perhitungan penentuan ukuran sampel, banyaknya kendaraan bermotor yang harus diobservasi menjadi elemen sampel adalah 500 kendaraan. Untuk dapat memilih 500 kendaraan roda empat dari 1500 kendaraan roda empat yang ada diperlukan sebuah teknik sampel. Proses pengambilan sebagian anggota populasi untuk mendapatkan sampel disebut sebagai *sampling*.

Sampling terbagi menjadi dua yaitu *non-probability sampling* dan *probability sampling*. Sebuah proses pengambilan sebagian anggota populasi dimana masing-masing elemen populasi tidak memiliki kesempatan/peluang sama untuk terpilih sebagai elemen sampel disebut sebagai *non-probability sampling*. Sebaliknya sebuah proses pengambilan sebagian anggota populasi dimana masing-masing elemen populasi memiliki kesempatan/peluang yang sama untuk dijadikan sebagai elemen sampel disebut sebagai *probability sampling*.

Perbedaan antara *non-probability sampling* dan *probability sampling* adalah *non-probability sampling* tidak melibatkan pemilihan secara acak sedangkan *probabiliy sampling* melibatkan pemilihan secara acak. Berbeda dengan *probability sampling* yang dapat diketahui peluang atau kemungkinan bahwa sampel telah mewakili populasi dengan baik sehingga dapat diperkirakan interval konfidensi untuk statistik, *non-probability sampling* justru tidak dapat diketahui berapa besarnya peluang untuk kemungkinannya untuk mewakili populasi dengan baik, sehingga akan sulit untuk mengetahui seberapa baik penelitian yang kita lakukan. *Non-probability sampling* tidak dapat digunakan untuk mengambil kesimpulan dari sampel untuk keseluruhan populasi. Setiap generalisasi yang diperoleh dari *non-*

probability sampling harus disaring melalui satu pengetahuan tentang topik kajian dalam penelitian yang dilakukan.

Secara umum, peneliti cenderung lebih menyukai *probability sampling* yang melibatkan pemilihan secara acak dibandingkan dengan *non-probability sampling*, dengan pertimbangan bahwa *probability sampling* lebih akurat dan robust. Namun, dalam penelitian sosial mungkin ada situasi yang tidak layak, praktis atau secara teoritis tidak masuk akal untuk melakukan sampel acak, maka dipertimbangkan berbagai alternatif *non-probability sampling*.

Terdapat beberapa teknik sampling dalam *non-probability sampling* diantaranya :

1. *Accidental sampling*

   Accidental sampling memiliki nama lain yaitu *haphazard sampling*, sampling seadanya atau sampling kenyamanan. Sampling ini adalah sampling dimana satuan sampling diperoleh secara sembarangan atau seketemunya sehingga tidak dapat dibuktikkan bahwa sampel yang telah diambil adalah wakil dari populasi. Sebagai contoh, wawancara yang sering dilakukan oleh program televisi untuk mendapatkan berita yang cepat walaupun tidak representatif, membaca opini publik, penggunaan mahasiswa dalam banyak penelitian psikologis adalah terutama untuk kenyamanan (psikolog tidak yakin bahwa mahasiswa dapat mewakili penduduk yang besar), penelitian pada bidang Arkeolog dan Sejarah.

2. *Purposive sampling*

   Purposive sampling disebut juga sebagai sampling sengaja atau sampling tetap. Sampling ini adalah sampling dimana pemilihan sampel berdasarkan sampel yang sesuai dengan kajian yang akan diteliti. Hal ini digunakan terutama bila ada beberapa orang yang memiliki keahlian di daerah penelitian. Dengan *purposive sampling*, ada kemungkinan untuk mendapatkan pendapat dari target populasi tetapi cenderung juga pada s*ubgoups* secara berlebihan dalam populasi yang lebih mudah diakses.

Sedangkan dalam *probability sampling* terdapat 4 teknik sampling yaitu *simple random sampling, systematic sampling, stratified random sampling* dan *cluster sampling*.

1) *Simple random sampling*

Simple random sampling atau sering juga disebut sebagai sampling acak sederhana merupakan bentuk yang paling dasar dari *probability sampling*. *Simple random sampling* merupakan suatu proses memilih satuan sampling dari populasi sedemikian rupa sehingga

setiap satuan sampling dalam populasi mempunyai peluang yang sama besar untuk terpilih ke dalam sampel dan peluang itu diketahui sebelum pemilihan dilakukan.

Teknik pengambilan sampling acak sederhana (*simple random sampling*) adalah pengambilan sampel sebanyak *n* sedemikian rupa sehingga setiap unit dalam populasi mempunyai kesempatan yang sama untuk terambil dan setiap ukuran sampel *n* juga mempunyai kesempatan yang sama untuk terambil (DR. Boediono, 2014).

Salah satu cara sederhana untuk mendapatkan sampel acak adalah sebagai berikut, misalkan sebuah populasi berisikan 100 anggota misalnya pegawai, rumah, keluarga, toko, industri, kotakan sawah, mobil, radio, makanan dalam kaleng dan sebagainya. Dari populasi ini akan diambil sebuah sampel acak terdiri atas 30 anggota. Tiap anggota dalam populasi kita beri nomo 1, 2, ..., 100 untuk anggota terakhir. Tuliskan nomor-nomor ini masing-masing pada secarik kertas yang berukuran dan beridentitas, lalu gulung. Setelah diaduk dengan baik, seseorang yang matanya tertutup mengambil satu-satu sampai 30 kali. Nomor-nomor yang tertulis pada kertas yang terambil akan memberikan sampel acak terdiri atas 30 anggota dari populasi yang diberikan (Sudjana, 2005).

Cara yang lebih bersifat ilmiah untuk mendapatkan sebuah sampel acak yaitu menggunakan daftar bilangan acak atau menggunakan kalkulator. Proses sampling acak sederhana digunakan apabila memenuhi beberapa kondisi sebagai berikut :

- Variabel yang akan diteliti keadaannya relatif homogen dan tersebar merata di seluruh populasi
- Apabila disusun secara lengkap kerangka sampling (daftar tiap anggota dalam populasi) yang menyangkut setiap satuan pengamatan yang ada dalam populasi.

Langkah-langkah pemilihan sampling acak sederhana adalah sebagai berikut :

- Tentukan secara tegas populasi sasaran, misalnya masyarakat di Daerah A.
- Buat kerangka sampling

No	Nama	Alamat
1	Awal	Jl. Merkuri Raya 23
2	Arya	Jl. Jakarta 24
⋮	⋮	⋮
100	Ending	Jl. Cikaso 23

- Tentukan ukuran sampel *n*

Penentuan ukuran sampel bergantung pada tujuan penelitiannya, berikut adalah penentuan ukuran sampel dengan tujuan penelitian menduga nilai rata-rata atau proporsi populasi sebagai berikut :

Tujuan Penelitian	Ukuran Sampel
Menaksir atau menduga nilai rata-rata populasi (μ)	$n_0 = \left(\frac{z_{\alpha/2} S}{\delta}\right)^2$ $\quad n = \frac{n_0}{1+\frac{n_0}{N}}$ S : Simpangan baku untuk variabel yang diteliti dalam populasi δ : *Bound of error* yang dikehendaki $z_{\alpha/2}$: konstanta yang diperoleh dari tabel normal standar N : Ukuran Populasi
Menaksir atau menduga nilai proporsi populasi (π)	$n_0 = \left(\frac{z_{\alpha/2}}{2\delta}\right)^2$ $\quad n = \frac{n_0}{1+\frac{n_0 - 1}{N}}$

- Lakukan proses pengambilan sampel

  Apabila target populasi telah ditentukan secara tegas dan dari populasi ini akan disusun sebuah sampel acak sederhana, maka selanjutnya harus dilakukan proses pemilihan dari anggota sampelnya melaui tabel acak atau dengan angka acak dalam kalkulator.

2) *Systematic sampling*

Sebuah sampel yang diperoleh dari penyeleksian satu unsur secara acak dari *k* unsur yang pertama dalam sebuah kerangka sampling dan setiap unsur ke-*k* kemudian disebut satu dalam *k* sampel sistematik. Jadi, suatu proses memilih dikatakan sampling sistematik apabila dalam pemilihan itu dilakukan pemilihan sistematik setelah terpilih bilangan acak dengan syarat bahwa peluang terpilihnya 1/*N*.

Sampling sistematik digunakan apabila bisa disusun kerangka sampling lengkap dan keadaan variabel yang sedang diteliti relatif homogen dan tersebar merata di seluruh populasi. Sampling sistematik memberikan sebuah alternatif yang berguna dari sampling acak sederhana untuk alasan sebagai berikut :

- Sampling sistematik lebih mudah untuk dilakuan dan oleh sebab itu lebih sedikit subjek yang melakukan kesalahan wawancara daripada sampling acak sederhana.
- Sampling sistematik sering memberikan informasi yang lebih banyak mengenai biaya per unit/satuan daripada yang diberikan sampling acak sederhana.

Pada umunya sampling sistematik merupakan penyeleksian secara acak pada suatu unsur dari *k* unsur yang pertama dan kemudian penyeleksian pada setiap unsur *k* sesudahnya. Prosedur ini lebih mudah dibentuk dan biasanya akan meminimalisisr kesalahan yang mungkin dilakukan oleh pewawancara daripada proses sampling acak sederhana. Sebagai contoh, akan menjadi lebih sulit apabila menggunakan sampling acak sederhana untuk menyeleksi *n*=50 pembeli pada sebuah *Mall*. Pewawancara tidak menentukan pembeli-pemblei mana yang termasuk dalam sampelnya, karena ia tidaak memiliki kerangka sampling serta tidak mengetahui ukuran populasi *N*. Sebagai solusinya, ia dapat mengambil sampel secara sistematik, katakanlah 1 dari 20 pembeli, hingga persyaratan sampelnya bisa didapatkan. Ini akan menjadi sebuah prosedur yang mudah bahkan untuk pewawancara yang tidak berpengalaman sekalipun dapat melakukannya.

3) *Stratified random sampling*

Misalkan di suatu daerah, pendapatan masyarakat bersifat heterogen yaitu tergolong "tinggi", "menengah" atau "rendah" dan melalui sampling acak sederhana akan diambil sampel dalam usaha menaksir rata-rata pendapatan masyarakat tersebut. Maka, ada kemungkinan yang terambil ke dalam sampel walaupun dilakukan secara acak, kebanyakan atau hanyalah mereka yang tergolong berpengahasilan rendah saja. Bila rata-rata pendapatan dihitung dari sampel ini, maka rata-rata tadi akan merupakan taksiran/dugaan yang rendah (*under estimate*). Dengan demikian sampling acak sederhana akan memberikan presisi atau kekonsistenan (keseragaman) dari nilai penaksir yang rendah. Oleh karena itu diperlukan sebuah metode sampling lain yang dapat menghasilkan presisi yang lebih tinggi.

Agar diperoleh presisi yang tinggi, sampel yang terambil haruslah sampel yang di dalamnya berisi masyarakat dari semua golongan pendapatan (tinggi, menengah dan rendah). Sampel seperti ini diperoleh melalui *stratified random sampling* atau sampling acak stratifikasi.

Dalam sampling acak stratifikasi, populasi *N* dibagi ke dalam beberapa kelompok sedemikian sehingga setiap kelompok mempunyai karakteristik yang homogen. Kelompok-

kelompok semacam ini disebut sebagai strata (tunggalnya stratum) dan dalam masing-masing stratum sampel diambli secara acak, yaitu melalui sampling acak sederhana.

Dalam contoh, populasi dibagi dalam tiga strata, stratum pertama adalah masyarakat yang tergolong berpenghasilan tinggi, stratum kedua yang berpenghasilan menengah dan stratum ketiga masyarakat yang berpenghasilan rendah.

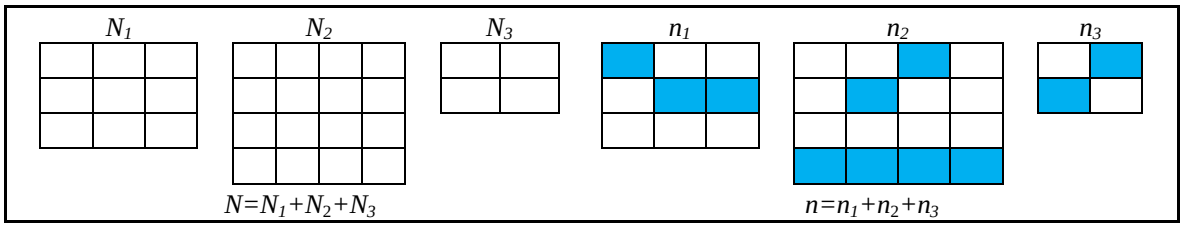
Langkah-langkah pemilihan sampel dalam *stratified random sampling* adalah sebagai berikut :

- Tentukan populasi sasaran dan tentukan populasi keseluruhan (N)
- Berdasarkan kriteria tertentu, populasi dibagi ke dalam L buah strata
- Untuk setiap strata lakukan pendaftaran satuan sampling sehingga untuk setiap strata diperoleh kerangka sampling masing-masing dengan ukuran strata masing-masing.
- Dari populasi tersebut kemudian ditentukan ukuran sampel N yang disebut *overall sample size*. Menentukan ukuran sampel n tentu saja harus berdasarkan kriteria tertentu.
- Ukuran-ukuran sampel sebesar n selanjutnya dialokasikan (disebarkan) ke seluruh strata yang kemudian disebut alokasi sampel (*sample allocation*).

$$\left.\begin{array}{l}\text{Stratum I} : n_1\\ \text{Stratum II} : n_2\\ \text{Stratum III} : n_3\\ \vdots\\ \text{Stratum L} : n_L\end{array}\right\} \text{sedemikian rupa sehingga } n=\sum_{i=1}^{L} n_i$$

- Dari setiap stratum kemudian dipilih satuan sampling melalui teknik sampel acak sederhana.

Sebagai combaran pembagntoh, di bawah ini diberikan gambaran populasi menjadi tiga biah stratum yang kemudian dilakukan proses *stratified random sampling*.



4) *Cluster sampling*

Andaikan seorang peneliti ingin mengetahui rata-rata pendapatan kepala kelurga di sebuah kota besar. Apabila sampling acak sederhana atau sampling acak stratifikasi digunakan, maka peneliti harus mempunyai kerangka sampling yang berisikan daftar keluarga di kota tersebut. daftar keseluruhan nama kelurga di kota yang besar pasti akan sulit diperoleh, kalaupun ada dan sampling acak sederhana dilakukan maka sampel masyarakat yang terambil bisa tersebar ke semua penjuru kota, dan ini akan melibatkan biaya pengambilan sampel yang tinggi. Daftar yang mungkin bisa diperoleh adalah daftar nama-nama kelurahan di kota tersebut. Kelurahan adalah kumpulan kepala keluarga. Oleh karena itu kelurahan dipandang sebagai klaster.

Proses pengambilan sampling klaster dilakukan dengan memperhatikan kerangka sampling yang berisikan dafatar klaster. Dalam contoh di atas daftar nama kelurahan. Pengambilan sampel kemudian dilakukan dengan mengambil secara acak klaster-klaster. Unit sampling yang berisikan klaster-klaster dinamakan unit sampling utama (USU). Apabila semua unit observasi dalam USU menjadi anggota sampel maka proses pengambilan sampel dilakukan dengan sampling klaster satu tahap. Namun apabila USU dibagi lagi ke dalam unit yang lebih kecil, misalnya kelurahan dibagi lagi ke dalam rukun-rukun warga (RW) maka rukun warga disebut unit sampling ke dua (USD). Apabila semua unit observasi dalam USD menjadi anggota sampel, maka proses pengambilan sampel dilakukan dengan sampling klaster dua tahap.

Langkah-langkah pemilihan sampel dalam sampling klaster adalah sebagai berikut :

- Populasi dibagi-bagi ke dalam *N* buah klaster atau unit sampling utama. Keadaan variabel yang diteliti dalam setiap klaster diusahakan heterogen artinya tidak seragam.
- Gunakan sampling acak sederhana untuk memilih *n* buah klaster
- Pemilihan hanya dilakukan sekali yaitu memilih klaster, oleh karena itu semua unit sampling kedua yang ada dalam klaster yang terpilih diperiksa.

Hal lain yang perlu dilakukan dalam sebuah pengumpulan data adalah menentukan alat pengumpulan datanya. Terdapat beberapa macam alat pengumpulan data, sebagai berikut “

1. Kuesioner
2. Wawancara
3. Observasi langsung atau pengamatan langsung
4. Melalui pos, telepon atau alat komunikasi lainnya

5. Alat ukur seperti meterean, timbangan, termometer, dll.

Setelah data terkumpul, maka selanjutnya hal yang perlu dilakukan dalam sebuah penelitian adalah mengolah data tersebut. Tujuan dari pengolahan data adalah untuk mendapatkan data statistik misalnya total, rata-rata, persentase, angka indeks, simpangan baku, koefisien korelasi, koefisien regresi, dll., yang dapat digunakan untuk melihat atau menjawab persoalan secara kelompok, bukan satu per satu secara individu. Metode pengolahan data bisa dilakukan secara manual atau menggunakan elektronik (kalkulator, komputer)

1.5. Skala Pengukuran Variabel

Variabel adalah karakteristik yang dimiliki satuan pengamatan yang berubah-ubah keadaanya. Variabel berdasarkan skala pengukurannya terbagi menjadi empat yaitu skala nominal, skala ordinal, skala interval dan skala rasio.

1.5.1. Skala Nominal

The nominal scale identifies observed values by name or classification (Rudolf, J.F., 203) atau dengan kata lain sebuah variabel dikatakan berskala nominal jika memiliki ciri-ciri sebagai berikut :

- Keadaan variabel dinyatakan dalam klasifikasi atau kategori
- Apabila untuk setiap klasifikasi yang berbeda, dicantumkan bilangan yang fungsinya hanya sebagai lambang pembeda.
- Pada variabel berskala nominal tidak berlaku hukum aritmatika, karena angka yang dicantumkan hanya sebagai lambang pembeda.

Sebagai contoh variabel jenis kelamin dapat dikategorikan sebagai berikut :

Angka (1) untuk kategori laki-laki

Angka (2) untuk kategori perempuan

Contoh lain, variabel jawaban responden pada sebuah item pertanyaan dapat dikategorikan sebagai berikut :

Angka (1) untuk jawaban “ya”

Angka (2) untuk jawaban “tidak”

1.5.2. Skala Ordinal

The ordinal scale distinguishes between measurements on the basis of the relative ampunts of some characteristic the possess. Usually the ordinal scale refers to measurements that make only "greater", "less", or "equal" comparisons between consecutive measurements (Rudolf, J.F.,2003). Dengan kata lain, sebuah variabel dikatakan berskala ordinal jika memiliki ciri-ciri sebagai berikut :

- Keadaan variabel dinyatakan dalam klasifikasi atau kategori
- Setiap klasifikasi diberi lambang bilangan
- Fungsi lambang bilangan adalah sebagai lambang pembeda dan memperlihatkan urutan peringkat
- Tidak berlaku hukum artimatika

Sebagai contoh, variabel jawaban responden pada sebuah item pertanyaan dapat dikategorikan sebagai berikut :

Angka (1) untuk jawaban "Sangat Tidak Setuju"

Angka (2) untuk jawaban "Tidak Setuju"

Angka (3) untuk jawaban "Netral"

Angka (4) untuk jawaban "Setuju"

Angka (5) untuk jawaban "Sangat Setuju"

1.5.3. Skala Interval

The interval scale of measurement also uses the concept of distance or measurement and requires a "zero" point, but the definition of zero may be arbitrary (Rudolf,J.F.,2003). Sebuah variabel dikatakan berskala interval jika memiliki ciri-ciri sebagai berikut :

- Keadaan variabel dinyatakan dalam bilangan atau angka
- Fungsi bilangan adalah sebagai lambang pembeda dan menunjukkan peringkat dengan catatan makin besar bilangan, makin tinggi peringkatnya (tidak bisa dibalik)
- Memperlihatkan jarak atau interval
- Titik nol bukan merupakan titik absolut tetapi titik yang ditentukan berdasarkan perjanjian
- Hukum aritmatika berlaku

Sebagai contoh, variabel yang menunjukkan temperatur Celcius 100,...,0 dimana $0^{\circ}C \neq 0^{\circ}F$

1.5.4. Skala Rasio

The ratio scale of measurement uses the concept of a unit of distance or measurement and rquires a uniqe definition of zero value (Rudolf,J.f.,2003). Sebuah variabel dikatakan berskala rasio jika memiliki ciri-ciri sebagai berikut :

- Keadaan variabel dinyatakan dalam bilangan atau angka
- Fungsi bilangan adalah sebagai lambang pembeda dan menunjukkan peringkat dengan catatan makin besar bilangan makin tinggi peringkatnya (tidak bisa dibalik)
- Memperlihatkan jarak/interval
- Titik nol nya merupakan titik absolut, artinya 0=kosong
- Hukum aritmatika berlaku

Sebagai contoh variabel yang menunjukkan usia, berat badan, tinggi badan, jumlah produk yang terjual, dkk.

LATIHAN

1. Jelaskan pengertian dan keguanaan dari Statistika?
2. Apakah yang dimaksud dengan data? Hal apakah yang harus diperhatikan tentang data ini?
3. Apa yang dimaksud dengan populasi? Berikan contohnya dengan batas-batas yang jelas?
4. Andaikan seorang produsen ingin mengetahui bagaimana kesenangan masyarakat Indonesia terhadap barang yang dihasilkan pabriknya. Jelaskan populasinya!
5. Pemerintah ingin mengetahui berapa rupiah penghasilan rata-rata setiap bulan dari pegawai-pegawai di seluruh Indonesia. Seorang pejabat mengatakan, bahwa untuk itu cukup dicatat atau diselidiki gaji para pegawai dari Kantor Bendahara Negara sebagai populasinya. Beri komentar tentang pernyataan pejabat tersebut!
6. Apakah yang dimaksud dengan statistika deskriptif? Apa pula yang dimaksud dengan statistika inferensial? dengan menggunakan contoh, bedakanlah antara keduanya!
7. Jelaskan apa yang dimaksud dengan :
   a. Data intern
   b. Data ekstern
   c. Data primer
   d. Data sekunder
   e. Sensus
   f. Sampel
   g. Sampling
   h. Data kualitatif
   i. Data kuantitatif
8. Mengapa penelitian secara sensus tidak selalu dapat dilakukan?
9. Andaikan kita sebagai pengusaha ingin mengetahui sikap 30 pegawai yang bekerja di perusahaan kita. Apakah dalam hal ini akan dilakukan sensus ataukah sampling? Apakah akan diadakan penyelidikan langsung wawancara atau melalui kuesioner? Jelaskan pendapat anda!
10. Berikan contoh-contoh penelitian dimana kita akan terpaksa harus menggunakan sampling dan bukan sensus!
11. Apa yang dimaksud dengan *non-probability sampling*? Sebutkan contohnya!
12. Jelaskan perbedaan antara *stratified sampling* dan *cluster sampling*. Berikan contohnya!
13. Sebutkan dan jelaskan jenis-jenis skala pengukuran variabel! Berikan contohnya!

Distribusi Frekuensi & Penyajian Data

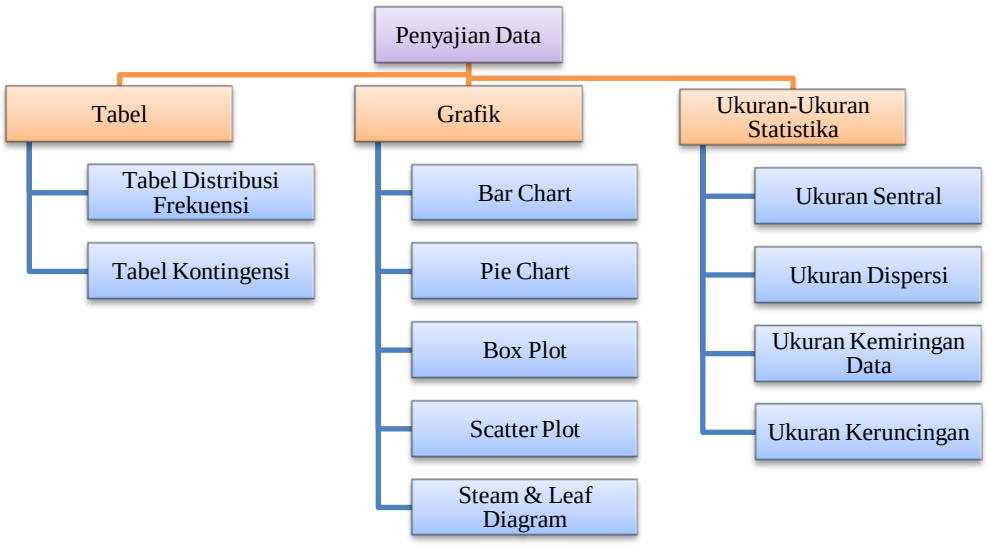
2.1. Penyajian Data

Data yang telah dikumpulkan, baik berasal dari populasi ataupun sampel, untuk keperluan laporan atau analisis selanjutnya, perlu diatur, disusun dan disajikan dalam bentuk yang jelas dan baik. Dengan kata lain data perlu disajikan secara sistematis dan rapi sehingga dapat dengan mudah dan cepat dimengerti oleh orang yang berkepentingan dengan data tersebut, misalnya para pengambil keputusan seperti manager, direktur dan lain-lain.

Misalkan sebuah perguruan tinggi mengirimkan 20 orang staf pengajarnya untuk mengikuti tes TOEFL. Hasil dari tes tersebut adalah sebagai berikut :

467	480	570	525	567	402	575
500	520	435	600	444	560	480
523	456	469	490	489	457	

Jika, kita hanya menyajikan data seperti yang terlihat di atas, maka informasi yang terdapat pada data tersebut akan sulit untuk dimengerti. Misalnya, berapakah rata-rata skor TOEFL dari 20 orang staf tersebut? Berapa orang yang mendapatkan nilai skor 548-584?. Agar informasi yang terdapat dalam sebuah data dapat dengan mudah dan cepat dimengerti maka diperlukan sebuah penyajian data. Data dapat disajikan dalam bentuk tabel, grafik maupun dalam bentuk ukuran-ukuran statistika.



Penyajian data melalui sebuah tabel dan grafik, saling berkaitan. Hal ini dikarenakan, pada dasarnya sebeluat dibuat grafik terlebih dahulu dibuat tabel. Penyajian data melalui grafik dikatakan lebih komunikatif karena dalam waktu singkat seseorang dapat dengan mudah memperoleh gambaran dan kesimpulan mengenai suatu keadaan. Dengan membaca data pada tabel dan grafik, para eksekutif akan dengan cepat dan mudah mengetahui situasi dan kondisi perusahaannya, sehingga dapat diambil tindakan-tindakan atau kebijakan-kebijakan yang tepat.

2.1.1. Penyajian Data dengan Tabel

Tabel atau daftar merupakan kumpulan angka yang disusun menurut kategori-kategori atau karakteristik-karakteristik data sehingga memudahkan analisis (DR. Boediono, 2014). Misalnya, jumlah penduduk menurut jenis kelamin, jumlah kendaraan menurut umur, mesin dan ukuran kendaraan. Terdapat tiga jenis tabel yang bisa digunakan untuk menyajikan data yaitu tabel satu arah, tabel dua arah dan tabel tiga arah. Untuk lebih jelasnya, perhatikan contoh tabel-tabel berikut.

Tabel 2.1. Tabel Satu Arah
Rata-Rata Banyaknya Anggota Rumah Tangga yang pernah yang pernah mengakses Internet dalam Tiga Bulan Terakhir Tahun 2014

PROVINSI	RATA-RATA
DKI JAKARTA	1.98
JAWA BARAT	1.70
JAWA TENGAH	1.52
DI YOGYAKARTA	1.63
JAWA TIMUR	1.53
BANTEN	1.84

Sumber : https://www.bps.go.id/linkTableDinamis/view/id/989

Tabel 2.2. Tabel Dua Arah
Rata-Rata Banyaknya Anggota Rumah Tangga yang pernah yang pernah mengakses Internet dalam Tiga Bulan Terakhir Tahun 2014

PROVINSI	Kota	Desa
DKI JAKARTA	1.98	-
JAWA BARAT	1.78	1.35
JAWA TENGAH	1.64	1.34
DI YOGYAKARTA	1.66	1.51
JAWA TIMUR	1.66	1.31
BANTEN	1.91	1.32

Sumber : https://www.bps.go.id/linkTableDinamis/view/id/989

Tabel 2.3. Tabel Tiga Arah
Rata-Rata Persentase Waktu Penyiaran dalam Seminggu Menurut
Jenis Program/Acara dan Jenis Kegiatan, 2012-2013

	Penyiaran Radio Pemerintah		Penyiaran Radio Swasta		Penyiaran dan Pemrograman Televisi Pemerintah		Penyiaran dan Pemrograman Televisi Swasta	
	2012	2013	2012	2013	2012	2013	2012	2013
Berita dan Informasi	27,19	25,66	15,74	15,79	36,67	33,50	22,82	25,49
Pendidikan dan Ilmu Pengetahuan	10,03	10,29	8,70	7,65	12,17	8,75	9,31	8,89
Seni dan Budaya	8,36	8,36	6,38	8,43	7,67	6,87	8,62	7,72
Agama	8,39	8,78	9,50	9,04	10,17	10,63	9,56	9,84
Olahraga	3,53	4,10	2,65	3,16	5,17	2,12	4,13	3,95
Musik	27,50	27,75	38,40	37,66	7,67	14,25	13,74	15,78
Sandiwara	1,06	0,90	0,80	0,79	5,00	1,38	11,79	9,37
Hiburan Lainnya	4,64	4,37	6,16	6,11	4,50	7,00	8,10	8,40
Talk Show	5,61	6,59	6,42	5,52	7,83	6,88	9,38	8,08
Lainnya	3,69	3,20	5,23	5,85	3,17	8,62	2,54	2,48

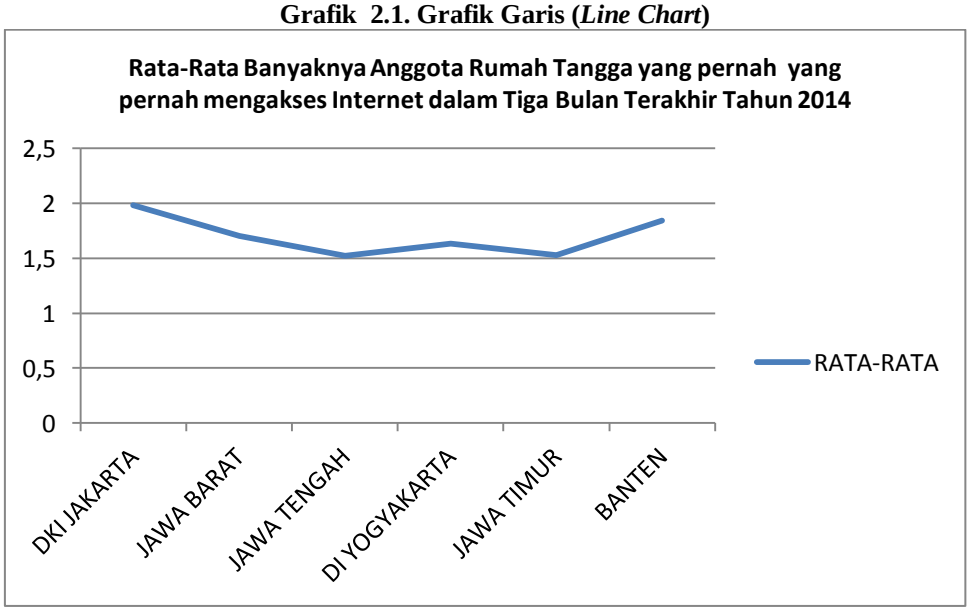
Sumber : https://www.bps.go.id/linkTabelStatis/view/id/1843

2.1.2. Penyajian Data dengan Grafik

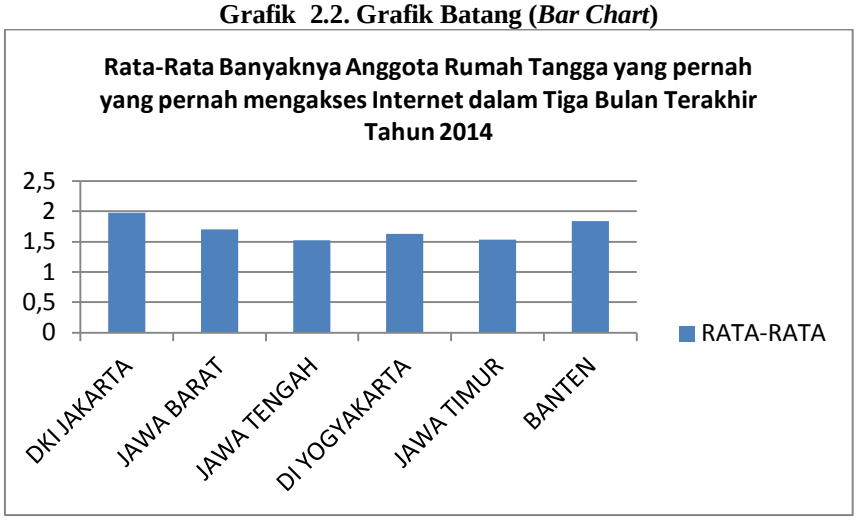
Seperti yang telah disebutkan sebelumnya, penyajian data dengan grafik lebih komunikatif dan dalam waktu yang singkat dapat memberikan gambaran suatu keadaan yang memerlukan sebuah keputusan. Secara visual grafik merupakan gambar –gambar yang menunjukkan data berupa angka dan biasanya dibuat berdasarkan tabel yang telah ada sebelumnya. Terdapat beberapa grafik yang dapat digunakan dalam penyajian sebuah data yaitu grafik garis (*line chart*), grafik batang (*bar chart*), grafik lingkaran (*pie chart*), *box-plot*, *scatter plot* dan *steam & leaf diagram*.

Perhatikan contoh-contoh grafik berikut :

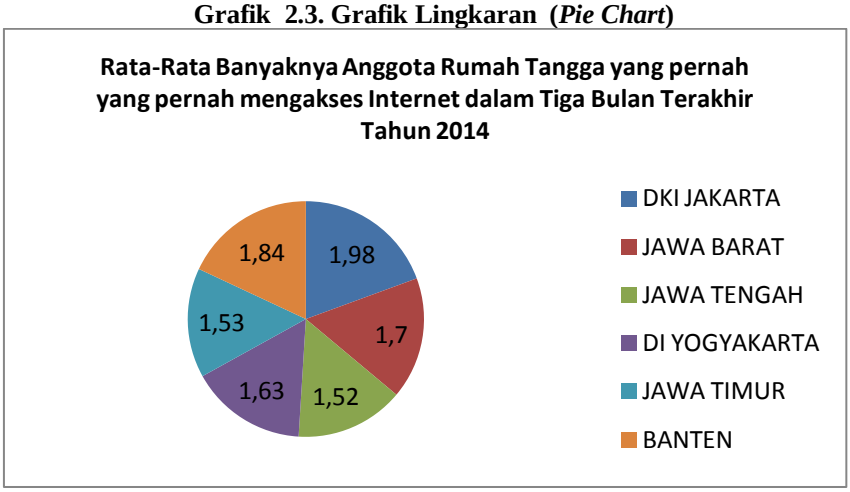
Grafik 2.1. Grafik Garis (*Line Chart*)



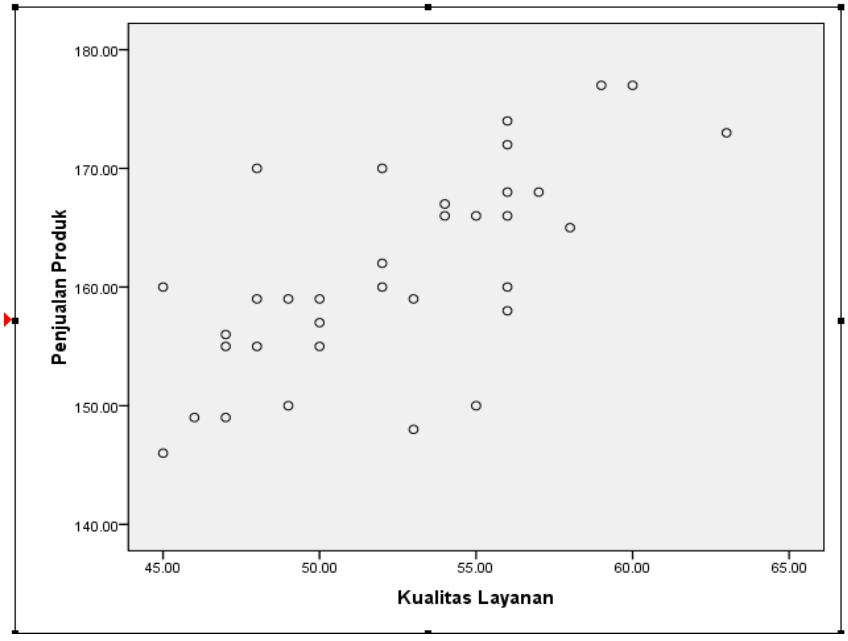
Grafik 2.2. Grafik Batang (*Bar Chart*)



Grafik 2.3. Grafik Lingkaran (*Pie Chart*)



Grafik 2.4. *Scatter Plot*



Grafik 2.5. ***Steam and Leaf Diagram***

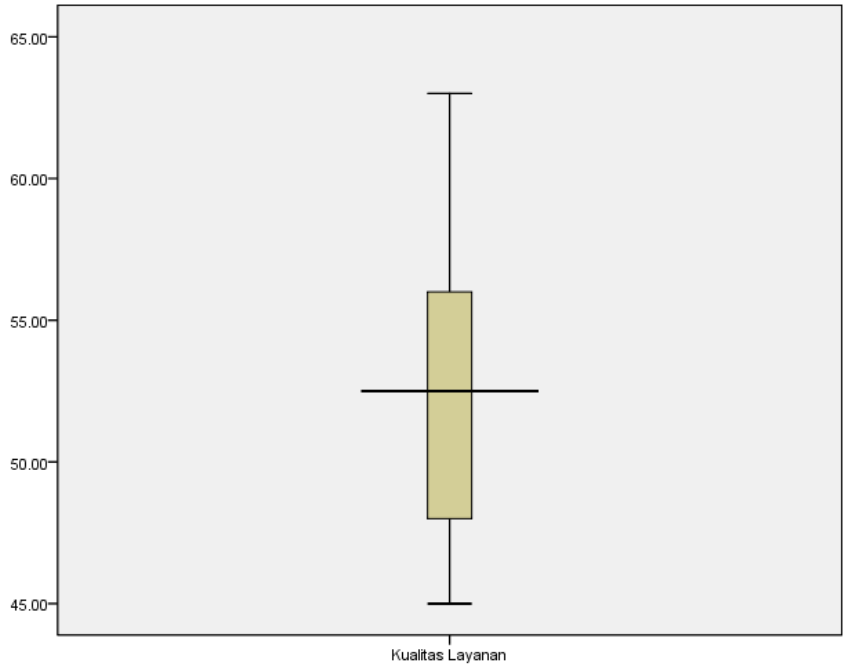
```
Kualitas Layanan Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

      ,00        4 .
    11,00        4 .  55677788899
    10,00        5 .  0002223344
    11,00        5 .  55666666789
     2,00        6 .  03

 Stem width:     10,00
 Each leaf:       1 case(s)
```

Grafik 2.6. ***Box-Plot***



2.2. Distribusi Frekuensi

A frequency distribution is a listing of frequencies of all categories of observed value of variable (Rudolf J.Freud, 2003). Sedangkan menurut DR. Boediono (2014), suatu pengelompokkan atau penyusunan data menjadi tabulasi data memakai kelas-kelas data dan dikaitkan dengan masing-masing frekuensinya disebut distribusi frekuensi atau tabel frekuensi. Tabel berikut adalah contoh tabel frekuensi dari data tinggi badan Mahasiswa di sebuah perguruan tinggi.

Tinggi Badan	Frekuensi
151-153	3
154-156	7
157-159	12
160-162	18
163-165	2
166-168	17
169-171	11
172-174	5

Dari tabel frekuensi di atas, kita dapat dengan mudah memperoleh informasi mengenai kebanyakan tinggi badan mahasiswa di perguruan tinggi tersebut berada pada rentang 160-162 cm, atau terdapat 18 orang yang memiliki tinggi badan pada rentang 160-162 cm. Dalam statistika, rentang nilai tinggi badan tersebut disebut sebagai kelas modus.

Untuk dapat menyajikan sebuah data (data mentah) menjadi sebuah distribusi frekuensi atau tabel frekuensi, berikut adalah langkah-langkahnya :

1. Susun data mentah dari nilai terendah ke nilai tertinggi
2. Hitunglah nilai *range*
3. Tentukan jumlah kelas
4. Tentukan interval kelas
5. Bentuk kelas-kelas distribusi frekuensi
6. Memasukkan nilai-nilai observasi ke dalam kelas-kelas yang sesuai
7. Menghitung *mid point*, tepi kelas dan frekuensi kumulatif.

Perhatikan contoh berikut, dengan menggunakan data skor TOEFL dari 20 orang staf pada sub bab sebelumnya, kita akan membuat tabel distribusi frekuensi dengan langkah seperti yang sudah disebutkan sebelumnya.

1. Urutkan nilai data dari nilai terkecil ke nilai tertinggi

   402 ; 435 ; 444 ; 456 ; 457 ; 467 ; 469 ; 480 ; 480; 489 ; 490; 500 ; 520 ; 523 ; 525 ; 560 ; 567 ; 570 ; 575 ; 600
2. Hitung nilai *range*

   Range (*r*) = Nilai tertinggi – nilai terendah

   Range (*r*) = 600 – 402 = 198

3. Hitung jumlah kelas

   Jumlah kelas (k) = 1 + 3,322 log n

   Jumlah kelas (k) = 1 + 3,322 log *(20)* = 5,322 ~ 6

4. Hitung Interval Kelas (C_i)

$$C_i = \frac{r}{k} = \frac{198}{5,322} = 37,2 \sim 37$$

5. Membentuk kelas-kelas distribusi frekuensi

   Untuk membentuk kelas-kelas interval pilih ujung bawah kelas interval pertama dengan mengambil nilai data terendah atau nilai data yang lebih kecil dari data terendah tetapi selisihnya harus kurang dari interval kelas yang ditentukan.

   Data terkecil = 402

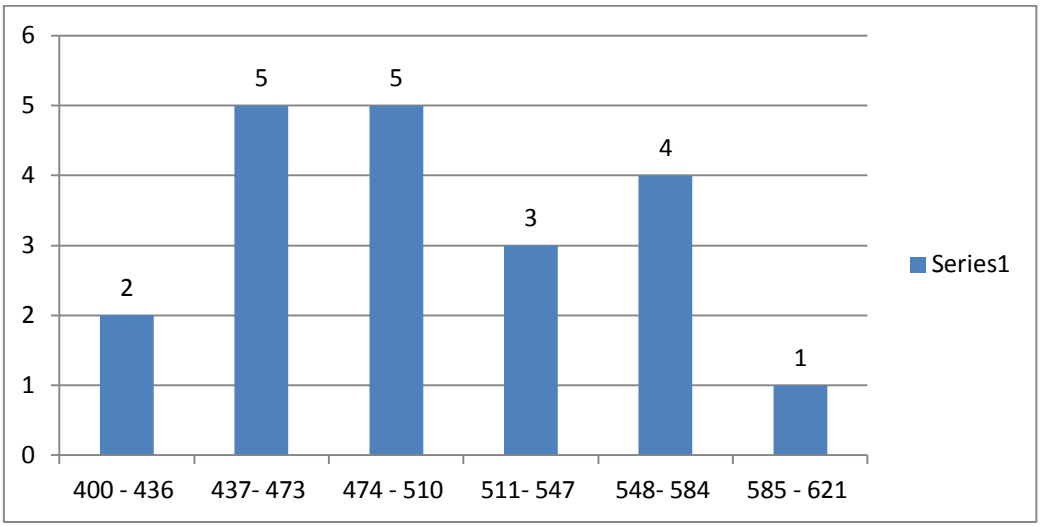
   Ujung bawah kelas interval yang dipilih = 400

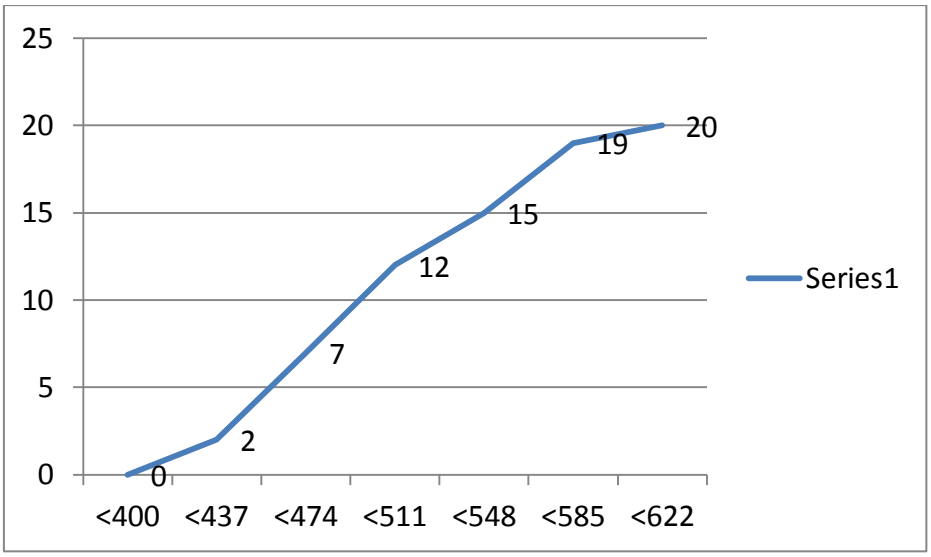
6. Bentuk kelas-kelas interval

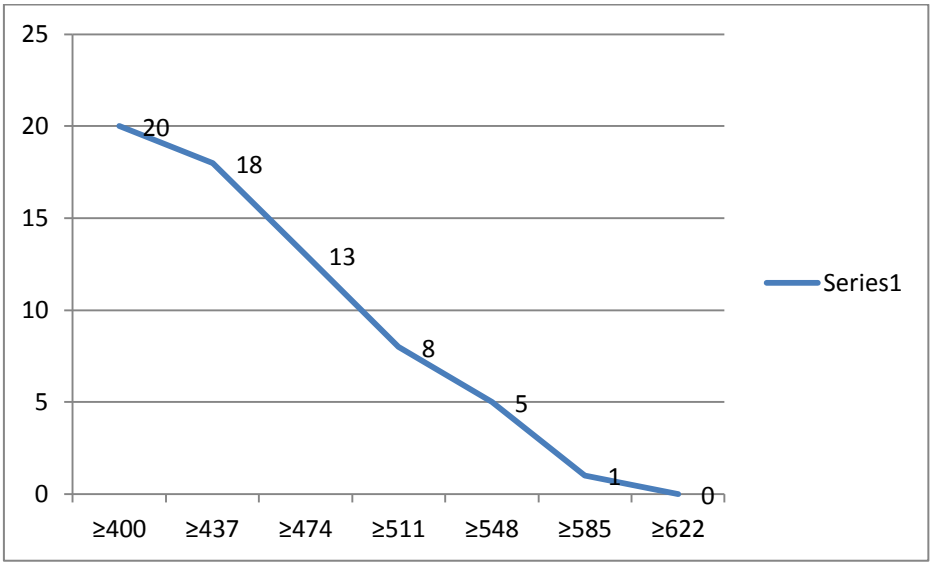
No Kelas	Ujung bawah kelas Interval	Ujung Atas Kelas Interval	Kelas Interval
1	**400**	400 + (37-1) = 436	400 - 436
2	436 + 1 = 437	437 + (37-1) = 473	437- 473
3	473 + 1 = 474	474 + (37-1) = 510	474 - 510
4	510 + 1 = 511	511 + (37-1) = 547	511- 547
5	547 + 1 = 548	548 + (37-1) = 584	548- 584
6	584 + 1 = 585	585 + (37-1) = 621	585 - 621

7. Memasukkan nilai-nilai observasi ke dalam kelas-kelas yang sesuai serta menghitung *mid point*, tepi kelas dan frekuensi kumulatif.

No Kelas	Kelas Interval	Tepi Kelas	Batas Kelas	Frekuensi	*X*	fk<		fk>		Frekuensi Relatif
		399.5								
1	400 - 436	436.5	399.5-436.5	2	418	<400	0	≥400	20	10%
2	437- 473	473.5	436.5-473.5	5	455	<437	2	≥437	18	25%
3	474 - 510	510.5	473.5-510.5	5	492	<474	7	≥474	13	25%
4	511- 547	547.5	510.5-547.5	3	529	<511	12	≥511	8	15%
5	548- 584	584.5	547.5-584.5	4	566	<548	15	≥548	5	20%
6	585 - 621	621.5	584.5-621.5	1	603	<585	19	≥585	1	5%
						<622	20	≥622	0	


6
5
4
3
2
1
0
5
5
4
3
2
1
400 - 436
437- 473
474 - 510
511- 547
548- 584
585 - 621
Series1


25
20
15
10
5
0
0
2
7
12
15
19
20
<400
<437
<474
<511
<548
<585
<622
Series1


25
20
15
10
5
0
20
18
13
8
5
1
0
≥400
≥437
≥474
≥511
≥548
≥585
≥622
Series1

LATIHAN

1. Mengapa data yang diperoleh melalui sensus ataupun sampling perlu disajikan dengan memakai tabel dan grafik? Jelaskan!
2. Mengapa penyajian data dengan grafik lebih baik dibandingkan dengan tabel?
3. Sebutkan beberapa cara untuk menyajikan data dengan tabel?
4. Sebutkan beberapa cara untuk menyajikan data dengan grafik?
5. Berikan beberapa contoh tabel satu arah, dua arah dan tiga arah untuk menggambarkan suatu karakteristik tertentu!
6. Data berikut merupakan hasil penjualan sebuah produk di 45 Toko :

   96 139 112 118 94 93 142 135
   136 127 88 94 132 125 119 117
   107 111 143 148 127 125 125 120
   117 95 155 104 103 97 113 155
   106 113 156 96 103 139 124 120
   108 112 134 138 89

   Buatlah Tabel Distribusi frekuensi dan grafiknya.

Ukuran Deskriptif

3.1. Ukuran Pemusatan Data

Ukuran pemusatan data (nilai sentral) menunjukkan dimana suatu data memusat atau suatu kumpulan pengamatan memusat (mengelompok) atau merupakan nilai tunggal yang mewakili semua data atau kumpulan pengamatan dan nilai tersebut menunjukkan pusat data (DR. Boediono, 2014).

Macam-macam ukuran pemusatan data diantaranya rata-rata hitung, median, modus, rata-rata ukur, rata-rata harmonis, kuartil, desil dan persentil.

3.1.1. Rata-Rata Hitung

Rata-rata hitung dikenal juga sebagai rata-rata atau *mean*. *The mean is the sum of all the observed values divided by the number of values* (Rudolf, J.F.,2003). Nilai rata-rata hanya dapat diperoleh dari data kuantitatif. Rata-rata atau lengkapnya rata-rata hitung, untuk data kuantitatif yang terdapat dalam sebuah sampel dihitung dengan jalan membagi jumlah nilai data oleh banyak data (Sudjana, 2005). Dengan demikian rata-rata didefinisikan sebagai:

$$Rata - rata = \frac{Jumlah\ nilai\ data}{banyak\ data} \tag{3.1}$$

Misalkan, nilai-nilai data kuantitatif dinyatakan sebagai x_1, x_2, x_3,.., x_n, apabila dalam kumpulan data itu terdapat n buah nilai, maka rata-rata yang dinotasikan sebagai $\bar{x}$ didefinisikan menjadi :

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{3.2}$$

Perhatikan contoh berikut, misalkan nilai ujian statistika dari sebagian mahasiswa dalam sebuah kelas adalah 70,75,60,65, 80 maka nilai rata-rata nya adalah :

$$\bar{x} = \frac{70+75+60+65+80}{5} = 70$$

Bila sebuah data dimana masing-masing nilai data mengulang dengan frekuensi tertentu, katakanlah nilai x_1 ada sebanyak f_1, x_2 ada sebanyak f_2, ..., dan x_n ada sebanyak f_n, maka nilai rata-rata nya didefinisikan sebagai :

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} \tag{3.3}$$

Misalkan pada suatu ujian Bahasa Inggris, ada 3 mahasiswa mendapat nilai 60, 5 mahasiswa mendapat nilai 65, 4 mahasiswa mendapat nilai 80, 1 mahasiswa mendapat nilai 50 dan 2 mahasiswa mendapat nilai 95. Maka nilai rata-ratanya adalah :

$$\bar{x} = \frac{(3\times 60)+(5\times 65)+(4\times 80)+(1\times 50)+(2\times 95)}{3+5+4+1+2} = \frac{1065}{15} = 71$$

Untuk data yang disusun dalam tabel distribusi frekuensi, rata-rata dapat dihitung melalui dua cara yaitu menggunakan persamaan (3.3) dimana x_i merupakan nilai tengah kelas ke-i atau dengan menggunakan koding sebagai berikut :

$$\bar{x} = x_0 + c\left(\frac{\sum_{i=1}^{n} f_i u_i}{\sum_{i=1}^{n} f_i}\right) \tag{3.4}$$

dengan : x_0 : nilai tengah kelas (paling tengah) yang berhimpit dengan nilai u=0

c : lebar kelas

u : kode kelas

nilai u ditentukan dengan aturan sebagai berikut :

- Untuk kelas (paling tengah) yaitu x_0 nilai u yang diberikan adalah nol (0)
- Kode kelas yang lebih kecil dari kelas x_0 bernilai negatif yaitu -1, -2, -3, dst.
- Kode kelas yang lebih besar dari kelas x_0 bernilai positif yaitu 1, 2, 3, dst.

Perhatikan contoh berikut, misalkan modal (dalam jutaan rupiah) dari 40 perusahaan disajikan dalam tabel distribusi frekuensi berikut :

Modal	Frekuensi (f_i)	Nilai Tengah (x_i)	$f_i\, x_i$
112-120	4	116	464
121-129	5	125	625
130-138	8	134	1072

Modal	Frekuensi (f_i)	Nilai Tengah (x_i)	$f_i\, x_i$
139-147	12	143	1716
148-156	5	152	760
157-165	4	161	644
166-174	2	170	340
	$\sum_{i=1}^{7} f_i = 40$		$\sum_{i=1}^{7} f_i x_i = 5621$

Sehingga rata-ratanya dapat dihitung sebagai berikut :

$$\bar{x} = \frac{\sum_{i=1}^{7} f_i x_i}{\sum_{i=1}^{7} f_i} = \frac{5621}{40} = 140,525$$

Jika data modal tersebut dihitung menggunakan persamaan (3.4), maka :

Modal	Frekuensi (f_i)	Nilai Tengah (x_i)	Kode Kelas (u_i)	$f_i\, u_i$
112-120	4	116	-3	-12
121-129	5	125	-2	-10
130-138	8	134	-1	-8
139-147	12	143	0	0
148-156	5	152	1	5
157-165	4	161	2	8
166-174	2	170	3	6
	$\sum_{i=1}^{7} f_i = 40$			$\sum_{i=1}^{7} f_i u_i = -11$

Dengan demikian nilai rata-ratanya :

$$\bar{x} = 143 + 9\left(\frac{-11}{40}\right) = 143 - 2,475 = 140,525$$

Selain dalam bentuk frekuensi, nilai-nilai data dari data kuantitatif masing-masing mempunyai bobot atau timbangan. Katakanlah nilai x_1 mempunyai bobot w_1, x_2 mempunyai bobot w_2, ..., dan x_n mempunyai bobot w_n, maka nilai rata-rata nya didefinisikan sebagai :

$$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \qquad (3.5)$$

Misalkan, pada akhir semester untuk mata kuliah Riset Operasional diketahui bahwa siswa bernama Amir mempunyai nilai terstruktur dengan rincian UAS adalah 65, UTS adalah 70, dan nilai tugas adalah 85. Pihak Universitas menentukan bahwa bobot nilai untuk UAS adalah 3, bobot UTS adalah 2 dan bobot nilai tugas adalah 1. Berdasarkan bobot masing-masing nilai tersebut maka rata-rata nilai akhir semester yang diperoleh oleh Amir adalah sebagai berikut :

$$\bar{x} = \frac{\sum_{i=1}^{3} w_i x_i}{\sum_{i=1}^{3} w_i} = \frac{(3\times 65)+(2\times 70)+(1\times 85)}{3+2+1} = \frac{420}{6} = 70$$

Pada dasarnya, perhitungan rata-rata melalui persamaan (3.3) dan (3.5) adalah sama. Perbedaan dari kedua persamaan terletak pada frekuensi dan bobot. Perhatikan contoh kedua berikut, andaikan terdapat 300 orang pegawai, 800 orang masing-masing berupah Rp. 5.260,- per bulan, 1000 orang masing-masing berupah Rp. 5.475,- per bulan dan sisanya masing-masing berupah Rp. 5.778,- per bulan. Upah rata-rata ke-3000 pegawai berikut :

$$\bar{x} = \frac{\sum_{i=1}^{3} w_i x_i}{\sum_{i=1}^{3} w_i} = \frac{(800\times 5260)+(1000\times 5474)+(1200\times 5778)}{3000} = 5538,87$$

3.1.2. Median

Median adalah nilai yang terletak di tengah suatu data yang telah diurutkan dari nilai terkecil hingga terbesar. *The median of a set values is defined to be the middle value when the measurements are arranged from lowest to highest, that is, 50% of the measurement lie above it and 50% fall below it* (Rudolf, J.F.,2003). Sejalan dengan yang diungkapkan Rudolf, menurut DR.Boediono (2014) median adalah nilai yang paling tengah untuk data ganjil atau rata-rata dari dua nilai tengah jika banyaknya data genap. Sehingga median yang dinotasikan sebagai *Med* dapat didefinisikan sebagai :

$$Med = \begin{cases} X_{\frac{n+1}{2}} & \text{, untuk data ganjil} \\ \dfrac{X_{\frac{n}{2}} + X_{\frac{n+2}{2}}}{2} & \text{, untuk data genap} \end{cases} \qquad (3.6)$$

Perhatikan contoh berikut, median dari data : 3,4,4,5,6,8,8,9,10 adalah :

$$X_{\frac{n+1}{2}} = X_{\frac{9+1}{2}} = X_5$$

Artinya median dari data di atas adalah data ke-5 yaitu 6.

Contoh berikutnya, tentukanlah median dari data upah 8 karyawan (dalam ribuan rupiah) berikut ini, 20,80,75,60,50,85,45,90.

Untuk menentukan median, terlebih dahulu urutkan data dari terkecil sampai terbesar sebagai berikut , 20,45,50,60,65,75,80,85,90. Oleh karena banyaknya data genap yaitu 8 maka mediannya adalah :

$$\frac{X_{\frac{n}{2}} + X_{\frac{n+2}{2}}}{2} = \frac{X_{\frac{8}{2}} + X_{\frac{8+2}{2}}}{2} = \frac{X_4 + X_5}{2}$$

Dengan kata lain, mediannya adalah nilai rata-rata dari data ke-4 dan ke-5 yaitu :

$$Med = \frac{X_4 + X_5}{2} = \frac{60+75}{2} 67,5$$

Artinya, median upah 8 karyawan tersebut adalah Rp.67.500,-

Untuk data yang dinyatakan dalam bentuk tabel distribusi frekuensi, mediannya dihitung menggunakan persamaan berikut :

$$Med = L_0 + c\left(\frac{\frac{n}{2} - F}{f}\right) \qquad (3.7)$$

Med = Median
L_0 = Batas bawah kelas median
c = Lebar kelas
n = Banyaknya data
F = Jumlah frekuensi semua kelas sebelum kelas yang mengandung median
f = Frekuensi Kelas median

Tentukan median dari data modal 40 perusahaan berikut :

Modal	Frekuensi (f_i)
112-120	4
121-129	5
130-138	8
139-147	12
148-156	5
157-165	4
166-174	2

Terlebih dahulu tentukan kelas interval dimana mediannya terletak. Dari tabel di atas, dapat dilihat bahwa median terletak pada :

$$\frac{X_{\frac{n}{2}} + X_{\frac{n+2}{2}}}{2} = \frac{X_{\frac{40}{2}} + X_{\frac{40+2}{2}}}{2} = \frac{X_{20} + X_{21}}{2}$$

Artinya, median adalah nilai rata-rata dari data ke 20 dan 21 yang terletak pada kelas interval 139-147. Sehingga diperoleh bahwa :

L_0 = 138,5
c = 9
n = 40
F = 4+4+8=17
f = 12

Maka :

$$Med = 138,5 + 9\left(\frac{\frac{40}{2} - 17}{12}\right) = 140,75$$

3.1.3. Modus

Untuk menyatakan gejala yang paling sering terjadi atau paling banyak muncul dipakai ukuran pemusatan data yang disebut modus. Untuk data kuantitatif, modul adalah nilai data yang paling banyak muncul atau nilai data yang mempunyai frekuensi paling tinggi. Suatu kelompok data mungkin mempunyai modus atau mungkin juga tidak mempunyai modus. Dengan kata lain, modus suatu data tidak selalu ada. Bila suatu data mempunyai modus, maka modusnya bisa lebih dari satu atau modusnya tidak tunggal.

Untuk data kuantitatif yang disusun dalam bentuk tabel distribusi frekuensi, modusnya diperoleh menggunakan persamaan berikut :

$$Mod = L_0 + c\left(\frac{b_1}{b_1 + b_2}\right) \qquad (3.8)$$

Mod = Modus

L_0 = Batas bawah kelas modus

c = Lebar kelas

b_1 = selisih antara frekuensi kelas modus dengan frekuensi tepat satu kelas sebelum kelas modus

b_2 = selisih antara frekuensi kelas modus dengan frekuensi tepat satu kelas sesudah kelas modus

Tentukan modus dari data modal 40 perusahaan berikut :

Modal	Frekuensi (f_i)
112-120	4
121-129	5
130-138	8
139-147	12
148-156	5
157-165	4
166-174	2

Terlebih dahulu tentukan kelas interval dimana modusnya terletak. Dari tabel di atas, dapat dilihat bahwa median terletak pada kelas interval 139-147. Sehingga diperoleh bahwa :

$L_0 = 138{,}5$

$c = 9$

b_1 = 12-8=4

b_2 = 12-5=7

Maka $Mod = 138{,}5 + 9\left(\frac{4}{4+7}\right) = 138{,}5 + 3{,}27 = 141{,}77$

Berikut adalah kelebihan dan kekurangan dari mean, median dan modus :

Ukuran Pemusatan	Kelebihan	Kekurangan
Rata-Rata Hitung	1. Mempertimbangkan semua nilai 2. Dapat menggambarkan mean populasi 3. Variasinya paling stabil 4. Cocok untuk data homogen	1. Peka atau mudah terpengaruh oleh nilai ekstrim 2. Kurang baik untuk data heterogen
Median	1. Tidak peka atau tidak terpengaruh oleh nilai ekstrim 2. Cocok untuk data heterogen	1. Tidak mempertimbangkan semua nilai 2. Kurang dapat menggambarkan mean populasi
Modus	1. Tidak peka atau tidak terpengaruh oleh nilai ekstrim 2. Cocok untuk data homogen ataupun heterogen	1. Kurang menggambarkan mean populasi 2. Modus bisa lebih dari satu

Sementara hubungan antara ketiga ukuran pemusatan data tersebut didefinisikan sebagai berikut :

$$\bar{X} - Mo = 3\left(\bar{X} - Med\right) \tag{3.9}$$

3.1.4. Rata-Rata Ukur

Rata-rata ukur digunakan untuk menghitung rata-rata tingkat perubahan atau *rate of change* dan didefinisikan sebagai :

$$G = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdots x_n}$$

Perhatikan contoh berikut adalah data banyaknya penabung (nasabah) di sebuah Bank :

Tahun	Jumlah	Rasio pertambahan
1980	1000	
1981	2000	2x
1982	20000	10x

Berdasarkan data yang disajikan pada tabel di atas, Bank tersebut ingin mengetahui rata-rata rasio pertambahan banyaknya penabung (nasabah) sebagai berikut :

$$G = \sqrt[2]{2 \times 10} = 4,47$$

Artinya, rata-rata rasio pertambhan penabung (nasabah) pada Bank tersebut adalah 4,47 atau 4-5 orang per tahun.

Rata-rata ukur juga dapat dihitung melalui persamaan berikut :

$$G = Anti\log\left[\frac{\left(\sum_{i=1}^{n}\log x_i\right)}{n}\right] \quad \text{atau} \quad G = Anti\log\left[\frac{\left(\sum_{i=1}^{n} f_i \log x_i\right)}{n}\right]$$

3.1.5. Rata-Rata Harmonis

Rata-rata harmonis merupakan rata-rata yang digunakan jika suatu kelompok data mempunyai ciri-ciri tertentu yang merupakan bilangan pecahan atau bilangan desimal. Rata-rata harmonis didefinisikan sebagai :

$$R_H = \frac{n}{\sum_{i=1}^{n}\left(\frac{1}{x_i}\right)} \qquad \text{atau} \qquad R_H = \frac{\sum_{i=1}^{n} f_i}{\sum_{i=1}^{n}\left(\frac{f_i}{x_i}\right)}$$

Sebagai contoh, rata-rata harmonis dari data :

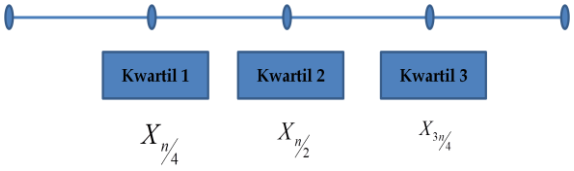
$$\frac{1}{3};\frac{2}{5};\frac{3}{7};\frac{4}{9}$$

Adalah :

$$R_H = \frac{n}{\sum_{i=1}^{4}\left(\frac{1}{x_i}\right)} = \frac{4}{\left(3+\frac{5}{2}+\frac{7}{3}+\frac{9}{4}\right)} = 0,397$$

3.1.6. Kuartil

Nilai kuartil adalah nilai yang membagi nilai observasi suatu data yang telah diurukan dari nilai terendah sampai nilai tertinggi menjadi empat bagian :



Berdasarkan gambar di atas, jelas bahwa dalam membagi sebuah data menjadi 4 buah bagian sama akan terdapat 3 kuartil (Q). Kuartil 1 (Q_1) adalah bilangan sehingga 25% dari data lebih kecil atau sama dengan bilangan itu sendiri. Kuartil 2 (Q_2) adalah bilangan yang sama dengan median, yaitu bilangan yang membagi data menjadi 2 bagian yang sama sehingga 50% data berada di atas bilangan tersebut dan 50% lainnya berada di bawahnya. Sementara

kuartil 3 (Q_3) adalah bilangan sehingga 75% dari data lebih kecil atau sama dengan bilangan tersebut. Kuartil ke-*i* dari sebuah data didefinisikan sebagai :

$$Q_i = \text{Nilai yang ke-}\ \frac{i(n+1)}{4}\ ;\ \text{i=1,2,3}$$

Perhatikan data upah bulanan karyawan (dalam ribuan rupiah) berikut :

40, 30, 50, 65, 45, 55, 70, 60, 80, 35, 85, 95, 100

Untuk dapat mencari nilai-nilai kuartil dari data tersebut, terlebih dahulu data diurutkan dari nilai terendah sampai dengan nilai tertinggi sebagai berikut :

30, 35, 40, 45, 50, 55, 60, 65, 75, 80, 85, 95, 100

Maka nilai kuartil-kuartilnya adalah :

$$Q_1 = \text{Nilai yang ke-}\ \frac{1(13+1)}{4} = \text{Nilai yang ke-}\ \frac{14}{4}$$
$$= \text{Nilai yang ke-}\ 3\frac{1}{2}$$
$$= \frac{\text{Nilai yang ke-}\ 3 + \text{Nilai yang ke-}\ 4}{2}$$
$$= \frac{40+45}{2}$$
$$= 42,5$$

$$Q_2 = \text{Nilai yang ke-}\ \frac{2(13+1)}{4} = \text{Nilai yang ke-}\ \frac{28}{4}$$
$$= \text{Nilai yang ke-}\ 7$$
$$= 60$$

$$Q_3 = \text{Nilai yang ke-}\ \frac{3(13+1)}{4} = \text{Nilai yang ke-}\ \frac{42}{4}$$
$$= \text{Nilai yang ke-}\ 10\frac{1}{2}$$
$$= \frac{\text{Nilai yang ke-}\ 10 + \text{Nilai yang ke-}\ 11}{2}$$
$$= \frac{80+85}{2}$$
$$= 82,5$$

Nilai kuartil untuk data berkelompok atau data yang disajikan dalam bentuk distribusi frekuensi diperoleh melalui persamaan berikut :

$$Q_i = L_0 + c\left(\frac{\frac{i \cdot n}{4} - F}{f}\right)$$; i=1,2,3

L_0 = batas bawah kelas kuartil

c = lebar kelas

F = jumlah frekuensi semua kelas sebelum kelas kuartil ke-i

f = frekuensi kelas kuartil ke=i

Sebagai contoh, berikut adalah data modal :

Modal	Frekuensi (f_i)	Nilai Tengah (x_i)
112-120	4	116
121-129	5	125
130-138	8	134
139-147	12	143
148-156	5	152
157-165	4	161
166-174	2	170

Q_1 adalah bilangan yang membagi data menjadi 25% ke bawah dan 75% ke atas, sehingga dengan n=40 Q_1 terletak pada kelas 130-138, maka :

L_0 = 129,5

c = 9

F = 4+5=9

f = 8

$$Q_1 = 129,5 + 9\left(\frac{\frac{1 \cdot 40}{4} - 9}{8}\right) = 129,5+9\left(\frac{10-9}{8}\right) = 130,625$$

Q_2 adalah bilangan yang membagi data menjadi 50% ke bawah dan 50% ke atas, sehingga dengan n=40 Q_2 terletak pada kelas 139-147, maka :

L_0 = 138,5

c = 9

F = 4+5+8=17

f = 12

$$Q_2 = 138,5 + 9\left(\frac{\frac{2\cdot 40}{4} - 17}{12}\right) = 138,5+9\left(\frac{20-17}{12}\right) = 140,75$$

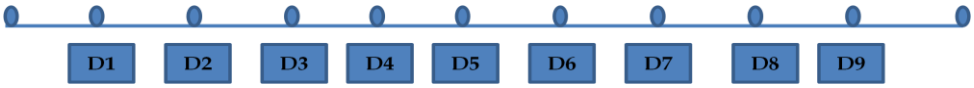
Q_3 adalah bilangan yang membagi data menjadi 75% ke bawah dan 25% ke atas, sehingga dengan n=40 Q_3 terletak pada kelas 148-156, maka :

L_0 = 147,5
c = 9
F = 4+5+8+12=29
f = 5

$$Q_3 = 147,5 + 9\left(\frac{\frac{3\cdot 40}{4} - 29}{5}\right) = 147,5+9\left(\frac{30-29}{5}\right) = 149,3$$

3.1.7. Desil

Nilai desil adalah nilai-nilai yang membagi nilai observasi suatu data yang telah diurutkan menjadi 10 bagian yang sama :



$$D_i = \text{Nilai yang ke-}\ \frac{i(n+1)}{10}\ ;\ \text{i=1,2,3,...,9}$$

Sebagai contoh, gunakan data yang sama pada perhitungan kuartil, yaitu data upah pegawai yang telah diurutkan sebagai berikut :

30, 35, 40, 45, 50, 55, 60, 65, 75, 80, 85, 95, 100

Nilai Desil ke-3 dari data tersebut adalah :

$$D_3 = \text{Nilai yang ke-}\ \frac{3(13+1)}{10} = \text{Nilai yang ke-}\ \frac{42}{10}$$

$$= \text{Nilai yang ke-}\ 4\frac{1}{5}$$

$$= \text{Nilai yang ke-}\ 4 + \frac{1}{5}(\text{Nilai yang ke-}\ 5\ - \text{Nilai yang ke-}\ 4)$$

$$= 45 + \frac{1}{5}(50-45)$$

$$= 46$$

Artinya, 30% dari data upah lebih kecil dari atau sama dengan 46 (ribuan rupiah).

Nilai Desil ke-7 dari data tersebut adalah :

$$D_7 = \text{Nilai yang ke-}\ \frac{7(13+1)}{10} = \text{Nilai yang ke-}\ \frac{98}{10}$$

$$= \text{Nilai yang ke-}\ 9\frac{8}{10}$$

$$= \text{Nilai yang ke-}\ 9 + \frac{8}{10}\left(\text{Nilai yang ke-}\ 10\ - \text{Nilai yang ke-}\ 9\right)$$

$$= 70 + \frac{8}{10}(80 - 70)$$

$$= 78$$

Artinya, 70% dari data upah lebih kecil dari atau sama dengan 78 (ribuan rupiah).

Nilai Desil untuk data berkelompok atau data yang disajikan dalam bentuk distribusi frekuensi diperoleh melalui persamaan berikut :

$$D_i = L_0 + c\left(\frac{\frac{i \cdot n}{10} - F}{f}\right) ; \text{i=1,2,3,...,9}$$

L_0 = batas bawah kelas desil ke-i
c = lebar kelas
F = jumlah frekuensi semua kelas sebelum kelas desil ke-i
f = frekuensi kelas desil ke=i

Masih menggunakan data yang sama yaitu data modal :

Modal	Frekuensi (f_i)	Nilai Tengah (x_i)
112-120	4	116
121-129	5	125
130-138	8	134
139-147	12	143
148-156	5	152
157-165	4	161
166-174	2	170

D_3 adalah bilangan yang membagi data menjadi 30% ke bawah dan 70% ke atas, sehingga dengan n=40 D_3 terletak pada kelas 130-138, maka :

L_0 = 129,5
c = 9
F = 4+5=9
f = 8

$$D_1 = 129,5 + 9\left(\frac{\frac{3\cdot 40}{4} - 9}{8}\right) = 129,5+9\left(\frac{12-9}{8}\right) = 132,875$$

3.1.8. Persentil

Nilai persentil adalah nilai-nilai yang membagi observasi suatu data yang telah diurutkan dari nilai terkecil hingga terbesar menjadi 100 bagian yang sama. Dengan demikian data akan memiliki 99 nilai persentil. Cara perhitungan persentil sama dengan perhitungan kuartil dan desil. Perbedaanya hanya persamaan yang digunakan. Persentil ke-*i* dari sebuah data didefinisikan sebagai :

$$P_i = \text{Nilai yang ke-}\ \frac{i(n+1)}{100}\ ;\ \text{i=1,2,3,...,99}$$

Nilai Persentil untuk data berkelompok atau data yang disajikan dalam bentuk distribusi frekuensi diperoleh melalui persamaan berikut :

$$D_i = L_0 + c\left(\frac{\frac{i\cdot n}{100} - F}{f}\right)\ ;\ \text{i=1,2,3,...,99}$$

L_0 = batas bawah kelas persentil ke-i
c = lebar kelas
F = jumlah frekuensi semua kelas sebelum kelas persentil ke-i
f = frekuensi kelas persentil ke=i

3.2. Ukuran Penyebaran Data

Ukuran penyebaran suatu kelompok data terhadap pusat data disebut dispersi atau variasi atau keragaman data (DR. Boediono, 2014) Ukuran dispersi terbagi menjadi dua yaitu dispersi mutlak dan dispersi relatif. Dispersi mutlak diantanya range, simpangan rata-rata dan varians. Sedangkan dispersi relatif salah satunya adalah koefisien variasi.

3.2.1. Range

Range merupakan selisih nilai maksimum dengan nilai nilai minimum.

3.2.2. Simpangan Rata-Rata

Simpangan rata-rata adalah jumlah nilai mutlak dari selisih semua nilai dengan nilai rata-rata dibagi banyaknya data, atau untuk data tidak berkelompok didefiinisikan sebagai :

$$SR = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$

Perhatikan contoh berikut adalah data produksi sebuah mesin dengan merk tertentu :

20, 30, 50, 70, 80.

Untuk dapat menghitung nilai simpangan rata-rata dari tersebut, terlebih dahulu lakukanperhitungan nilai rata-rata sebagai berikut :

$$\bar{x} = \frac{20+30+50+70+80}{5} = 50$$

Maka simpangan rata-rata dari data di atas adalah :

$$SR = \frac{\sum_{i=1}^{5} |x_i - \bar{x}|}{n}$$

$$= \frac{|20-50|+|30-50|+|50-50|+|70-50|+|80-50|}{5}$$

$$= \frac{30+20+0+20+30}{5}$$

$$= \frac{100}{5}$$

$$= 20$$

Untuk data berkelompok didefinisikan sebagai :

$$SR = \frac{\sum_{i=1}^{n} f_i |x_i - \bar{x}|}{\sum_{i=1}^{n} f_i}$$

Gunakan data mengenai modal sebagai contoh perhitungan simpangan rata-rata berikut :

Modal	Frekuensi (f_i)	Nilai Tengah (x_i)	$\lvert x_i - \bar{x} \rvert$	$f_i \lvert x_i - \bar{x} \rvert$
112-120	4	116	24,525	98,100
121-129	5	125	15,525	77,625
130-138	8	134	6,525	52,200

Modal	Frekuensi (f_i)	Nilai Tengah (x_i)	$\lvert x_i - \bar{x} \rvert$	$f_i \lvert x_i - \bar{x} \rvert$
139-147	12	143	2,475	29,700
148-156	5	152	11,475	57,375
157-165	4	161	20,475	81,900
166-174	2	170	29,475	58,950
Total	**40**			**455,850**

$$SR = \frac{\sum_{i=1}^{7} f_i \lvert x_i - \bar{x} \rvert}{\sum_{i=1}^{7} f_i} = \frac{455,850}{40} = 11,396$$

3.2.3. Varians

Varians adalah rata-rata kuadrat selisih atau kuadrat simpangan dari semua nilai data terhadap rata.rata hitung. Varians untuk sampel dinotasikan sebagai S^2 sedangkan untuk populasi dinotasikan sebagai σ^2, yang didefinisikan sebagai :

$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

Perhatikan contoh berikut adalah data produksi sebuah mesin dengan merk tertentu :

20, 30, 50, 70, 80.

Untuk dapat menghitung nilai varians dari tersebut, terlebih dahulu lakukanperhitungan nilai rata-rata sebagai berikut :

$$\bar{x} = \frac{20+30+50+70+80}{5} = 50$$

Maka varians dari data di atas adalah :

$$\begin{aligned} S^2 &= \frac{\sum_{i=1}^{5} (x_i - \bar{x})^2}{n-1} \\ &= \frac{(20-50)^2 + (30-50)^2 + (50-50)^2 + (70-50)^2 + (80-50)^2}{5-1} \\ &= \frac{900+400+0+400+900}{4} \\ &= 650 \end{aligned}$$

Sementara, untuk data berkelompok, varians didefinisikan sebagai :

$$S^2 = \frac{\sum_{i=1}^{n} f_i \left(x_i - \bar{x}\right)^2}{\left(\sum_{i=1}^{n} f_i\right) - 1}$$

Gunakan data mengenai modal sebagai contoh perhitungan varians berikut :

Modal	Frekuensi (f_i)	Nilai Tengah (x_i)	$\left(x_i - \bar{x}\right)^2$	$f_i \left(x_i - \bar{x}\right)^2$
112-120	4	116	601,4756	2405,9024
121-129	5	125	241,0256	1205,1280
130-138	8	134	42,5756	340,6048
139-147	12	143	6,1256	73,5072
148-156	5	152	131,6756	658,3780
157-165	4	161	419,2256	1676,9024
166-174	2	170	868,7756	1737,5513
Total	**40**			**8097,5513**

$$S^2 = \frac{\sum_{i=1}^{7} f_i \left(x_i - \bar{x}\right)^2}{\left(\sum_{i=1}^{7} f_i\right) - 1} = \frac{8097,9741}{40-1} = 207,64$$

3.2.4. Standard Deviasi

Standard deviasi adalah akar pangkat dua dari varians. Standard deviasi disebut juga sebagai simpangan baku yang didefinisikan sebagai :

$$S = \sqrt{\frac{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2}{n-1}}$$

Menggunakan hasil perhitungan varians mengenai jumlah produksi mesin, maka diperoleh standard deviasi dari data tersebut yaitu :

$$S = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1}}$$
$$= \sqrt{\frac{\sum_{i=1}^{5}\left(x_i - \bar{x}\right)^2}{n-1}}$$
$$= \sqrt{\frac{(20-50)^2 + (30-50)^2 + (50-50)^2 + (70-50)^2 + (80-50)^2}{5-1}}$$
$$= \sqrt{\frac{900+400+0+400+900}{4}}$$
$$= \sqrt{650}$$
$$= 25,495$$

Cara perhitungan yang sama untuk data berkelompok, standard deviasi atau simpangan baku didefinisikan sebagai :

$$S = \sqrt{\frac{\sum_{i=1}^{n} f_i\left(x_i - \bar{x}\right)^2}{\left(\sum_{i=1}^{n} f_i\right) - 1}}$$

3.2.5. Koefisien Variasi

Dispersi relatif digunakan untuk membandingkan variablitas nilai observasi pada dua atau lebih kelompok data dengan memperhatikan besar kecilnya nilai observasi pada kelompok data tersebut.

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

Keterangan :

KV : Koefisien variasi

$\bar{x}$: rata-rata

S : standar deviasi

Perhatikan kasus berikut, terdapat dua buah bola lampu. Lampu jenis A secara rata-rata mampu menyala selama 1.500 jam dengan simpangan baku S_1=275 jam. Sedangkan lampu jenis B secara rata-rata mampu menyala selama 1.750 jam dengan simpangan baku S_2=300 jam. Lampu manakah yang memiliki kualitas yang lebih baik?

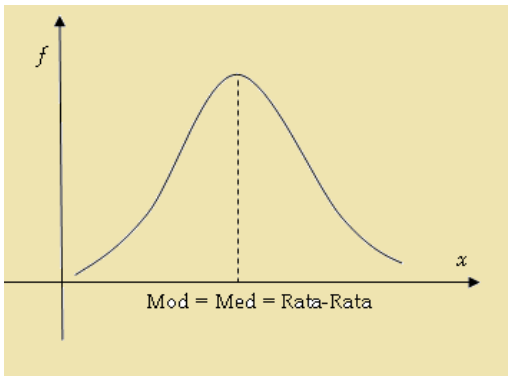
Lampu Jenis A, $KV_A = \frac{S_1}{\bar{x}_1} \times 100\% = \frac{275}{1.500} \times 100\% = 18,3\%$

Lampu Jenis B, $KV_B = \frac{S_2}{\bar{x}_2} \times 100\% = \frac{300}{1.750} \times 100\% = 17,1\%$

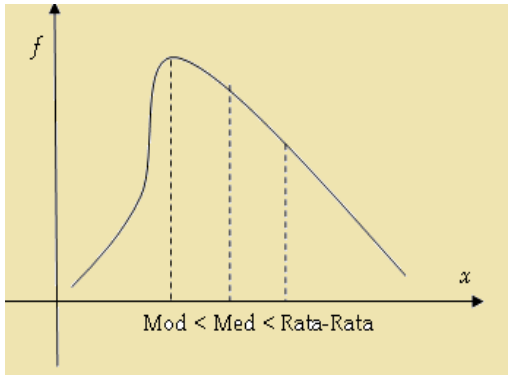
Berdasarkan hasil perhitungan nilai koefisien variasi pada masing-masing jenis lampu, lampu jenis B jelas memiliki nilai KV yang lebih kecil dibandingkan dengan KV lampu jenis A. Dengan demikian lampu jenis B memiliki kualitas yang lebih seragam sedangkan lampu jenis A memiliki kualitas yang bervariasi. Oleh karena itu, dapat disimpulkan bahwa lampu jenis A memiliki kualitas yang lebih baik dibandingkan dengan lampu jenis B.

3.3. Kemiringan Distribusi Data

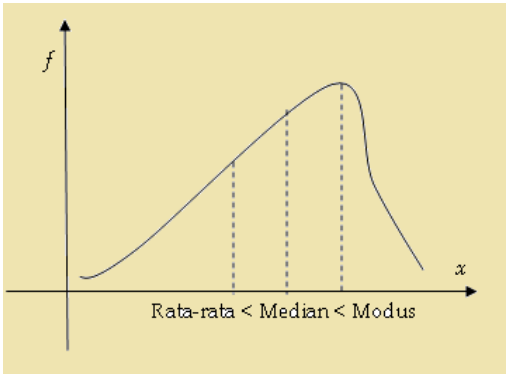
Kemiringan adalah derajat atau uuran dari ketidaksimetrian (asimetri) suatu distribusi data. Tiga kemiringan distribusi data yaitu simetris, miring ke kanan dan miring ke kiri.



(a). Distribusi Simetri



(b). Distribusi Miring ke Kanan



(b). Distribusi Miring ke Kiri

Terdapat beberapa cara yang dipakai untuk menghitung derajat kemiringan distribusi data, yaitu :

1. Karl Pearson

   Menurut Pearson, derajat kemiringan (α) didefinisikan sebagai :

$$\alpha = \frac{(\bar{x} - Mo)}{S} \quad \text{atau} \quad \alpha = \frac{3(\bar{x} - Med)}{S}$$

   Bila α=0 atau mendekati nol maka dikatakan distribusi data simetri, bila α < 0 atau bertanda negatif maka distribusi data dikatakan miring ke kiri sedangkan bila α > 0 atau bertanda positif maka distribusi data dikatakan miring ke kanan.

2. Momen

   Cara lain yang dapat digunakan untuk menghitung derajat kemiringan distribusi data adalah rumus momen yang didefinisikan sebagai berikut :

$$\alpha_3 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^3}{nS^3}$$

   Untuk data berkelompok, rumus momen yang digunakan yaitu :

$$\alpha_3 = \frac{\sum_{i=1}^{n} f_i (x_i - \bar{x})^3}{\left(\sum_{i=1}^{n} f_i\right) S^3}$$

   Bila α_3=0 atau mendekati nol maka dikatakan distribusi data simetri, bila α_3 < 0 atau bertanda negatif maka distribusi data dikatakan miring ke kiri sedangkan bila α_3> 0 atau bertanda positif maka distribusi data dikatakan miring ke kanan.

3. Rumus Bowley

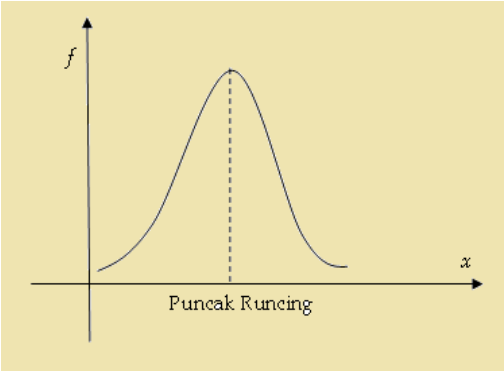
Cara ketiga yang dapat digunakan untuk menghitung derajat kemiringan distribusi data adalah rumus Bowley yang didefinisikan sebagai berikut :

$$\alpha = \frac{Q_3 + Q_1 - Q_2}{Q_3 - Q_1}$$

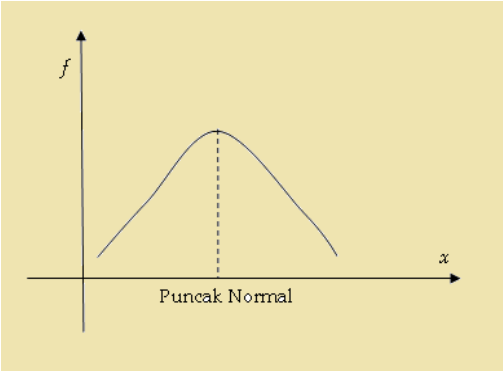
Jika distribusinya simetri, Q_3-Q_2=Q_2-Q_1 maka α=0. Sementara jika distribusinya miring, maka ada dua kemungkinan yaitu Q_1 = Q_2 sehingga α=1 dan Q_2=Q_3 sehingga α=-1

3.4. Keruncingan Distribusi Data

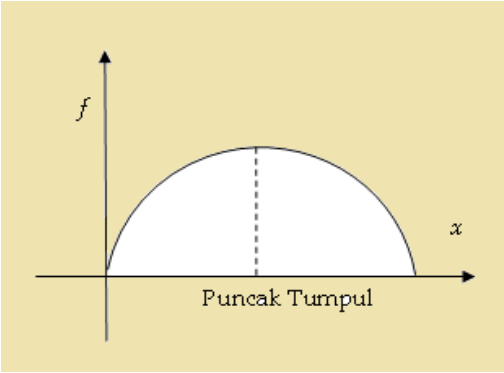
Keruncingan distribusi data (kurtosis) adalah derajat ukuran tinggi rendahnya puncak suatu distribusi data terhadap distribusi normalnya data (DR. Boediono, 2014). Terdapat tiga keruncingan distribusi data yaitu leptokurtis, mesokurtis dan platikurtis. Leptokurtis artinya distribusi data memiliki puncak yang relatif tinggi atau runcing, perhatikan gambar berikut :



Mesokurtis artinya distribusi data memiliki puncak yang normal, tidak terlalu tinggi (runcing) maupun tidak terlalu rendah. Perhatikan gambar berikut :



Sedangkan, platikurtis artinya distribusi data memiliki puncak yang terlalu rendah atau terlalu mendatar dan dapat digambarkan sebagai berikut :



Untuk data yang tidak berkelompok, derajat keruncingan distribusi (α_4) dihitung melalui persamaan berikut :

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

Sementara, untuk data yang berkelompok, α_4 dihitung melalui persamaan berikut :

$$\alpha_4 = \frac{\sum_{i=1}^{n} f_i (x_i - \bar{x})^4}{\left(\sum_{i=1}^{n} f_i\right) S^4}$$

Jika $\alpha_4=3$, maka keruncingan distribusi data disebut mesokurtis, Jika $\alpha_4>3$, maka keruncingan distribusi data disebut leptokurtis dan jika $\alpha_4<3$ maka keruncingan distribusi data disebut sebagai platikurtis.

CASE STUDY

Kumpulkan data mengenai rata-rata lamanya (dalam satuan jam) mahasiswa di STMIK Sumedang mengakses Internet melalui fasilitas *Wifi* kampus. Kemudian sajikan data yang telah dikumpulkan melalui sebuah tabel atau grafik dan hitunglah :

a) Rata-Rata
b) Median
c) Modus
d) Kuartil 1 (Q_1) , Kuartil 2 (Q_2) dan Kuartil 3 (Q_3)
e) Persentil 30
f) *Range*
g) Simpangan Baku (Standard Deviasi)
h) Koefisien Variasi
i) Koefisien Kemiringan dan tentukan kecenderungan kemiringan dari distribusi data yang diperoleh
j) Koefisien Keruncingan dan tentukan kecenderungan keruncingan puncak dari distribusi data yang diperoleh.

Teori dan Distribusi Probabilitas

4.1. Teori Probabilitas

4.1.1. Definisi Probabilitas

The probability of an event is the proportion (relative prequency) of times that the event is expected to occur when an experiment is repeated a large number of times under identical condition (Rudolf, J.Freund, 2003). Menurut Mendenhall dan Reinmunt (1982) dalam Supranto (2008), *probability is a measure of a likelihood of the occurance of a random event.* Kemudian menurut DR. Boediono (2014), derajat atau tingkat kepastian atau keyakinan dari munculnya hasil percobaan statistik disebut probabilitas atau peluang. Berdasarkan ketiga definisi tersebut, probabilitas atau peluang pada dasarnya memberikan sebuah nilai atau besaran pada sebuah kejadian yang masih dalam ruang lingkup pembicaraan. Dengan kata lain probabilitas atau peluang adalah sebuah kemungkinan terjadinya suatu *event* yang dinyatakan dalam sebuah nilai atau besaran.

Perumusan probabilitas terbagi menjadi tiga, yaitu perumusan klasik, perumusan empiris dan perumusan subjektif.

- Perumusan klasik

  Dalam perumusan klasik, probabilitas sebuah peristiwa dinyatakan sebagai hasil bagi antara jumlah persitiwa yang mungkin terjadi dengan jumlah semua peristiwa yang mungkin terjadi. Jika peristiwa tersebut dimisalkan sebagai peristiwa *A*, terjadi dalam *m* cara dari seluruh *n* cara yang mungkin terjadi dan masing-masing *n* cara tersebut memiliki kesempatan yang sama untuk muncul, maka probabilitasnya dapat dinyatakan sebagai :

$$P(A) = \frac{m}{n}$$

- Perumusan empiris

  Dalam perumusan empiris, perhitungan probabilitas dilakukan berdasarkan frekuensi relatif dari terjadinya suatu kejadian dengan syarat banyaknya pengamatan atau banyaknya sampel *n* sangat besar. Bila *n* bertambah besar sampai tak terhingga, maka probabilitas dari kejadian *A* sama dengan nilai limit dari frekuensi relatif kejadian *A* tersebut (DR. Boediono, 2014). Jika kejadian *A* terjadi sebanyak *m* kali dari keseluruhan pengamatan sebanyak n , dimana *n* sangat besar atau mendekati tak hingga, maka probabilitas kejadian *A* didefinisikan sebagai :

$$P(A) = \lim_{n \to \infty} \frac{m}{n}$$

- Perumusan subjektif

  Dalam perumusan subjektif, probabilitas dirumuskan berdasarkan keyakinan dan pandangan pribadi terhadap probabilitas terjadinya suatu peristiwa. Sebagai contoh, seorang guru meyakini bahwa tingkat kelulusan siswa SMA dalam Ujian Nasional tahun 2015 adalah 100%.

4.1.2. Probabilitas Peristiwa Sederhana

Probabilitas peristiwa sederhana :

$$P(a) = \frac{\sum Peristiwa}{RuangSampel}$$

Probabilitas peristiwa sederhana

- Nilai probabilitasnya harus berkisar antara $0 \leq p \leq 1$
- Jumlah semua nilai probabilitas sama dengan 1

Dalam menghitung probabilitas, terdapat beberapa langkah yang dapat dilakukan :

1. Tentukan percobaan yang dilakukan
2. Tentukan ruang sampel *S*
3. Tentukan peluang tiap titik sampel
4. Tentukkan peristiwa *A* yang menjadi perhatian
5. Hitung probabilitas peristiwa *A*.

Sebagai contoh, jika peristiwa *A* adalah munculnya mata dadu 1 atau 2 atau 3 pada pelemparan sebuah mata dadu, maka hitunglah probabilits *A*.

Percobaan yang dilakukan adalah pelemparan mata dadu

Rungan sampel *S* : 1,2,3,4,5,6

Titik sampel *A* = 1,2,3

$$P(A) = P(1) + P(2) + P(3)$$
$$= \frac{1}{6} + \frac{1}{6} + \frac{1}{6}$$
$$= \frac{1}{2}$$
$$P(A) = \frac{\sum Peristiwa(A)}{RuangSampel} = \frac{1+1+1}{6} = \frac{1}{2}$$

Contoh lain, pada sebuah kotak terdapat 10 buah bola pingpong, 3 berwarna putih, 2 hitam dan sisanya merah. Jika kita mengambil sebuah bola secara random atau acak, maka probabilitas menemukan bola berwarna merah dapat dihitung sebagai berikut :

A : peristiwa menemukan bola merah

$$P(A) = \frac{\sum Peristiwa(A)}{RuangSampel} = \frac{5}{10}$$

4.1.3. Probabilitas Peristiwa *Mutually-Exclusif*

If two events cannot occur simultaneously, that is, one "excludes" the other, then two events are said to be mutually exclusif (Rudolf J. Freund, 2003). Dengan kata lain, dua peristiwa dikatakan saling eksklusif jika terjadinya peristiwa yang satu menyebabkan tidak terjadinya peristiwa yang lain (Sudjana, 2004). Bila A dan B dua kejadian sebarang pada S dan berlaku $A \cap B = \emptyset$ maka A dan B dikatakan dua kejadian saling lepas atau saling bertentangan atau saling terpisah (DR.Boediono, 2014). Jika peristiwa A dan B adalah dua peristiwa *mutually exclusif* maka besar probabilitas peristiwa A atau B adalah :

$$P(A \cup B) = P(A) + P(B)$$

Misalkan, pada pelemparan dadu satu kali, jika A adalah peristiwa munculnya mata dadu 3 dan B adalah peristiwa munculnya mata dadu 5, berapakah probabilitas munculnya mata dadu 3 **atau** 5 :

$$P(A \cup B) = P(A) + P(B)$$
$$= 1/6 + 1/6$$
$$= 1/3$$

Pada pelemparan dua buah dadu, tentukanlah probabilitas munculnya muka dua dadu dengan jumlah 7 atau 11? Ruang Sampel dari pelemparan dua buah dadu tersebut dapat dilihat pada tabel berikut :

Mata Dadu	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Dengan *A* adalah kejadian munculnya jumlah 7 dan *B* adalah kejadian munculnya jumlah 11.

A={(6,1),(5,2),(4,3),(3,4),(2,5),(5,2)}

B={(6,5),(5,6)}

$$P(A \cup B) = P(A) + P(B) = 4/36 + 2/36 = 1/6$$

4.1.4. Probabilitas Peristiwa non *Mutually-Exclusif*

Dua peristiwa dikatakan non-*mutually exclusif* apabila keduanya dapat terjadi pada waktu yang bersamaan, atau A∩B ≠ Ø, sehingga probabilitas terjadinya peristiwa *A* atau peristiwa *B* didefinisikan sebagai :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Misalkan, dari sebuah survey terhadap 100 responden, diketahui bahwa 60 responden suka film *action*, 50 suka drama dan 10 suka keduanya. Jika dari 100 responden tersebut diambil secara acak, berapakah probabilitas menemukan responden yang suka film *action* atau responden yang suka drama?

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{60}{100} + \frac{50}{100} - \frac{10}{100} = 1$$

4.1.5. Probabilitas Peristiwa Independen

Two events A and B are said to be independent if the probability of A occuring is in no way affected by event B having occured or vice versa (Rudolf. J.Freund, 2003) atau dua peristiwa dikatakan bebas atau independen atau bebas statistis jika terjadinya atau tidak terjadinya peristiwa yang satu tidak mempengaruhi atau dipengaruhi oleh yang lainnya

(Sudjana, 2004). Dua kejadian *A* dan *B* dalam ruang sampel *S* dikatakan saling bebas jika kejadian *A* tidak mempengaruhi kejadian *B* dan sebaliknya kejadian *B* tidak mempengaruhi kejadian *A* (DR.Boediono, 2014). Jika peristiwa *A* dan *B* adalah dua peristiwa saling bebas maka probabilitas peristiwa *A* dan *B* didefinisikan sebagai :

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

4.1.6. Probabilitas Peristiwa Dependen

Dua peristiwa yang terjadi berurutan dapat dikatakan dependen jika peristiwa pertama mempengaruhi peristiwa kedua sehingga besar probabilitas peristiwa *A* dan peristiwa *B* adalah :

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

Probabilitas terjadinya kejadian *A* bila kejadian *B* telah terjadi disebut probabilitas bersyarat yang ditulis P(*A*|*B*) dan dirumuskan sebagai berikut :

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

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

Misalkan, di sebuah kotak terdapat 3 buah bola putih dan 7 bola hitam. Jika diambil 2 bola satu per satu tanpa disertai pengembalian, maka berapa probabilitas bola pertama yang terambil adalah putih dan bola kedua adalah hitam?

$$P(A \cap B) = P(A) \times P(B|A) = \frac{3}{10} \times \frac{7}{(10-1)} = \frac{21}{90} = \frac{7}{30}$$

Misalkan diberikan populasi sarjana di suatu kota dibagi menurut jenis kelamin dan status pekerjaan sebagai berikut :

Jenis Kelamin	Bekerja	Menganggur	Jumlah
Laki-Laki	460	40	500
Wanita	140	260	400
Jumlah	600	300	900

Dari data tersebut, kemudian akan diambil seorang dari mereka untuk ditugaskan melakukan promosi barang di suatu kota. Bila ternyata yang dipilih adalah dalam status telah bekerja, berapakah probabilitas bahwa dia (a) laki-laki dan (b) wanita.

A : kejadian terpilihnya sarjana telah bekerja
B : kejadian bahwa dia laki-laki
C : kejadian bahwa dia perempuan
$n(A \cap B) = 460$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{460/900}{600/900} = \frac{460}{600} = \frac{23}{30}$$

$$P(C|A) = \frac{P(A \cap C)}{P(A)} = \frac{140/900}{600/900} = \frac{140}{600} = \frac{7}{30}$$

4.1.7. Ruang Sampel

Seperti yang sudah diuraikan sebelumnya, dalam menghitung probabilitas sebuah peristiwa kita harus dapat menghitung ruang sampel. Ruang sampel S dapat dihitung melalui tiga cara yaitu aturan perkalian, permutasi dan kombinasi.

Dalam menghitung ruang sampel S dengan menggunakan aturan perkalian, jika kita memiliki m unsur $a_1, a_2, ..., a_m$ dan n unsur $b_1, b_2, ..., b_n$ yang mengandung unsur dari setiap kelompok akan membentuk $m.n$ pasang.

Perhatikan percobaan yang terdiri atas pencatatan hari lahir untuk masing-masing dari 20 orang yang dipilih secara acak. Jika kita mengabaikan tahun kabisat dan andaikan ada hanya 365 hari lahir yang mungkin berbeda, carilah banyaknya titik dalam ruang sampel terhadap percobaan ini.Jika kita andaikan bahwa setiap himpunana hari lahir yang mungkin berpeluang sama, berapakah peluang bahwa setiap orang mempunyai hari lahir yang berbeda?

Banyaknya titik sampel :

$$n(A) = \underset{\text{orang pertama}}{(365)} \quad \underset{\text{orang ke-dua}}{(364)} \quad \underset{\text{orang ke=tiga}}{(363)} \cdots \underset{\text{orang ke-dua puluh}}{(346)}$$

Peluang bahwa setiap orang mempunyai hari lahir yang berbeda :

$$P(A) = \frac{n(A)}{N} = \frac{(365)(364)\cdots(346)}{(365)^{20}} = 0,5886$$

Selanjutnya, dalam menghitung ruang sampel menggunakan permutasi artinya menghitung jumlah cara menyusun sebuah objek dengan memperhatikan urutannya. Cara menyusun *r* objek dari *n* objek didefinisikan sebagai berikut :

$$P_r^n = \frac{n!}{(n-r)!}$$

Andaikan operasi merakit dalam perencanaan pabrik mengandung empat tahap, yang dibentuk dalam suatu barisan. Jika pemilik pabrik ingin mencoba untuk membndingkan waktu merakit untuk masing-masing barisan, berapa banyak barisan yang berbeda akan dilibatkan dalam percobaan?

$$P_4^4 = \frac{4!}{(4-4)!} = 24$$

Cara menghitung ruang sampel dengan menggunakan kombinasi, artinya menghitung banyaknya cara menyusun suatu objek tanpa memperhatikan urutannya. Cara menyusun *r* objek dari *n* objek tanpa memperhatikan urutannya didefinisikan sebagai :

$$C_r^n = \frac{n!}{r!(n-r)!}$$

Misalkan, 10 orang hadir dalam sebuah pesta. Jika mereka berjabat tangan satu per satu, maka berapa kali jabat tangan yang terjadi dalam pesta tersebut.

$$C_2^{10} = \frac{10!}{2!(10-2)!} = 45$$

4.1.8. Rumus Bayes

Secara umum, bila A_1, A_2, ...,A_n kejadian saling lepas dalam ruang sampel *S* dan *B* kejadian lain yang sembarang dalam *S*, maka probabilitas kejadian bersyarat $A_i|B$ dirumuskan sebagai berikut :

$$P(A_i|B) = \frac{P(B \cap A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\underbrace{\sum_{i=1}^{n} P(B|A_i)P(A_i)}_{\text{probabilitas marginal kejadian B}}}$$

Misalkan, Sebuah perusahaan memiliki 3 mesin *A*, *B* dan *C* yang masing-masing memproduksi 20%, 30% dan 50% dari total produksi. Dari pengalaman diperoleh informasi bahwa :

dari mesin *A* terdapat 5% produk yang rusak

dari mesin *B* terdapat 4% produk yang rusak

dari mesin *C* terdapat 3% produk yang rusak

Bila suatu produk yang diambil secara acak dan ternyata rusak. Berapa peluang bahwa produk tersebut berasal dari mesin A?

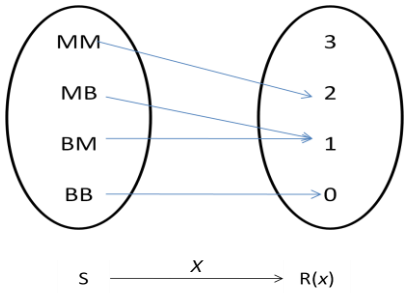
$$P(A)=0,2 \qquad P(B)=0,3 \qquad P(C)=0,5$$
$$P(R|A)=0,05 \qquad P(R|B)=0,04 \qquad P(R|C)=0,03$$
$$P(A|R)=\frac{P(R\cap A)}{P(R)}=\frac{P(R|A)P(A)}{P(R|A)P(A)+P(R|B)P(B)+P(R|C)P(C)}$$
$$=\frac{0,05(0,2)}{0,05(0,2)+0,04(0,3)+0,03(0,5)}$$
$$=0,27027$$

4.2. Distribusi Probabilitas

4.2.1. Definisi Variabel Acak

A random variabel is a rule that assigns a numerical value to an outcome of interest (Rudolf. J.Freund,2003) atau variabel acak diartikan sebagai deskripsi numerik dari hasil percobaan, biasanya menghubungkan nilai-nilai numerik dengan setiap kemungkinan hasil percobaan (J. Supranto, 2009).

Misalkan, pada sebuah pengundian atau pelemparan dua mata uang logam, distribusi kemungkinan untuk muncul muka dapat digambarkan melalui diagram Venn berikut :



Nilai *X* ditentukan terhadap titik sampel yang dalam hal ini tergantung kepada jumlah muka yang diakibatkan oleh setiap titik :

X(MM)=2

X(MB)=1

X(BM)=1

X(BB)=0

X dikatakan sebagai variabel acak, yang menyatakan jumlah muka yang muncul dari pelemparan 2 mata uang logam yang dapat mengambil 3 nilai *x*=0,1,2. Maka distribusi kemungkinannya adalah :

x	0	1	2
P(*x*)	1/4	2/4	1/4

Variabel acak terbagi menjadi dua, yaitu variabel acak diskrit dan vriabel acak kontinu. *A discrete random variabel is one that can take on only a countable number of values* (Rudolf, J.Freund, 2003) atau variabel yang hanya memiliki nilai 0,1,2,3,... dimana untuk tiap nilai variabel terdapat peluangnya (Sudjana, 2005). Variabel acak diskrit hanya dapat mengambil nilai-nilai tertentu yang terpisah, yang umumnya dihasilkan dari perhitungan suatu objek (J.Supranto, 2009). Berikut adalah beberapa contoh variabael acak diskrit :

Percobaan	Variabel Acak	Kemungkinan Nilai-nilai Variabel Acak
Penelitian terhadap 50 produk baru	Banyaknya produk yang rusak	0,1,2,3,…,50
Pencatatan pengunjung restoran pada suatu hari	Banyaknya pengunjung	0,1,2,3,

Sementara, *a continuous random variable is one that can take on any value in an interval* (Rudolf, J.Freund, 2003) atau merupakan variabel acak yang tidak diskrit, artinya jika *X* adalah variabel kontinu maka nilai *X* berada dalam interval $-\infty<x<\infty$ atau batas-batas lain (Sudjana, 2005). Varaibel acak kontinu adalah variabel acak yang nilai-nilainya dihasilkan dari hasil pengukuran pada suatu objek (J.Supranto, 2009). Berikut adalah beberapa contoh variabel acak kontinu :

Percobaan	Variabel Acak	Kemungkinan nilai-nilai Variabel Acak
Membangun proyek perkantoran baru setelah 6 bulan	Persentase proyek yang diselesaikan	$0 \leq x \leq 100$
Isi Botol Minuman jadi (maks = 600 ml)	Jumlah mililiter	$0 \leq x \leq 600$
Penimbangan 20 paket kemasan (mks = 2 Kg)	Berat sebuah paket kemasan (kg)	$0 \leq x \leq 2$

4.2.2. Distribusi Probabilitas Variabel Acak Diskrit

A probability distribution is a definition of the probabilties of the values of a random variable (Rudolf, J.Freund,2003). Distribusi probabilitas variabel acak menggambarkan bagaimana suatu probabilitas didistribusikan terhadap nilai-nilai dari variable acak tersebut (J.Supranto, 2009).

Sama seperti variabel acak, distribusi probabilitas dikelompokkan menjadi dua yaitu distribusi probabilitas variabel acak diskrit yang dinotasikan sebagai $p(x)=P(X=x)$ dan distribusi probabilitas variabel acak kontinu yang dinotasikan sebagai $f(x)$.

Syarat yang harus dipenuhi pada sebuah distribusi probabilitas variabel acak diskrit :

1. $p(x)\geq 0$
2. $0\leq p(x)\leq 1$
3. $\sum p(x)=1$

Fungsi probabilitas kumulatif variabel acak diskrit dituliskan sebagai berikut :

$$F(x) = P(X\leq x)$$

Nilai harapan (Ekspektasi/Rata-rata) :

$$\mathrm{E}(X) = \sum x\ \mathrm{p}(x)$$

Nilai varians :

$$\sigma^2 = E\left[X-\mu\right]^2$$
$$= \sum_{i=1}^{n}\left(x_i-\mu\right)^2 p\left(x_i\right)$$

Data berikut merupakan hasil penjualan mobil selama 300 hari pada PT. Indah Caraka Motor, Jakarta. Data yang dicatat adalah jumlah mobil yang terjual dalam sehari.

Jumlah mobil yang terjual dalam sehari	Jumlah Hari
0	54
1	117
2	72
3	42
4	12
5	3
Total	**300**

Probabilitas untuk masing-masing jumlah mobil yang terjual dalam sehari yaitu :

x	Frekuensi	$p(x)=\frac{frekuensi}{Total}$	$x * p(x)$
0	54	0,18	0
1	117	0,39	0,39
2	72	0,24	0,48
3	42	0,14	0,42
4	12	0,04	0.,6
5	3	0,01	0,05
Total	**300**	**1,00**	**1,5**

Berdasarkan tabel, dapat diketahui bahwa probabilitas tidak satu pun mobil terjual dalam sehari adalah 0,18 atau kemungkinannya adalah 18%, sedangkan probabilitas terjualnya 1 unit mobil dalam sehari adalah 0,39 atau kemungkinannya 39%, probabilitas terjualnya 2 unit mobil dalam sehari adalh 0,24 atau kemungkinannya 24%, probabilitas terjualnya 3 unit mobil dalam sehari adalah 0,14 atau kemungkinannya 14%, probabilitas terjualnya 4 unit mobil dalam sehari adalah 0,04 atau kemungkinannya adalah 4% dan probabilitas terjualnya 5 unit mobil dalam sehari adalah 0,01 atau kemungkinannya 1%. Dari tabel yang sama, kita juga dapat mengetahui ekspektasi (harapan) rata-rata terjualnya mobil dalam 300 hari yaitu 1-2 unit mobil.

Perhatikan contoh berikut, seorang mandor pabrik merencanakan mempunyai tiga pekerja pria dan tiga pekerja wanita untuknya. Ia ingin memilih dua pekerja untuk pekerjaan khusus. Dia tidak menginginkan kesalahan dalam pemilihannya. Ia memutuskan untuk memeilih dua pekerja secara acak. Misalkan, X menyatakan banyaknya wanita dalam pemilihannya, carilah distribusi peluang untuk X.

Untuk menyelesaikan persoalan tersebut, pertama kita harus menentukan ruang sampel S yaitu banyaknya cara memilih dua pekerja dari 6 pekerja. Oleh karena urutan pemilihan pekerja tidak diperhatikan maka perhitungan banyaknya cara memilih dua pekerja dari 6 pekerja dilakukan dengan menggunakan aturan kombinasi sebagai berikut :

$$C_2^6=\frac{6!}{2!4!}=15$$

Selanjutnya, tentukan kemungkinan wanita terpilih dalam pemilihan dua pekerja adalah sebagai berikut :

Wanita	Pria
0	2
1	1
2	0

Dengan *X* menyatakan banyaknya wanita dalam pemilihan, maka probabilitas masing-masing kemungkinan tersebut dapat dihitung sebagai berikut :

- Probabilitas tidak satu pun wanita terpilih atau keduanya pria

$$p(0) = P(X = 0) = \frac{C_0^3 C_2^3}{C_2^6} = \frac{3}{15} = \frac{1}{5}$$

- Probabilitas satu wanita terpilih dan 1 pria terpilih

$$p(1) = P(X = 1) = \frac{C_1^3 C_1^3}{C_2^6} = \frac{9}{15} = \frac{3}{5}$$

- Probabilitas keduanya wanita atau tidak satupun pria terpilih

$$p(2) = P(X = 2) = \frac{C_2^3 C_0^3}{C_2^6} = \frac{3}{15} = \frac{1}{5}$$

Dari probabilitas-probabilitas tersebut, kita dapat dengan jelas melihat pola distribusi peluangnya yang dapat didefinisikan sebagai :

$$p(x) = \frac{C_x^3 C_{2-x}^3}{C_2^6}$$

Untuk x=0,1,2

4.2.2.1. Distribusi Binomial

Distribusi binomial diterapkan pada percobaan atau fenomena tau kejadian dengan ciri-ciri :

- Percobaan atau fenomena atau kejadiaan terdiri atas *n* tindakan (*trial*) yang identik.
- Setiap percobaan atau fenomena atau kejadian hanya mempunyai dua kemungkinan hasil, misalkan “Sukses” dan “Gagal”.

- Peluang sukses pada percobaan tunggal sama dengan p dan tetap sama dari percobaan ke percobaan, peluang gagal sama dengan $(1-p)=q$
- Setiap percobaan bersifat independen dan dengan pengembalian.
- Variabel acak X adalah banyaknya sukses percobaan selama n tindakan.

Probabilitas sebuah variabel acak X yang berdistribusi binomial didefinisikan sebagai berikut :

$$p(x) = C_x^n p^x (1-p)^{n-x}$$

Untuk $x = 0,1,2,3,\ldots,n$

$p(x)$: probabilitas peristiwa sukses terjadi sebanyak x kali

C : kombinasi x dari n

n : banyaknya percobaan

p : probabilitas sukses

$(1-p)$: probabilitas gagal

Rata-rata nya didefinisikan sebagai :

$$\mu = n \times p$$

Standard deviasi nya didefinisikan sebagai :

$$\sigma = \sqrt{n \times p \times (1-p)}$$

Perhatikan contoh berikut, probabilitas seorang penembak akan menembak tepat pada sasaran adalah 0m7. Jika penembak tersebut diberi kesempatan untuk menembak 10 kali, berapakah probabilitas 5 tembakan tepat sasaran.

$$\begin{aligned} p(5) &= C_5^{10} (0,7)^5 (1-0,7)^{10-5} \\ &= \frac{10!}{5!5!}(0,16807)(0,00243) \\ &= \frac{10 \times 9 \times 8 \times 7 \times 6}{5 \times 4 \times 3 \times 2 \times 1}(0,000408) \\ &= 0,102919 \end{aligned}$$

Rata-rata banyaknya tembakan tepat sasaran *np*=10. 0,7=7 kali

4.2.2.2. Distribusi Poisson

Distribusi poisson diterapkan pada percobaan atau fenomena atau kejadian dengan ciri-ciri :

- Banyaknya percobaan yang dilakukan (*n*) relatif besar
- Kejadian yang diteliti merupakan kejadian yang jarang terjadi sehingga probabilitasnya kecil
- Diketahui parameter intensitas percobaan (rata-rata kejadian)

Probabilitas sebuah variabel acak *X* yang berdistribusi poisson didefinisikan sebagai berikut :

$$P(x) = \frac{\mu^x e^{-\mu}}{x!}$$

dengan :

μ : rata-rata kejadian (intensitas kejadian)
x : jumlah sukses
e : bilangan alam (epsilon) = 2,7182

Misalkan, berdasarkan pengalaman, setiap kali mencetak 10.000 lembar kertas terdapat 100 lembar yang rusak. Pada suatu waktu, perusahaan mencetak 1.000 lembar kertas. Berapakah probabilitas mendapatkan tepat 5 lembar kertas yang rusak? Dan berapakah probabilitas mendapatkan paling banyak 2 lembar kertas yang rusak?
Berdasarkan pengalaman, maka probabilitas rusaknya lembar kertas yang dicetak adalah:

$$p = \frac{100}{10.000} = 0,01$$

Dengan *n*=1.000 dan rata-rata $\mu = n \times p = 1000 \times 0,01 = 10$ maka probabilitas mendapatkan tepat 5 lembar kertas yang rusak adalah :

$$P(5) = \frac{10^5 e^{-10}}{5!} = 0,03783$$

Sedangkan probabilitas mendapatkan paling banyak 2 lembar kertas yang rusak adalah :

$$P(x \le 2) = P(0) + (1) + (2)$$
$$= \frac{10^0 e^{-10}}{0!} + \frac{10^1 e^{-10}}{1!} + \frac{10^2 e^{-10}}{2!}$$
$$= 0,002768$$

4.2.2.3. Distribusi Hipergeometrik

Misalkan kita mempunyai suatu populasi sebanyak N yang terdiri atas dua jenis. Jenis merah sebanyak N_1 dan sisanya jenis putih sebanyak N-N_1. Kemudian diambil sampel secara acak sebanyak n buah tanpa pengembalian. Misalkan $X=k$ menyatakan banyaknya jenis merah yang terambil, maka probabilitas untuk memperoleh sampel jenis merah sebanyak $X=k$ adalah :

$$P(X = k) = \frac{\binom{N_1}{k}\binom{N - N_1}{n-k}}{\binom{N}{n}}$$

Rata-Rata :

$$\mu = \frac{nk}{N}$$

Varians :

$$\sigma^2 = \left(\frac{N-n}{N-1}\right)(n)\left(\frac{k}{n}\right)\left(1-\frac{k}{n}\right)$$

Simpangan Baku/Standard Deviasi :

$$\sigma = \sqrt{\left(\frac{N-n}{N-1}\right)(n)\left(\frac{k}{n}\right)\left(1-\frac{k}{n}\right)}$$

Misalkan dalam suatu rak terdapat 50 kain batik yang diantaranya 5 rusak. Bila diambil kain sebanyak 4 helai secara acak, hitunglah probabilitas untuk memperoleh 0,1,2,3,4 helai kain yang rusak?

Untuk memudahkan perhitungan probbilitas pada persoalan di atas, maka terlebih dahulu definisikan setiap informasi yang terdapat dalam persoalan sebagai berikut :

N = banyaknya kain =50

N_1 = banyaknya kain rusak = 5

N_2 = banyaknya kain baik = 45

n = sampel kain yang diambil =4

$X=k$ = menyatakan banyaknya kain rusak yang diperoleh, maka k=0,1,2,3,4

$$P(X=0)=\frac{\binom{5}{0}\binom{45}{4}}{\binom{50}{4}}=0,64696 \qquad P(X=1)=\frac{\binom{5}{1}\binom{45}{3}}{\binom{50}{4}}=0,30808$$

$$P(X=2)=\frac{\binom{5}{2}\binom{45}{2}}{\binom{50}{4}}=0,04299 \qquad P(X=3)=\frac{\binom{5}{3}\binom{45}{1}}{\binom{50}{4}}=0,00195$$

$$P(X=4)=\frac{\binom{5}{4}\binom{45}{0}}{\binom{50}{4}}=0,00002$$

Perhatikan contoh kedua berikut, suatu masalah penting dihadapi oleh direktur personalia antara lain pemilihan terbaik dalam kelompok unsur berhingga ditunjukkan oleh situasi berikut. Dari satu kelompok atas 20 sarjana teknik, 10 dipilih untuk dipekerjakan. Berapakah peluang bahwa 10 orang yang terpilih termasuk semua di dalamnya 5 sarjana teknik terbaik dalam kelompok atas 20?

Sama seperti pada persoalan sebelumnya, definisikan setiap informasi yang terdapat dalam soal sebagai berikut :

N = Kelompok atas Sarjana Teknik =20

N_1 = Sarjana Teknik Terbaik= 5

n = Sarjana Teknik yang dipekerjakan =10

$X=k$ = menyatakan banyaknya sarjana teknik terbaik yang terpilih maka k=5

Dengan demikian, peluang bahwa 10 orang yang terpilih termasuk semua di dalamnya 5 sarjana teknik terbaik dalam kelompok atas 20 adalah :

$$P(X=5)=\frac{\binom{5}{5}\binom{15}{5}}{\binom{20}{10}}=0,0162$$

Pada kenyataanya, dalam sebuah populasi tidak hanya terdapat satu atau dua jenis, bisa lebih dari dua. Misalkan, X_1 adalah banyaknya jenis 1, X_2 adalah banyaknya

jenis 2 dan X_3 adalah banyaknya jenis 3. Maka probabilitas terpilihanya $X_1=k_1$, $X_2=k_2$, dan $X_3=k_3$ didefinisikan sebagai :

$$P(X=k)=\frac{\binom{N_1}{k_1}\binom{N_2}{k_2}\binom{N_3}{k_3}}{\binom{N_1+N_2+N_3}{k_1+k_2+k_3}}$$

Misalkan, dalam sebuah kotak terdapat 5 kelereng merah, 10 kelereng putih dan 12 kelereng biru. Bila diambil 3 kelereng, probabilitas untuk memperoleh tiga jenis kelereng tersebut yang terdiri tas 1 kelereng merah, 1 kelereng putih dan 1 kelereng biru adalah

$$P(X_1=1,X_2=1,X_3=1)=\frac{\binom{N_1}{k_1}\binom{N_2}{k_2}\binom{N_3}{k_3}}{\binom{N_1+N_2+N_3}{k_1+k_2+k_3}}=\frac{\binom{5}{1}\binom{10}{1}\binom{12}{1}}{\binom{27}{3}}=0,2051$$

4.2.3. Distribusi Probabilitas Variabel Acak Kontinu

Sama seperti pada distribusi probabilitas variabel acak diskrit, terdapat syarat yang harus dipenuhi pada distribusi probabilitas variabel acak kontinu, sebagai berikut :

- $f(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x)dx=1$

Fungsi probabilitas kumulatif variabel acak kontinu :

$$F(x)=P(X\leq x)=\int_{-\infty}^{x} f(x)dx$$

Nilai harapan (ekspektasi) rata-rata didefinisikan sebagai :

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

Varians nya adalah :

$$\sigma^2 = E\left[X-\mu\right]^2$$
$$= E\left[X^2\right] - E\left[X\right]$$

Sebagai contoh, jumlah jam, diukur dalam satuan 100 jam, suatu keluarga akan menggunakan mesin pengisap debu setahun berbentuk variabel acak kontinu dengan fungsi distribusi probbilitas sebagai berikut :

$$f(x) = \begin{cases} x & 0 < x < 1 \\ 2-x & 1 \le x < 2 \\ 0 & x, lainnya \end{cases}$$

Probabilitas atau peluang bahwa dalam setahun keluarga itu akan menggunakan mesin pengisap debu kurang dari 120 jam yaitu :

$$f\left(x < \frac{120}{100}\right) = f\left(x < 1,2\right)$$
$$= f\left(0 < x < 1,2\right)$$
$$= f\left(0 < x < 1\right) + f\left(1 < x < 1,2\right)$$
$$= \int_0^1 x\,dx + \int_1^{1,2} (2-x)\,dx$$
$$= \left[\frac{1}{2}x^2\right]_0^1 + \left[2x - \frac{1}{2}x^2\right]_1^{1,2}$$
$$= \left[\left(\frac{1}{2}\bullet 1^2\right) - \left(\frac{1}{2}\bullet 0^2\right)\right] + \left[\left((2\bullet 1,2) - \left(\frac{1}{2}\bullet 1,2^2\right)\right) - \left((2\bullet 1) - \left(\frac{1}{2}\bullet 1^2\right)\right)\right]$$
$$= \left[\frac{1}{2}\right] + \left[1,68 - 1,5\right]$$
$$= 0,5 - 0,18$$
$$= 0,32$$

4.2.3.1. Distribusi Normal

Distribusi noraml ditemukan oleh Cark Gauss pada abad ke-18 sehingga sering juga disebut sebagai distribusi Gauss. Distribusi normal adalah distribusi variabel kontinu yang memiliki kurva berbentuk lonceng dan simetris (Lukas, S.A., 2009). Sebuah sampel acak *X* berdistribusi normal jika fungsi distribusi peluangnya didefinisikan sebagai :

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

Dengan :

$-\infty < x < \infty$

$-\infty < \mu < \infty$

$\sigma > 0$

$\pi = 3,1415...$

$e = 2,71828...$

Distribusi normal memiliki sifat-sifat sebagai berikut (Sudjana, 2004) :

1. Grafiknya selalu ada di atas suatu sumbu mendatar *X*
2. Simetris terhadap $X=\mu$
3. Mempunyai satu modus, yakni nilai terbesar untuk Y, dicapai pada $X=\mu$ yang besarnya $0,3989/\sigma$
4. Grafiknya mendekati sumbu datar *X* yang mulai dari $X=\mu+3\sigma$ kanan dan $X=\mu-3\sigma$ ke kiri
5. Luas daerah di bawah lengkungan selalu sama dengan satu unit persegi.

4.2.3.2. Distribusi Normal Standard

Distribusi normal standard adalah distribusi normal yang memiliki rata-rata nol dan standard deviasi atau simpangan baku sama dengan satu (Lukas, S.A.,2009). Fungsi distribusi normal standar diperoleh dari hasil transformasi berikut :

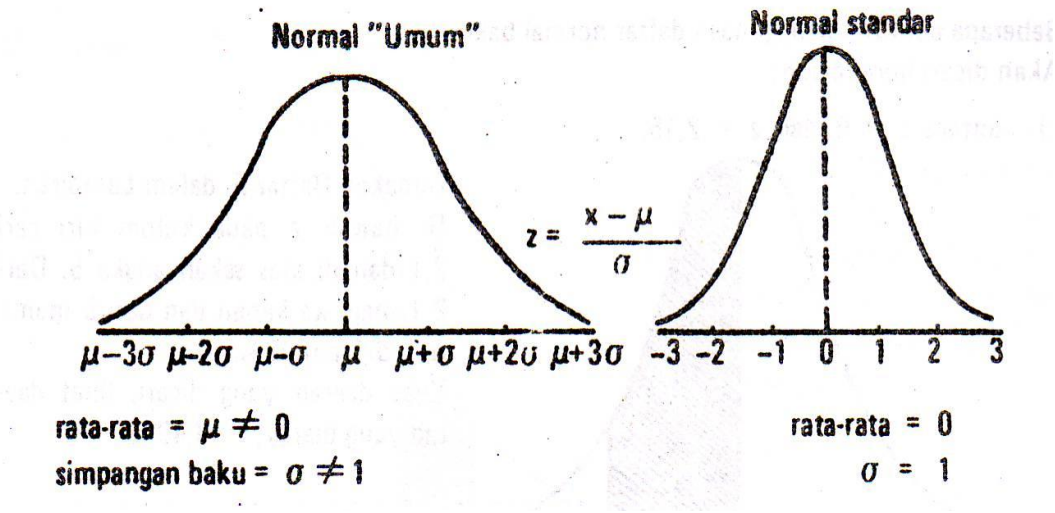
$$z = \frac{x-\mu}{\sigma}$$

Sehingga fungsi distribusi normal standar didefinisikan sebagai :

$$f(z \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

dengan $-\infty < z < \infty; \mu = 0; \sigma = 1; \pi = 3,1415...; e = 2,71828...$

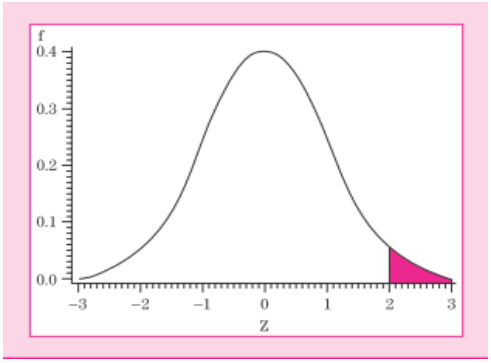
Perbedaan antara distribusi normal umum dan normal standar dapat dilihat pada kurva berikut :



Luas daerah bagian atas sumbu datar dan di bawah lengkungan normal standard yang telah dihitung hingga 4 desimal dan disusun dalam sebuah daftar, yaitu tabel distribusi normal standard (Sudjana, 2004). Ketentuan perhitungan luas distribusi normal standard :

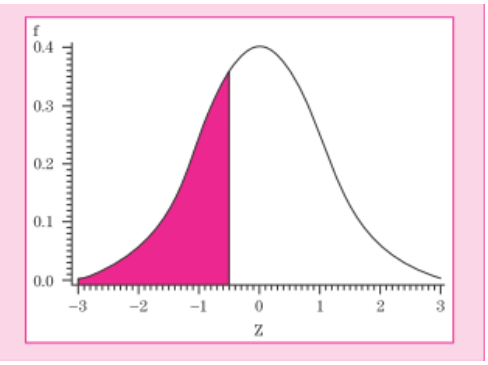
1. P(Z>z)=P(Z<-z)
2. P(Z<z)=1-P(Z>z)
3. Memungkinkan untuk operasi penambahan atau pengurangan pada sebuah kombinasi nilai.

Luas daerah normal standar dari 2 ke kanan sebagai berikut :



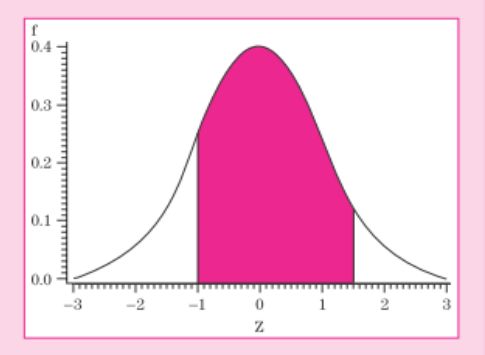
P(Z>2)=0,5-P(Z<2)=0,5-0,4772=0,0228

Luas daerah normal standar dari 0,5 ke kiri sebagai berikut :



P(Z<-0,5)=0,5-P(-0,5<Z<0) =0,5-0,1915 = 0,3085

Luas daerah normal standar antara -1 sampai dengan 1,5 adalah sebagai berikut :



P(-1<Z<1,5)=P(-1<Z<0)+P(0<Z<1,5)=0,3413+0,4332 =0,7745

Misalkan, nilai ujian masuk perguruan tinggi didistribusikan secara normal dengan rata-rata 75 dan simpangan baku 10. Berapakah probabilitas atau peluang seorang peserta ujian mencapai skor antara 80-90?

Untuk menyelesaikan persoalan di atas terdapat dua cara yaitu menggunakan distribusi normal atau distribusi normal standard. Jika persoalan diselesaikan dengan menggunakan distribusi normal, maka probabilitas atau peluang seorang peserta ujian mencapai skor antara 80-90 didefinisikan sebagai :

$$P(80 \le x \le 90) = \int_{80}^{90} \frac{1}{10\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-75}{10}\right)^2} dx$$

Integral di atas akan sulit diselesaikan, oleh karena alternatif untuk memudahkan menghitung besar probabilitas atau peluang seorang peserta ujian mencapai skor 80-90 dapat menggunakan distribusi normal standar dengan mentrasnformasikan nilai skor 80-90 menjadi nilai *z* sebagai berikut :

$$P(80 \le x \le 90) = P(\frac{80-75}{10} \le z \le \frac{90-75}{10})$$
$$= P(0,5 \le z \le 1,5)$$
$$= P(z \le 1,5) - P(z \le 0,5)$$
$$= 0,9332 - 0,6915$$
$$= 0,2417$$

Perhatikan contoh kedua berikut, masa hidup suatu batere yang dihasilkan oleh suatu perusahaan mendekati distribusi normal. Rata-rata masa hidupnya 300 jam sedangkan simpangan bakunya 35 jam. Tentukan :

a. Persentase hasil batu batere yang masa hidupnya antara 250 dan 350 jam
b. Persentasi hasil batu batere yang masa hidupnya paling sedikit 325 jam
c. Masa hidup batu batere jika diketahui bahwa terdapat 20% batu batere dengan masa hidup di atasnya.

Persentasi hasil batu batere yang masa hiduppnya 250 jam didefinisikan sebagai :

$$P(250 < x < 350) = P(\frac{250-300}{35} < z < \frac{350-300}{35})$$
$$= P(-1,43 < z < 1,43)$$
$$= P(-1,43 < z < 0) + P(0 < z < 1,43)$$
$$= 0,4236 + 0,4236$$
$$= 0,8472$$

Artinya, sekitar 84,72% dari produksi batu batere diharapkan akan mempunyai masa hidup antara 250 dan 350 jam.

Persentasi hasil batu batere yang masa hidupnya paling sedikit 325 jam didefinisikan sebagai :

$$P(x \ge 325) = P(z \ge \frac{325-300}{35})$$
$$= P(z \ge 0,71)$$
$$= 0,5 - P(0 < z < 0,71)$$
$$= 0,5 - 0,2612$$
$$= 0,2388$$

Artinya, terdapat sekitar 23,88% batu batere yang masa hidupnya paling sedikit 325 jam.

Masa hidup batu batere jika diketahui bahwa terdapat 20% batu batere dengan masa hidup di atasnya yaitu :

$$P(X > x) = 0,2$$

$$P(Z > \frac{x-300}{35}) = 0,2$$

$$P(0 < Z < \frac{x-300}{35}) = 0,5 - P(Z > \frac{x-300}{35})$$

$$P(0 < Z < \frac{x-300}{35}) = 0,5 - 0,2$$

$$P(0 < Z < \frac{x-300}{35}) = 0,3$$

$$z = 0,84$$

$$\frac{x-300}{35} = 0,84$$

$$x - 300 = 0,84 \times 35$$

$$x = 329,4$$

Artinya, masa hidup yang dimaksud adalah 329,4 jam.

LATIHAN

1. Sekelompok orang terdiri atas 50 orang dan 3 orang diantaranya lahir pada tanggal 31 Desember. Bila secara acak dipilih 5 orang, berapakah probabilitas orang yang terpilih itu :
   a. Tidak terdapat orang yang lahir pada tanggal 31 Desember?
   b. Tidak lebih dari satu orang yang lahir pada tanggal 31 Desember
2. Suatu percobaan mengenai ukuran ruang memori dengan menggunakan metode *quickshort* menyatakan bahwa ukurang penggunaan ruang memomri berdistribusi normal dengan rata-rata 510,8 *byte* dan simpangan baku 40,67 *byte*.
   a. Berapa persen dalam percobaan tersebut ditemukan ruang memori yang melebihi 600 *byte*.
   b. Jika ditemukan 10 buah percobaan mempunyai ruang memori berkisar antara 500-550 *byte*, berapakah jumlah percobaan yang telah dilakukan oleh peneliti?
   c. Jika dalam percobaan tersebut ditemukan bahwa 10% hasil terendah, berapakah ukuran memori tertingi dari kelompok hasil percobaan dengan ukuran memori terendah tersebut?
3. Andaikan lot sebesar 300 sekering listrik mengandung 5% rusak. Jika sampel dari 5 sekering diuji, carilah peluang pengamatan paling sedikit satu rusak!
4. Dari barang yang dihasilkan oleh semacam mesin ternyata 15% rusak. Kemudian diambil secara acak dari produksi barang itu sebanyak 30 buah untuk diselidiki. Berapa peluangnya dari barang yang diselidiki itu akan terdapat :
   a. Bagus semua
   b. Satu rusak
   c. Dua bagus
   d. Paling sedikit satu rusak
   e. Paling banyak dua rusak
5. Misalkan sebuah televisi diiklankan di surat kabar untuk dijual. Jika surat kabar yang memuat iklan tersebut memiliki 100.000 pembaca dan peluang seorang pembaca akan membalas iklan tersebut adalah 0,00002, maka :
   a. Berapa orangkah diharapkan akan membalas iklas itu?
   b. Berapa peluangnya bahwa yang membalas iklan hanya satu orang pembaca?
   c. Berapa peluangnya bahwa tidak satupun pembca yang membalas iklan tersebut?

6. Suatu perusahaan mengeluarkan produk baru *floopy disk* 5,25" dengan target konsumennya siswa SMU. Alat tersebut rata-rata berumur 11 bulan dengan simpangan baku 0,88. Jika pada produksi perdana perusahaan tersebut menghasilkan sebanyak 100.000 buah, berapakah :
    a. Peluang *floopy disk* tersebut berumur lebih dari 1 tahun.
    b. Peluang *floopy disk* tersebut berumur antara 9 sampai 13 bulan.
    c. Batas minimum umur dari 5% kelompok tertinggi *floopy disk* tersebut.
7. Suatu mesin dapat membuat suatu alat tahanan listrik dengan rata-rata tahanan 47,3 ohm dan simpangan baku 4,8 ohm. Tahanan tersebut menyebar mengikuti distribusi normal.
    a. Hitunglah probabilitas mesin tersebut yang mampu menghasilkan tahanan melebihi 50 ohm!.
    b. Jika ada 5.000 alat yang dihasilkan, berapa banyak alat yang mempunyai tahanan antara 40 sampai dengan 50 ohm?
    c. Jika ada 10% tahanan yang paling rendah, berapa sebenarnya nilai tahanan tertinggi untuk kelompok tersebut?.
8. Andaikan bahwa sistem acak dari patroli polisi direncanakan sehingga orang yang berpatroli dapat mengunjungi rute lokasi yang diberikan yaitu X=0,1,2,... kali per periode setengah jam, dan bahwa sistem disusun sehingga ia mengunjungi lokasi rata-rata sekali per periode waktu. Andaikan X mengikuti distribusi peluang Poisson, hitunglah peluang bahwa ia mengunjungi lokasi itu :
    a. Satu kali.
    b. Dua kali.
    c. Paling sedikit satu kali.

Distribusi Sampling

5.1. Definisi Distribusi Sampling

Seperti yang sudah diuraikan pada Bab 1, dalam sebuah penelitian, untuk memudahkan pengumpulan data, sampling lebih disukai dibandingkan dengan sensus. Hal ini dikarenakan dengan menggunakan sampling biaya, waktu dan tenaga yang dibutuhkan lebih sedikit dibandingkan dengan menggunakan sensus. Jika banyaknya objek dalam sebuah populasi adalah N sedangkan banyaknya objek yang akan dijadikan sampel adalah n, maka akan terdapat lebih dari sebuah sampel berukuran n yang mungkin bisa diambil dari populasi berukuran N . Hal ini disebabkan oleh adanya bermacam-macam kombinasi data sampel yang bisa terambil.

Dalam prakteknya hanya akan diambil sebuah sampel untuk digunakan dalam penelitian atau dengan kata lain hanya akan diambil satu buah kombinasi data sampel dari bermacam-macam kombinasi yang terbentuk. Sampel yang diambil biasanya dipilih secara acak, disebut dengan *sampel acak* atau *sampel random*.

Pada sampel yang bersangkutan akan diperoleh taksiran parameter populasi dari θ. Kumpulan nilai-nilai $\hat{\theta}$ pada sampel-sampel yang mungkin disebut sebagai distribusi sampling. Jika sampling dilakukan tanpa pengembalian, maka banyaknya sampel yang mungkin terbentuk adalah :

$$\binom{N}{n} = \frac{N!}{n!(N-n)!}$$

5.2. Distribusi Sampling Rata-Rata

Distribusi sampling rata-rata terbentuk pada pemilihan sampel yang bertujuan untuk menaksir atau menduga nilai rata-rata dari populasi. Tiap-tiap kemungkinan sampel akan memiliki rata-rata sampelnya. Anggap rata-rata ini sebagai data baru, maka akan terbentuk suatu kumpulan data yang terdiri dari rata-rata dari sampel-sampel. Dari kumpulan rata-rata tersebut dicari rata-rata dan simpangan bakunya, maka akan diperoleh rata-rata dari rata-rata, dinotasikan sebagai $\mu_{\bar{x}}$ dan simpangan baku dari rata-rata dinotasikan sebagai $\sigma_{\bar{x}}$.

Tabel berikut menyajikan data mengenai nilai intelegensi calon legislatif yang menggunakan ijazah palsu. Terdapat 5 calon legislatif yang menggunakan ijazah palsu dengan nilai intelegensi masing-masing 50,60,70,80 dan 90. Dari populasi 5 calon legislatif

tersebut, diambil 2 sampel secara berulang-ulang sampai semua kemungkinan sampel terambil.

No. Caleg	Nilai Intelegensi
1	50
2	60
3	70
4	80
5	90

Rata-rata nya adalah :

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{350}{5} = 70$$

Variansnya adalah :

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}} = \sqrt{\frac{(50-70)^2 + (60-70)^2 + (70-70)^2 + (80-70)^2 + (90-70)^2}{5}} = \sqrt{\frac{1000}{5}} = 14{,}14214$$

- Sampling dengan pengembalian

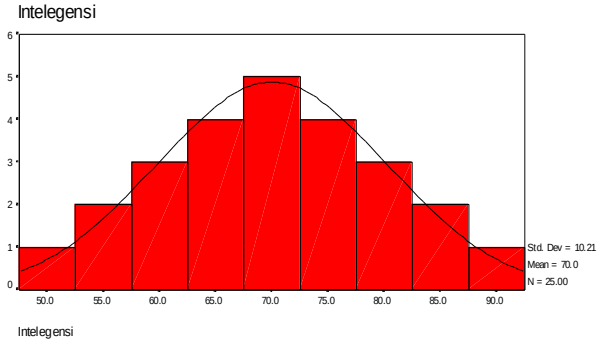
  Dalam sampling dengan pengembalian akan diperoleh sebanyak 5^5=25 buah kemungkinan sampel, yaitu :

Sampel	Caleg yang terpilih	Nilai Intelegensi	Rata-rata Nilai Intelegensi
1	1 ; 1	50 ; 50	50
2	1 ; 2	50 ; 60	55
3	1 ; 3	50 ; 70	60
4	1 ; 4	50 ; 80	65
5	1 ; 5	50 ; 90	70
6	2 ; 1	60 ; 50	55
7	2 ; 2	60 ; 60	60
8	2 ; 3	60 ; 70	65
9	2 ; 4	60 ; 80	70
10	2 ; 5	60 ; 90	75
11	3 ; 1	70 ; 50	60
12	3 ; 2	70 ; 60	65
13	3 ; 3	70 ; 70	70
14	3 ; 4	70 ; 80	75
15	3 ; 5	70 ; 90	80
16	4 ; 1	80 ; 50	65
17	4 ; 2	80 ; 60	70
18	4 ; 3	80 ; 70	75
19	4 ; 4	80 ; 80	80
20	4 ; 5	80 ; 90	85
21	5 ; 1	90 ; 50	70
22	5 ; 2	90 ; 60	75
23	5 ; 3	90 ; 70	80
24	5 ; 4	90 ; 80	85
25	5 ; 5	90 ; 90	90

Distribusi dari rata-rata tersebut dapat disajikan ke dalam bentuk berikut :

Rata-rata Nilai Intelegensi	Frekuensi	P(X)
50	1	0,04
55	2	0,08
60	3	0,12
65	4	0,16
70	5	0,2
75	4	0,16
80	3	0,12
85	2	0,08
90	1	0,04

Atau dalam grafik berikut :



Dari kumpulan rata-rata pada tabel sebelumnya, diperoleh jumlah rata-rata adalah 1750, maka rata-rata untuk ke-25 nilai rata-rata ini adalah :

$$\mu_{\bar{X}} = \frac{\sum_{i=1}^{25} \bar{X}_i}{25} = \frac{1750}{25} = 70$$

Sedangkan standard deviasi atau simpangan baku ke-25 nilai rata-rata tersebut dapat dihitung sebagai berikut :

$$\sigma_{\bar{X}} = \sqrt{\frac{\sum_{i=25}^{25} \left(\bar{X}_i - \mu_{\bar{X}}\right)^2}{25}}$$

$$= \sqrt{\frac{(50-70)^2 + (55-70)^2 + (60-70)^2 + \ldots (90-70)^2}{25}}$$

$$= \sqrt{\frac{2500}{25}} = 10$$

Dari hasil perhitungan rata-rata dan standard deviasi atau simpangan bakunya diperoleh bahwa :

Rata-Rata Populasi = Rata-Rata ke-25 Kombinasi Sampel

$$\mu = \mu_{\bar{X}} = 70$$

Sementara nilai simpangan baku dari populasi adalah 14,14214 sedangkan standard deviasi atau simpangan baku dari ke-25 rata-rata adalah 10, sehingga dapat dihitung bahwa :

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{14{,}14214}{\sqrt{2}} = 10$$

Dengan demikian, dapat disimpulkan bahwa :

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

Persamaan di atas juga berlan ku untuk kasus pengambilan sampel tanpa pengembalian jika *N* cukup besar dibandingkan dengan *n*, dalam hal ini jika :

$$\frac{n}{N} \leq 5\%$$

- Sampling tanpa pengembalian

Dalam sampling dilakukan tanpa pengembalian, maka banyaknya kemungkinan sampel yang terbentuk adalah :

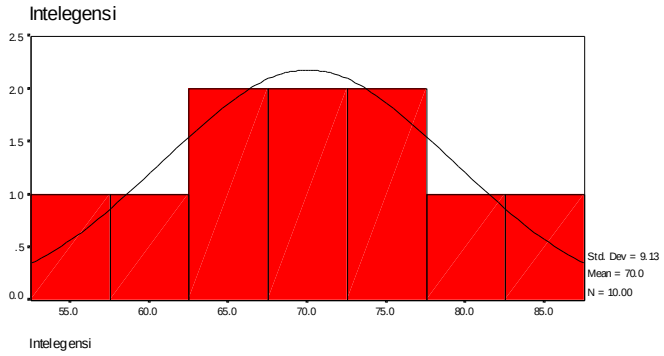
$$C_2^5 = \frac{5\ !}{(5-2)\ !\ \ 2\ !} = 10$$

Sampel	Caleg yang terpilih	Nilai Intelegensi	Rata-rata Nilai Intelegensi
1	1 ; 2	50 ; 60	55
2	1 ; 3	50 ; 70	60
3	1 ; 4	50 ; 80	65
4	1 ; 5	50 ; 90	70
5	2 ; 3	60 ; 70	65
6	2 ; 4	60 ; 80	70
7	2 ; 5	60 ; 90	75
8	3 ; 4	70 ; 80	75
9	3 ; 5	70 ; 90	80
10	4 ; 5	80 ; 90	85

Distribusi dari rata-rata tersebut dapat disajikan dalam bentuk berikut :

Rata-rata Nilai Intelegensi	Frekuensi	P(X)
55	1	0,1
60	1	0,1
65	2	0,2
70	2	0,2
75	2	0,2
80	1	0,1
85	1	0,1

Selanjutnya akan ditampilkan ke dalam bentuk grafik sebagai berikut :



Dari kumpulan rata-rata di atas, diperoleh jumlah rata-rata adalah 490, maka rata-rata untuk ke-25 rata-rata ini adalah :

$$\mu_{\bar{X}} = \frac{\sum_{i=1}^{10} \bar{X}_i}{10} = \frac{700}{10} = 70$$

Simpangan baku atau standard deviasi ke-25 rata-rata tersebut adalah :

$$\sigma_{\bar{X}} = \sqrt{\frac{\sum_{i=25}^{10} \left(\bar{X}_i - \mu_{\bar{X}} \right)^2}{10}}$$

$$= \sqrt{\frac{(55-70)^2 + (60-70)^2 + (65-70)^2 + \ldots + (85-70)^2}{10}}$$

$$= \sqrt{\frac{750}{10}} = 8,66$$

Dari hasil perhitungan rata-rata dan standard deviasi atau simpangan bakunya diperoleh bahwa :

Rata-Rata Populasi = Rata-Rata ke-25 Kombinasi Sampel

$$\mu = \mu_{\bar{X}} = 70$$

Sementara nilai simpangan baku dari populasi adalah 14,14214 sedangkan standard deviasi atau simpangan baku dari ke-25 rata-rata adalah 8,66, sehingga dapat dihitung bahwa :

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} = \frac{14,14214}{\sqrt{2}} \sqrt{\frac{5-2}{5-1}} = 8,66$$

Dengan demikian, dapat disimpulkan bahwa :

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

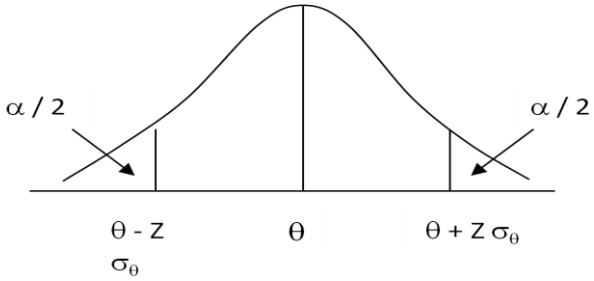
$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

Jika terdapat sebuah populasi dengan rata-rata μ atau proporsi π dan simpangan baku (standard deviasi) σ yang besarnya terhingga, maka distribusi sampel berdasarkan pengambilan sampel n secara acak dan berulang-ulang memiliki beberapa sifat yaitu :

- Rata-rata distribusi sampel untuk statistik akan sama dengan parameter populasi θ

- Simpangan baku statistik sampel akan sama dengan $\frac{\sigma}{\sqrt{n}}$. Ukuran ini juga dikenal sebagai *standard error* (*SE*) yang memegang persanan penting pada estimasi parameter dan uji statistik.
- Jika distribusi nilai pada populasi normal, maka distribusi sampel juga normal. Tetapi yang lebih penting adalah jika distribusi nilai pada populasi tidak normal, dengan jumlah sampel yang cukup besar maka distribusi sampel akan mendekati normal, tanpa tergantung dari distribusinya nilai parameter populasinya.

Dengan asumsi besar sampel yang cukup, distribusi sampel *x* dapat digambarkan sebagai berikut :



Kurva di atas menunjukkan nilai *standard error* dari rata-rata distribusi sampel. Nilai *α* merupakan taraf signifikansi yang menunjukkan derajat kekeliruan yang diberikan.

Misalkan, sebuah Bank menghitung tabungan seluruh nasabahnya, setelah perhitungan, Bank tersebut mendapati bahwa rata-rata tabungan setiap nasabahnya sebesar Rp. 2.000,- dengan standard deviasi Rp. 600,-. Apabila seorang peneliti mengambil sampel sebanyak 100 nasabah, berapakah probabilitas jika :

a. Rata-rata sampel akan terletak antara Rp. 1900 dan Rp. 2050?
b. Rata-rata sampel akan lebih besar dari Rp. 2050
c. Rata-rata sampel akan lebih kecil dari Rp. 1900
d. Rata-rata sampel akan atau lebih kecil dari Rp. 1900

Untuk dapat menghitung keempat nilai probabilitas tersebut, terlebih dahulu kita definisikan nilai rata-rata dari distribusi rata-rata dan simpangan bakunya sebagai berikut :

$$\mu_{\bar{x}} = \mu = 2000$$
$$\sigma_{\bar{x}} = \sigma \big/ \sqrt{n} = 600 \big/ \sqrt{100} = 60$$

Probabilitas jika rata-rata sampel akan terletak antara Rp. 1.900,- dan Rp. 2.050,- yaitu :

$$P(1900 < \mu_{\bar{x}} < 2050) = P\left(\frac{1900-2000}{60} < z < \frac{2050-2000}{60}\right)$$
$$= P(-1,67 < z < 0,83)$$
$$= 0,4525 + 0,2967$$
$$= 0,7492$$

Probabilitas jika rata-rata sampel akan lebih besar dari Rp. 2.050,- yaitu :

$$P(\mu_{\bar{x}} > 2050) = P\left(z > \frac{2050-2000}{60}\right)$$
$$= P(z > 0,83)$$
$$= 0,5 - 0,2967$$
$$= 0,2033$$

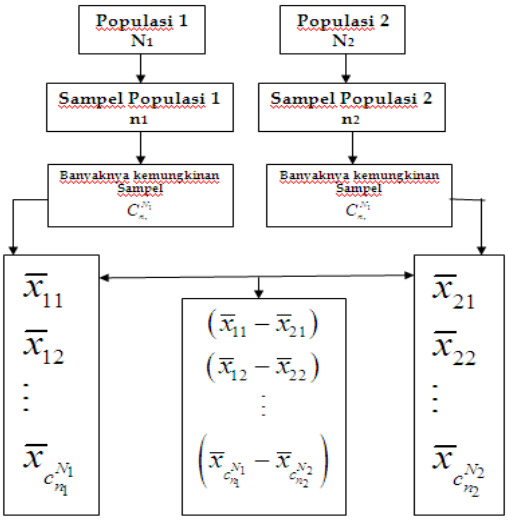
Probabilitas jika rata-rata sampel akan lebih kecil dari Rp. 1.900,- yaitu :

$$P(\mu_{\bar{x}} < 1900) = P\left(z < \frac{1900-2000}{60}\right)$$
$$= P(z < -1,67)$$
$$= 0,5 - 0,4525$$
$$= 0,0475$$

Probabilitas jika rata-rata sampel akan atau lebih kecil dari Rp. 1900

$$P(\mu_{\bar{x}} \leq 1900) = P\left(z \leq \frac{1900-2000}{60}\right)$$
$$= P(z \leq -1,67)$$
$$= 0,5 - 0,4525$$
$$= 0,0475$$

Jika terdapat dua buah populasi masing-masing berukuran N_1 dan N_2 kemudian masing-masing diambil sampel secara acak berukuran n_1 dan n_1, maka masing-masing populasi akan memiliki distribusi sampling rata-rata kemudian dicari selisihnya akan diperoleh sebuah distribusi sampling beda rata-rata.



Nilai rata-rata dan simpangan baku dari distribusi sampling beda rata-rata didefinisikan sebagai berikut :

$$\mu_{(\bar{x}_1-\bar{x}_2)} = \mu_1 - \mu_2$$

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

Perhatikan contoh berikut, lampu pijar merk "Ampuh" memiliki rata-rata daya tahan 4500 jam dengan standar deviasi (simpangan baku) 500 jam, sedangkan lampu pijar merk "Baik" memiliki rata-rata daya tahan 4000 jam dengan standar deviasi (simpangan baku) 400 jam. Jika diambil sampel masing-masing 100 buah lampu pijar dan diteliti, berapa probabilitas bahwa selisih rata-rata daya tahan kedua lampu pijar tersebut lebih besar dari 600 jam?

Untuk dapat menyelesaikan persoalan di atas, terlebih dahulu definisikan rata-rata dan simpangan baku dari selisih rata-rata daya tahan kedua lampu pijar tersebut sebagai berikut :

$$\mu_{(\bar{x}_1-\bar{x}_2)} = \mu_1 - \mu_2 = 4500 - 4000 = 500$$

$$\sigma_{(\bar{x}_1-\bar{x}_2)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{500^2}{100} + \frac{400^2}{100}} = 64$$

Maka probabilitas bahwa selisih rata-rata daya tahan kedua lampu pijar tersebut lebih besar dari 600 jam adalah :

$$P\left((\bar{x}_1 - \bar{x}_2) > 600\right) = P\left(z > \frac{600-500}{64}\right)$$
$$= P\left(z > 1,56\right)$$
$$= 0,5 - 0,4406$$
$$= 0,059$$

5.3. Distribusi Sampling Proporsi

Distribusi sampling proporsi terbentuk dari hasil pengambilan sampel secara acak yang bertujuan untuk menaksir atau menduga nilai proporsi suatu peristiwa dari sebuah populasi. Dimisalkan bahwa ukuran dari populasi adalah *N* dan ukuran sampel yang diambil adalah *n*. Apabila dari populasi tersebut terdapat *Y* buah peristiwa khusus yang akan diteliti, maka proporsi terjadinya peristiwa tersebut adalah :

$$\pi = \frac{Y}{N}$$

Berdasarkan sampel yang diambil, ternyata peristiwa khusus yang diperoleh ada sebanyak *x* buah, maka diperoleh statistik proporsi peristiwa tersebut adalah :

$$p = \frac{x}{n}$$

Jika proses pengambilan sampel dilakukan tanpa pengembalian atau jika kondisi populasi memiliki ukuran yang tidak terlalu besar dibandingkan dengan data sampelnya yaitu $\frac{n}{N} > 5\%$

Maka nilai rata-rata dan simpangan baku dari sebuah distribusi sampling proporsi didefinisikan sebagai :

$$\mu_p = p$$

$$\sigma_p = \sqrt{\frac{p(1-p)}{n}} \sqrt{\frac{N-n}{N-1}}$$

Jika proses pengambilan sampel dilakukan dengan pengembalian atau ukuran populasinya besar dibandingkan dengan ukuran sampel, yaitu $\frac{n}{N} \le 5\%$, maka nilai rata-rata dan simpangan baku dari sebuah distribusi sampling proporsi didefinisikan sebagai :

$$\mu_p = p$$

$$\sigma_p = \sqrt{\frac{p(1-p)}{n}}$$

Perhatikan contoh berikut, dari 1000 buah mobil yang diproduksi, diketahui 100 diantaranya cacat. Jika diambil sampel random sebanyak 500 buah mobil dari populasi tersebut dan diteliti berapa besar probabilitas proporsi mobil yang cacat lebih besar dari 12%?

Terlebih dahulu hitung proporsi mobil cacat dari populasi yaitu :

$$\pi = \frac{100}{1000} = 0,1$$

Dengan demikian akan diperoleh rata-rata proporsi mobil cacat dari sampel sebagai berikut :

$$\mu_{\bar{p}} = \pi = 0,1$$

Dengan

$$\frac{n}{N} = \frac{500}{1000} = 0,5$$

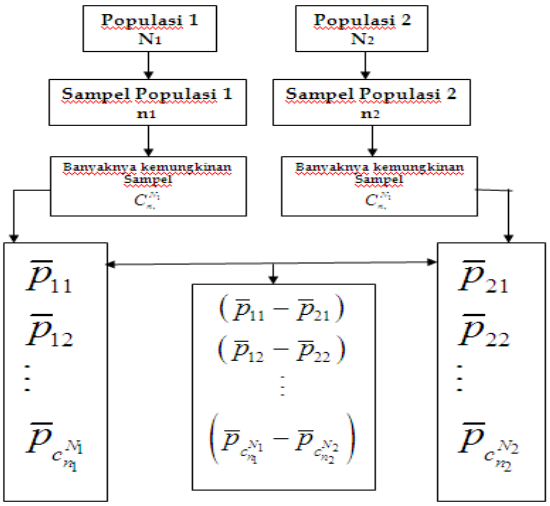
Lebih besar dari 5%, maka simpangan baku dari distribusi sampling proporsi dapat dihitung sebagai berikut :

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}\sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{0,1(1-0,1)}{500}}\sqrt{\frac{1000-500}{1000-1}} = 0,009492$$

Sehingga, probabilitas proporsi mobil yang cacat lebih besar dari 12% adalah sebagai berikut:

$$\begin{aligned} P(p > 0,12) &= P\left(z > \frac{0,12-0,1}{0,009492}\right) \\ &= P(z > 2,1071) \\ &= 0,5 - 0,482447 \\ &= 0,017553 \end{aligned}$$

Jika terdapat dua buah populasi masing-masing berukuran N_1 dan N_2 kemudian masing-masing diambil sampel secara acak berukuran n_1 dan n_1, maka masing-masing populasi akan memiliki distribusi sampling proporsi kemudian dicari selisihnya akan diperoleh sebuah distribusi sampling beda proporsi.



Nilai rata-rata dan simpangan baku dari distribusi sampling beda rata-rata didefinisikan sebagai berikut :

$$\mu_{(\bar{p}_1-\bar{p}_2)} = p_1 - p_2$$

$$\sigma_{(\bar{p}_1-\bar{p}_2)} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

$$\sigma_{(\bar{p}_1-\bar{p}_2)} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \sqrt{\frac{N-n}{N-1}}$$

Misalkan, berdasarkan sebuah penelitian, 1% dari orang yang tidak merokok terkena TBC dan dari setiap 100 orang perokok, 5 orang diantaranya terkena TBC. Jika diambil sampel masing-masing 100 orang perokok dan populasi orang tidak merokok. Berapakah probabilitas yang terkena TBC lebih besar dari 5%?

Anggap bahwa proporsi populasi perokok terkena TBC sebagai p_1 dan proporsi populasi bukan perokok terkena TBC adalah p_2, maka :

$$\begin{aligned} P\left((\bar{p}_1 - \bar{p}_2) > 0,05\right) &= P\left(z > \frac{0,05-0,04}{0,024}\right) \\ &= P\left(z > 0,42\right) \\ &= 0,5 - 0,1628 \\ &= 0,3372 \end{aligned}$$

LATIHAN

1. Jelaskan apa yang dimaksud dengan :
    a. Distribusi sampling
    b. Distribusi sampling rata-rata
    c. Distribusi sampling proporsi
2. Berapakah rata-rata dan simpangan baku untuk distribusi sampling :
    a. Rata-rata
    b. Selisih dua rata-rata
    c. Proporsi
    d. Selisih dua proporsi
3. *A* dan *B* menghasilkan dua macam kabel, yang daya tahannya masing-masing adalah 4.000 dan 4.500 kg dengan simpangan baku 300 dan 200 kg. Jika dari kabel yang dihasilkan oleh *A* diuji 100 potong sedangkan dari *B* diuji 50 potong maka tentukanlah peluang daya tahan *B* :
    a. Paling sedikit 600 kg lebih daripada daya tahan kabel *A*?
    b. Paling banyak 450 kg lebih daripada daya tahan kabel *A*?
4. Tinggi Badan Mahasiswa rata-rata mencapai 165 cm dan simpangan baku 8,4 cm. Telah diambil sebuah sampel acak terdiri atas 45 cm. Tentukan berapa peluang tinggi rata-rata dari k3 45 mahasiswa tersebut :
    a. antara 160 cm dan 168 cm
    b. Paling sedikit 166 cm
5. Ada petunjuk kuat bahwa 10% anggota masyarakat tergolong ke dalam golongan A. Sebuah sampel acak terdiri atas 100 orang telah diambil. Tentukan peluangnya bahwa 100 orang itu akan ada paling sedikit 15 orang dari golongan A.
6. Rata-Rata Tinggi mahasiswa laki-laki 163 cm dan simpangan bakunya 5,2 cm. Sedangkan untuk mahasiswa perempuan, parameter tersebut berturut-turut 152 cm dan 4,9 cm.
7. Dari kedua kelompok mahasiswa itu masing-masing diambil sebuah sampel acak, secara independen, berukuran sama yaitu 140 orang. Berapa peluang rata-rata tinggi mahasiswa laki-laki paling sedikit 10 cm lebihnya dari rata-rata tinggi mahasiswa perempuan?
8. Ada petunjuk kuat bahwa calon A akan mendapatkan suara 60% dalam pemilihan. Dua buah sampel acak secara independen telah diambil masing-masing terdiri atas 300 0rang.

Tentukan peluangnya akan terjadi perbedaan persentase tidak lebih dari 10% yang akan memilih A.

9. Ada semacam barang yang dihasilkan oleh dua pengusaha *A* dan *B*. Biasanya barang yang dihasilkan oleh pengusaha *A* ternyata terjadi kerusakan sekitar 5%, sedangkan oleh pengusaha *B* kerusakan terjadi sekitar 4%. Dari barang-barang yang dihasilkan oleh kedua pengusahan ini, diambil sampel acak, masing-masing terdiri atas 100 barang. Berapakah peluangnya bahwa kerusakan barang yang dihasilkan pengusaha *A* akan berbeda lebih dari 0,5% dibandingkan terhadap kerusakan barang yang dihasilkan oleh pengusaha *B*?
10. Dari suatu proses produksi semacam barang, ternyata 90% diproduksi tanpa cacat dan 10% lagi dalam keadaan rusak. Setiap hari kerja, selama proses berlangsung, diambil sebuah sampel acak terdiri atas 100 barang. Tentukanlah :
    a. Rata-rata persentase barang yang rusak dan simpangan bakunya?
    b. Peluang barang rusak dari sebuah sampel berukuran 100 paling kecil 15%?
    c. Berapa ukuran sampel paling sedikit, agar jika kita mengambil sampel cukup banyak dengan ukuran tersebut, persentase kerusakannya diharapkan akan berbeda tidak lebih dari 2%?

Pengujian Hipotesis

6.1. Definisi Hipotesis

Seperti yang sudah disebutkan pada bab sebelumnya, ruang lingkup statistika terbagi menjadi dua, yaitu statistika deskriptif dan statistika inferensial/induktif. Statistika yang berkenaan dengan cara penarikan kesimpulan berdasarkan data yang diperoleh dari sampel untuk menggambarkan karakteristik atau ciri-ciri dari suatu populasi atau sifatnya adalah konfirmatif disebut sebagai statistika inferensial/induktif.

Statistika inferensi pada dasarnya adalah suatu keputusan, perkiraan atau generalisasi tentang suatu populasi berdasarkan informasi yang terkandung dari suatu sampel. Seperti yang telah uraikan pada bab sebelumnya, ruang lingkup statistika inferensial/induktif terbagi menjadi dua, yaitu estimasi dan pengujian hipotesis. Dalam estimasi, kita akan mencari nilai dugaan atau taksiran dari nilai parameter populasi melalui nilai-nilai statistik yang diperoleh dari data sampel. Sementara dalam pengujian hipotesis kita akan menguji kebenaran suatu hipotesis.

Hipotesis adalah rumusan pernyataan ilmiah sebagai jawaban terhadap masalah serta masih memerlukan pengujian empiris. Sementara pengujian hipotesis adalah langkah-langkah yang dilakukan dengan tujuan untuk memutuskan apakah hipotesa tersebut diterima atau ditolak. Dengan demikian, jelas akan terdapat dua kemungkinan hasil dalam pengujian hipotesis yaitu menolak atau menerima hipotesis. Menolak hipotesis artinya bahwa hipotesis tidak benar sedangkan menerima hipotesis artinya tidak cukup bukti untuk menolak hipotesis.

6.2. Cara Menguji Hipotesis

Dalam menguji sebuah hipotesis atau pernyataan ilmiah, diperlukan sebuah langkah empiris dengan langkah-langkah sebagai berikut :

1. Merumuskan hipotesis yaitu H_0 dan H_1
2. Menentukan taraf signifikansi (α)
3. Menentukan nilai statistik uji
4. Menentukan nilai kritis
5. Membuat kesimpulan

6.2.1. Merumuskan Hipotesis

Dalam merumuskan hipotesis, harus dapat dibedakan H_0 dan H_1. Hipotesis nol (H_0) atau *The null hypothesis is a statement about the values of one or more parameters. This hypothesis represents the status quo and is usually not rejected unless the sample results strongly imply that it is false* (Rudolf. J.Freud, 2003). Hipotesis nol (H_0) digunakan untuk menyatakan kondisi parameter yang akan diuji. Pada umumnya menggunakan notasi "=" yang mengindikasikan kondisi yang sama atau tidak berubah.

Sedangkan hipotesis satu (H_1) atau sering juga disebut hipotesis alternatif atau hipotesis tandingan secara umum menyatakan bahwa hipotesis nol tidak benar. *The alternative hypothesis is a statement that contradicts the null hypothesis. This hypothesis is declared to be accepted if the null hypothesis is rejected. The alternative hypothesis is often called the research hypothesis because it usually implies that some action is to be performed,some money spent, or some established theory overturned* (Rudolf, J.F.,2003).

Berikut adalah tanda-tanda yang sering digunakan pada masing-masing hipotesis :

Tanda dalam Ho: **=, ≤,** atau ≥
Tanda dalam H_1: **≠, <, or** >

Berdasarkan tanda yang digunakan dalam H_1, pengujian hipotesis dibedakan menjadi dua. Jika hipotesis satu (H_1) atau hipotesis alternatif atau hipotesis tandingan menggunakan tanda "≠" artinya pengujian hipotesis yang dilakukan adalah pengujian hipotesis dua pihak. Sebaliknya jika hipotesis satu (H_1) atau hipotesis alternatif atau hipotesis tandingan menggunakan tanda "<" atau ">" artinya pengujian hipotesis yang dilakukan adalah pengujian hipotesis satu pihak. Pengujian hipotesis satu pihak ini, dibedakan menjadi dua, yaitu pihak kanan dan pihak kiri. Pengujian hipotesis disebut sebagai pengujian pihak kanan jika H_1 nya mengandung tanda ">", sebaliknya jika H_1 mengandung tanda "<" pengujian hipotesis yang dilakukan adalah pengujian pihak kiri.

Penentuan pengujian dua pihak, pihak kanan atau pihak kiri akan berpengaruh terhadap penentuan nilai kritis yang nantinya akan dibandingkan dengan nilai hitung (Statistik Uji) yang diperoleh.

Perhatikan contoh berikut, misalkan kita memiliki tiga pernyataan yang ingin diketahui kebenarannya, yaitu :

1. Peluang lahir bayi laki-laki adalah 0,5
2. Rata-rata nilai TOEFL mahasiswa adalah 450

3. Rata-rata nilai TOEFL mahasiswa adalah lebih dari 450

Dalam mendeskripsikan ketiga pernyataan ilmiah tersebut ke dalam bentuk hipotesis statistik, yang perlu diperhatikan adalah parameter apa yang diukur dan nilainya. Untuk pernyataan pertama, parameter yang diukur adalah sebuah *peluang*. Dalam statistika, *peluang* dinotasikan sebagai *p*, kemudian nilai parameter yang diperoleh dari pernyatan ilmiah tersebut adalah 0,5, artinya pernyataan mengandung tanda "=". Dengan demikian, informasi yang terkandung dalam pernyataan dijadikan sebagai hipotesis nol (H_0), artinya hipotesis satu atau tandingannya mengandung tanda "≠". Hipotesis statistika dari pernyataan pertama dapat didefinisikan sebagai :

H_0 : p=0,5 (Peluang lahir bayi laki-laki adalah 0,5)

H_1 : p≠0,5 (Peluang lahir bayi laki-laki bukan 0,5)

Dalam pernyataan ilmiah yang ke-2, parameter yang diukur adalah rata-rata. dalam statistika parameter rata-rata dinotasikan sebagai μ. Nilai dari rata-rata yang dinyatakan dalam pernyataan ilmiah adalah 450, artinya mengandung tanda "=". Sama halnya dengan pernyataan ke-1, informasi yang terkandung dalam pernyataan dijadikan sebagai hipotesis nol (H_0), artinya hipotesis satu atau tandingannya mengandung tanda "≠". Hipotesis statistika dari pernyataan pertama dapat didefinisikan sebagai :

H_0 : μ =450 (Rata-rata nilai TOEFL mahasiswa adalah 450)

H_1 : μ ≠450 (Rata-rata nilai TOEFL mahasiswa bukan 450)

Dalam pernyataan ilmiah yang ke-3, parameter yang diukur adalah rata-rata. dalam statistika parameter rata-rata dinotasikan sebagai μ. Nilai dari rata-rata yang dinyatakan dalam pernyataan ilmiah adalah lebih besar dari 450, artinya mengandung tanda ">". Dengan informasi yang terkandung dalam pernyataan dijadikan sebagai hipotesis satu (H_1), artinya hipotesis nol atau tandingannya mengandung tanda "≤". Hipotesis statistika dari pernyataan pertama dapat didefinisikan sebagai :

H_0 : μ ≤450 (Rata-rata nilai TOEFL mahasiswa kurang dari atau sama dengan 450)

H_1 : μ >450 (Rata-rata nilai TOEFL mahasiswa lebih dari 450)

6.2.2. Menentukan Taraf Signifikansi

Dalam sebuah pengujian hipotesis, terdapat dua jenis eror yaitu eror tipe I dan eror tipe II. *A type I error occurs when we incorrectly reject H_0, that is, when H_0 is actually true and our sample-based inference procedure rejects it and A type II error occurs when we*

incorrectly fail to reject H_0, that is, when H_0 is actually not true, and our inference procedure fails to detect this fact. (Rudolf,J.f.,2003). Peluang kemungkinan terjadinya eror tipe I dinyatakan sebagai α, sedangkan peluang kemungkinan terjadinya eror tipe II dinyatakan sebagai β.

Taraf signifikansi atau *the significance level of a hypothesis test is the maximum acceptable probability of rejecting a true null hypothesis* (Rudolf,J.F.,2003). Dengan demikian penentuan taraf signifikansi sama dengan menentukan nilai maksimum α yang masih dapat diterima.

Taraf signifikansi ditentukan oleh peneliti yang didasarkan pada pertimbangan mengenai keseriusan atau biaya pembuatan eror tipe I. Besar taraf signifikansi yang sering digunakan adalah 5% dan 1%. Besar taraf signifikansi yang digunakan akan berhubungan dengan derajat kepercayaan hasil pengujian yang diperoleh. Jika taraf signifikansi yang digunakan adalah 5%, maka derajat kepercayaan hasil yang diperoleh adalah (1-0,05) atau 95%. Sedangkan jika taraf signifikansi yang digunakan adalah 1%, maka derajat kepercayaan hasil yang diperoleh adalah (1-0,01) atau 99%.

6.2.3. Menentukan Nilai Statistik Uji

Statistik uji atau *the test statistic is a sample statistic whose sampling distribution can be specified for both the null and alternative hypothesis case (although the sampling distribution when the alternative hypothesis is true may ofthen be quite complex). After specifying the appropriate significance level of α, the sampling distribution of this statistic is used to define the rejection region* (Rudolf,J.F.,2003).

Berdasarkan definisi tersebut, nilai statistik uji ditentukan berdasarkan asumsi yang dipenuhi atau kondisi data sampel. Misalnya, dalam pengujian rata-rata satu sampel, nilai statistik uji ditentukan dengan asumsi apakah populasi mengikuti distribusi normal dan simpangan baku populasi diketahui atau tidak.

6.2.4. Menentukan Nilai Kritis

Nilai kritis atau *the rejection region (also called the critical region) is the range of values of a sample statistic that will lead to rejection of the null hypothesis* (Rudolf,J.f.,2003). Nilai kritis dihitung bergantung pada statistik uji yang dilakukan dan biasanya diperoleh dari tabel distribusi tertentu berdasarkan taraf signifikansi dan statistik uji yang digunakan.

6.2.5. Membuat Kesimpulan

Kesimpulan diperoleh dengan membandingkan nilai statistik uji dengan nilai kritis. Berikut adalah kesimpulan yang mungkin diperoleh dalam pengujian hipotesis (Rudolf,J.F.,2003) :

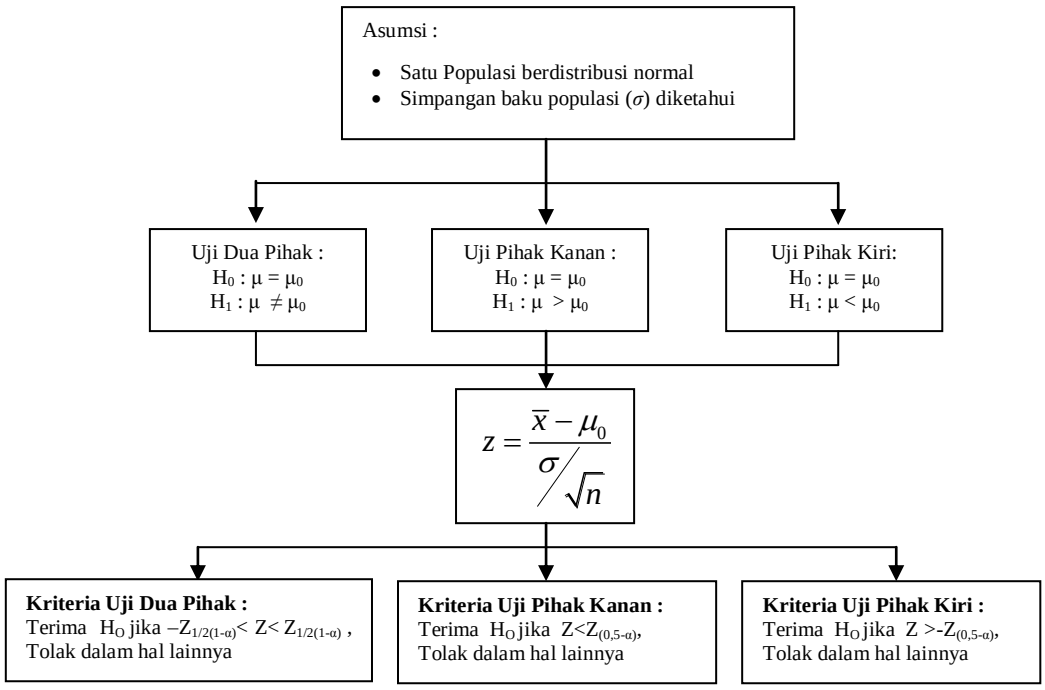
Kesimpulan	**Populasi**	
	H_0 Benar	**H_0 Tidak Benar**
H_0 diterima	Kesimpulan Benar	*Error* tipe II
H_0 ditolak	*Error* Tipe I	Kesimpulan Benar

6.3. Uji Rata-Rata

Pengujian hipotesis rata-rata dibagi berdasarkan banyaknya populasi dan asumsi yang dipenuhi sebagai berikut :

a) Uji Rata-Rata Satu Sampel

Uji rata-rata satu sampel terbagi menjadi dua berdasarkan simpangan baku populasinya diketahui atau tidak, sebagai berikut :



Perhatikan kasus berikut, pengusaha lampu pijar A mengatakan bahwa lampunya bisa tahan pakai sekitar 800 jam. Akhir-akhir ini timbul dugaan bahwa masa pakai

lampu itu telah berubah. Untuk menentukan hal ini, dilakukan penelitian dengan cara menguji 50 buah lampu. Ternyata rata-ratanya 792 jam. Dari pengalaman, diketahui bahwa simpangan baku masa hidup lampu 60 jam. Selidikilah dengan taraf signifikansi 5%, apakah kualitas lampu sudah berubah atau belum?

Berdasarkan kasus tersebut, maka hipotesis yang diuji adalah sebagai berikut :

H_0 : μ=800 (Rata-rata masa pakai lampu adalah 800 jam)

H_1 : $\mu \neq 800$ (Rata-rata masa pakai lampu bukan 800 jam)

$\alpha = 0{,}05$

Dari uraian kasus diperoleh informasi mengenai ukuran sampel (n) yaitu 50 buah lampu, rata-rata masa pakai dari kelima puluh lampu ($\bar{x}$) yaitu 792 jam dan simpangan baku dari populasi lampu (σ) yaitu 60 jam, dengan demikian statistik ujinya dapat dihitung sebagai berikut :

$$Z = \frac{792 - 800}{60/\sqrt{50}} = -0.94$$

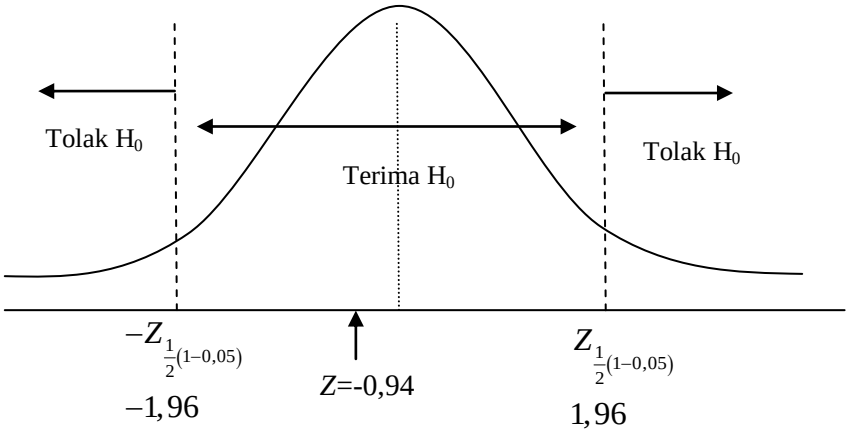
Oleh karena, pengujian hipotesis yang dilakukan adalah pengujian hipotesis dua pihak, maka nilai kritis yang diperoleh dari Tabel Distribusi Normal Standard adalah :

$$Z_{\frac{1}{2}(1-\alpha)} = Z_{\frac{1}{2}(1-0{,}05)} = 1{,}96$$

Kriteria Uji :

Terima H_O jika $-Z_{1/2(1-\alpha)} < Z < Z_{1/2(1-\alpha)}$, Tolak dalam hal lainnya

Untuk memudahkan dalam mengambil kesimpulan, gambarkan kurva daerah penolakan dan penerimaan H_0 sebagai berikut :



Berdasarkan nilai statistik uji yang diperoleh dan kriteria uji yang digunakan, dapat dilihat pada kurva bahwa nilai statistik uji (Z) yang bernilai -0,94 berada di daerah penerimaan H_0 sehingga dapat disimpulkan bahwa H_0 diterima. Dengan kata lain, kualitas lampu pijar A belum berubah atau rata-rata masa pakainya masih 800 jam.

Kasus selanjutnya, proses pembuatan barang rata-rata menghasilkan 15,7 unit per jam. Hasil produksi mempunyai varians sama dengan 2,3. Metode baru diusulkan untuk mengganti yang lama jika rata-rata per jam menghasilkan paling sedikit 16 unit. Untuk menentukan apakah metode diganti atau tidak, metode baru dicoba 20 kali dan ternyata rata-rata per jam menghasilkan 16,9 unit. Pengusaha bermaksud mengambil resiko 5% untuk menggunakan metode baru apabila metode ini rata-rata menghasilkan lebih dari 16 unit. Metode mana yang harus diputuskan oleh pengusaha tersebut?.

Hipotesis yang diuji adalah :

H_0 : μ=16 (Metode Baru menghasilkan rata-rata per jam 16 unit)

H_1 : μ>16 (Metode Baru menghasilkan rata-rata per jam lebih dari 16 unit)

$\alpha = 0,05$

Dari uraian kasus diperoleh informasi mengenai ukuran sampel (n) yaitu 20, rata-rata per jam yang dihasilkan ($\bar{x}$) yaitu 16,9 unit dan varians dari populasi (σ^2) yaitu 2,3 sehingga simpangan bakunya adalah $\sigma = \sqrt{2,3} = 1,5657$, dengan demikian statistik ujinya dapat dihitung sebagai berikut :

$$z = \frac{16,9 - 16}{1,5657 / \sqrt{20}} = 2,65$$

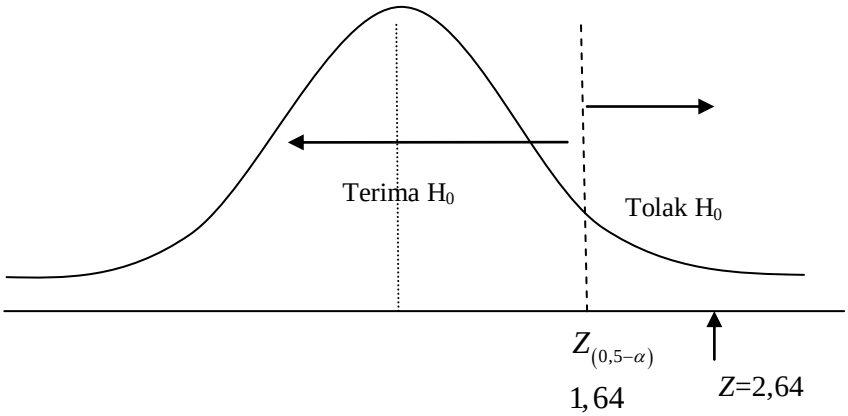
Oleh karena, pengujian hipotesis yang dilakukan adalah pengujian hipotesis pihak kanan, maka nilai kritis yang diperoleh dari Tabel Distribusi Normal Standard adalah :

$Z_{(0,5-\alpha)} = 1,64$

Kriteria Uji :

Terima H_O jika $Z < Z_{(0,5-\alpha)}$, Tolak dalam hal lainnya

Untuk memudahkan dalam mengambil kesimpulan, gambarkan kurva daerah penolakan dan penerimaan H_0 sebagai berikut :



Berdasarkan nilai statistik uji yang diperoleh dan kriteria uji yang digunakan, dapat dilihat pada kurva bahwa nilai statistik uji (Z) yang bernilai 2,64 berada di daerah penolakan H_0 sehingga dapat disimpulkan bahwa H_0 ditolak. Dengan kata lain, tidak terdapat cukup bukti untuk menolak pernyataan bahwa metode baru menghasilkan rata-rata per jam lebih dari 16 unit atau pengusaha tersebut dapat menggunakan metode baru untuk menggantikan metode lama.

Perhatikan pengujian hipotesis pihak kiri pada kasus berikut, sebuah iklan menyatakan bahwa sebuah baju merk A yang dihasilkan oleh sebuah pabrik cukup awet untuk dipakai. Iklan ini dibuat pengusaha berdasarkan kenyataan bahwa baju tersebut dapat digunakan dalam keadaan masih baik paling singkat selama tempo 180 hari. Untuk meneliti pernyataan yang dibuat dalam iklan itu telah diuji sebanyak 36 buah baju. Ketiga puluh enam baju itu digunakan orang (dalam kondisi baru) hingga mulai rusak. Jika dari ke 36 baju itu diperoleh rata-rata mulai rusak pada hari ke-174. Berdasarkan pengalaman diketahui bahwa simpangan baku pemakaian baju tersebut adalah 10 hari. Dengan menggunakan taraf signifikansi 5%, tentukan apakah penelitian yang dilakukan berhasil memperlihatkan bahwa baju itu cukup awet sesuai dengan yang dinyatakan dalam iklan atau tidak.

Hipotesis yang diuji adalah sebagai berikut :

H_0 : $\mu \geq 180$ (Baju merk A dapat digunakan dalam keadaan masih baik paling singkat selama tempo 180 hari)

H_1 : $\mu < 180$ (Baju merk A dapat digunakan dalam keadaan masih baik kurang dari tempo 180 hari)

$\alpha = 0,05$

Dari uraian kasus diperoleh informasi mengenai ukuran sampel (n) yaitu 36 baju, rata-rata masa mulai rusak ($\bar{x}$) hari ke-174 dan simpangan baku dari populasi baju (σ) yaitu 10 hari, dengan demikian statistik ujinya dapat dihitung sebagai berikut :

$$z = \frac{174 - 180}{10/\sqrt{36}} = -3,6$$

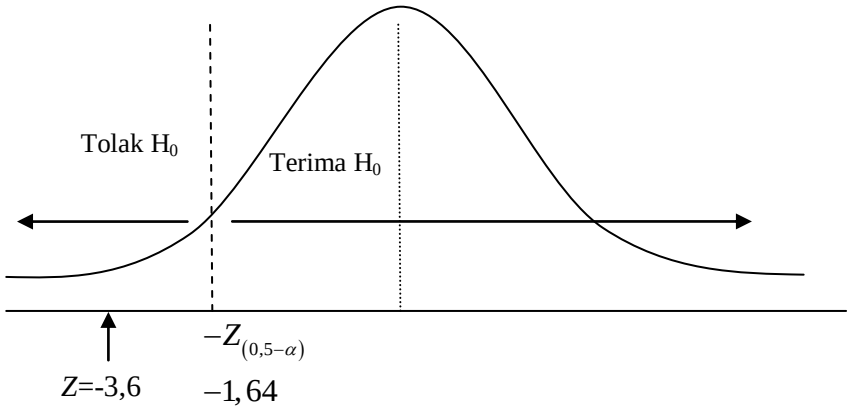
Oleh karena, pengujian hipotesis yang dilakukan adalah pengujian hipotesis pihak kiri, maka nilai kritis yang diperoleh dari Tabel Distribusi Normal Standard adalah :

$Z_{(0,5-\alpha)} = 1,64$

Kriteria Uji :

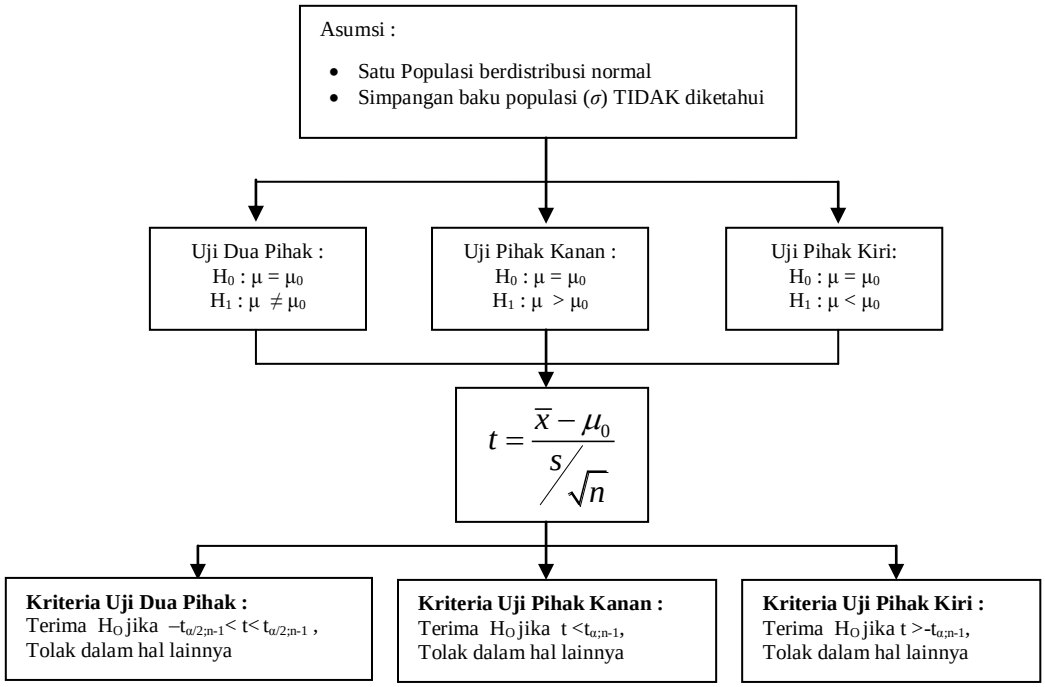
Terima H_O jika $Z > -Z_{(0,5-\alpha)}$, Tolak dalam hal lainnya

Untuk memudahkan dalam mengambil kesimpulan, gambarkan kurva daerah penolakan dan penerimaan H_0 sebagai berikut :



Berdasarkan nilai statistik uji yang diperoleh dan kriteria uji yang digunakan, dapat dilihat pada kurva bahwa nilai statistik uji (Z) yang bernilai -,36 berada di daerah penolakan H_0 sehingga dapat disimpulkan bahwa H_0 ditolak. Dengan kata lain, Baju merk A tidak cukup awet sesuai dengan yang dinyatakan dalam iklam atau pernyataan Baju merk A dapat digunakan dalam keadaan masih baik paling singkat selama tempo 180 hari dalam iklan tidak dapat diterima.

Ketiga contoh kasus di atas adalah contoh pengujian hipotesis dimana simpangan baku dari populasi (σ) diketahui. Pada kenyataannya, dalam sebuah penelitian, informasi mengenai simpangan baku populasi tidak diketahui. Dengan demikian, asumsi yang harus dipenuhi dalam menggunakan statistik uji Z tidak terpenuhi. Sebagai alternatifnya, untuk dapat menguji hipotesis dimana simpangan baku populasinya tidak diketahui, digunakan statistik uji t sebagai berikut :



Sama seperti kasus pertama pada pengujian hipotesis sebelumnya, yang berbeda adalah simpangan baku yang diketahui adalah simpangan baku dari sampel. Pengusaha lampu pijar A mengatakan bahwa lampunya bisa tahan pakai sekitar 800 jam. Akhir-akhir ini timbul dugaan bahwa masa pakai lampu itu telah berubah. Untuk menentukan hal ini, dilakukan penelitian dengan cara menguji 50 buah lampu. Ternyata rata-ratanya 792 jam dan simpangan baku nya 55 jam. Selidikilah dengan taraf signifikansi 5%, apakah kualitas lampu itu sudah berubah atau belum?

Berdasarkan kasus tersebut, maka hipotesis yang diuji adalah sebagai berikut :

H_0 : μ=800 (Rata-rata masa pakai lampu adalah 800 jam)

H_1 : $\mu \neq 800$ (Rata-rata masa pakai lampu bukan 800 jam)

$\alpha = 0{,}05$

Dari uraian kasus diperoleh informasi mengenai ukuran sampel (n) yaitu 50 buah lampu, rata-rata masa pakai dari kelima puluh lampu ($\bar{x}$) yaitu 792 jam dan simpangan bakunya (s) yaitu 55 jam, dengan demikian statistik ujinya dapat dihitung sebagai berikut :

$$t = \frac{792 - 800}{55/\sqrt{50}} = -1{,}029$$

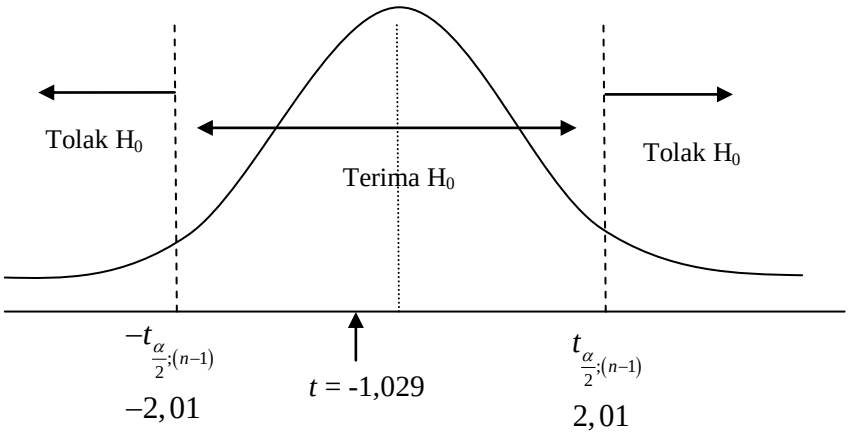
Oleh karena, pengujian hipotesis yang dilakukan adalah pengujian hipotesis dua pihak, maka nilai kritis yang diperoleh dari Tabel Distribusi *t-Student* adalah :

$t_{\frac{\alpha}{2};(n-1)} = t_{0,025;49} = 2,01$

Kriteria Uji :

Terima H_O jika $-t_{\alpha/2;n-1} < t < t_{\alpha/2;n-1}$, Tolak dalam hal lainnya

Untuk memudahkan dalam mengambil kesimpulan, gambarkan kurva daerah penolakan dan penerimaan H_0 sebagai berikut :



Berdasarkan nilai statistik uji yang diperoleh dan kriteria uji yang digunakan, dapat dilihat pada kurva bahwa nilai statistik uji (t) yang bernilai -1,029 berada di daerah penerimaan H_0 sehingga dapat disimpulkan bahwa H_0 diterima. Dengan kata lain, kualitas lampu pijar A belum berubah atau rata-rata masa pakainya masih 800 jam.

Perhatikan contoh kasus pengujian pihak kanan berikut, dikatakan bahwa dengan menyuntikkan semacam hormon tertentu kepada ayam akan menambah berat telurnya rata-rata 4,5 gram. Sampel acak yang terdiri atas 31 butir telur dari ayam yang telah diberi suntikan hormon tersebut memberikan rata-rata 4,9 gram dan simpangan baku 0,8 gram. Cukup beralasankah untuk menerima pernyataan bahwa pertambhan rata-rata berta telur lebih dari 4,5 gram?

Hipotesis yang diuji adalah :

H_0 : μ=4,5 (Pertambahan rata-rata berat telur adalah 4,5 gram)

H_1 : μ>4,5 (Pertambahan rata-rata berat telur lebih dari 4,5 gram)

$\alpha = 0,05$

Dari uraian kasus diperoleh informasi mengenai ukuran sampel (n) yaitu 31, rata-rata pertambahan berat telur yang dihasilkan ($\bar{x}$) yaitu 4,9 gram dan simpangan bakunya (s) adalah 0,8 gram. demikian statistik ujinya dapat dihitung sebagai berikut :

$$t = \frac{4,9 - 4,5}{0,8/\sqrt{31}} = 2,78$$

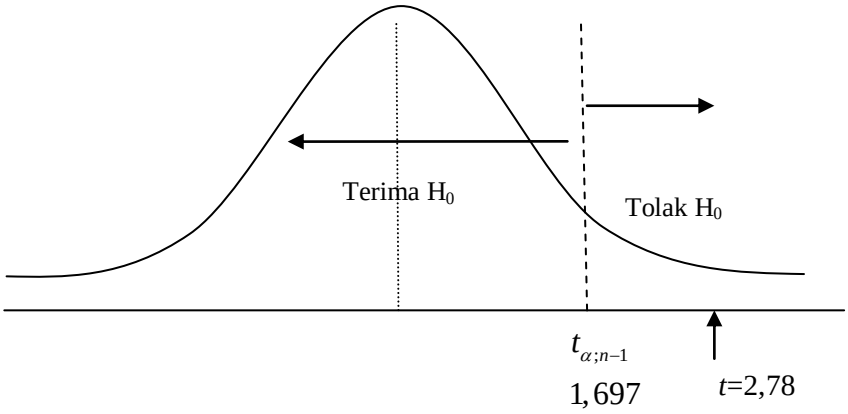
Oleh karena, pengujian hipotesis yang dilakukan adalah pengujian hipotesis pihak kanan, maka nilai kritis yang diperoleh dari Tabel Distribusi *t-Student* adalah :

$t_{\alpha;n-1} = t_{0,05;30} = 1,697$

Kriteria Uji :

Terima H_O jika $t < t_{\alpha;n-1}$, Tolak dalam hal lainnya

Untuk memudahkan dalam mengambil kesimpulan, gambarkan kurva daerah penolakan dan penerimaan H_0 sebagai berikut :



Berdasarkan nilai statistik uji yang diperoleh dan kriteria uji yang digunakan, dapat dilihat pada kurva bahwa nilai statistik uji (t) yang bernilai 2,78 berada di daerah penolakan H_0 sehingga dapat disimpulkan bahwa H_0 ditolak. Dengan kata lain, cukup beralasan untuk menerima pernyataan bahwa pertambahan rata-rata berat telur lebih besar dari 4,5 gram.

Contoh kasus berikut adalah pengujian hipotesis pihak kiri, akhir-akhir ini masyarakat mengeluh dan mengatakan bahwa isi bersih makanan A dalam kaleng tidak sesuai dengan yang tertulis pada etiketnya sebesar 5 ons. Untuk meneliti hal ini, 23 kaleng makanan A telah diteliti secara acak. Dari ke-23 isi kaleng tersebut, berat rata-ratanya 4,9 ons dan simpangan baku 0,2 ons. Dengan taraf signifikansi 5%, tentukan apa yang dapat anda katakan tentang keluhan masyarakat tersebut?

Hipotesis yang diuji adalah sebagai berikut :

H_0 : $\mu = 5$ (Isi bersih makanan A dalam kaleng sesuai dengan etiketnya yaitu 5 ons)

H_1 : $\mu < 5$ (Isi bersih makanan A dalam kaleng tidak sesuai dengan etiketnya yaitu kurang dari 5 ons)

$\alpha = 0{,}05$

Dari uraian kasus diperoleh informasi mengenai ukuran sampel (n) yaitu 23 kaleng, rata-rata isi bersih dalam kaleng ($\bar{x}$) adalah 4,9 dan simpangan bakunya (s) yaitu 0,2, dengan demikian statistik ujinya dapat dihitung sebagai berikut :

$$t = \frac{4{,}9 - 5}{0{,}2 / \sqrt{23}} = -2{,}398$$

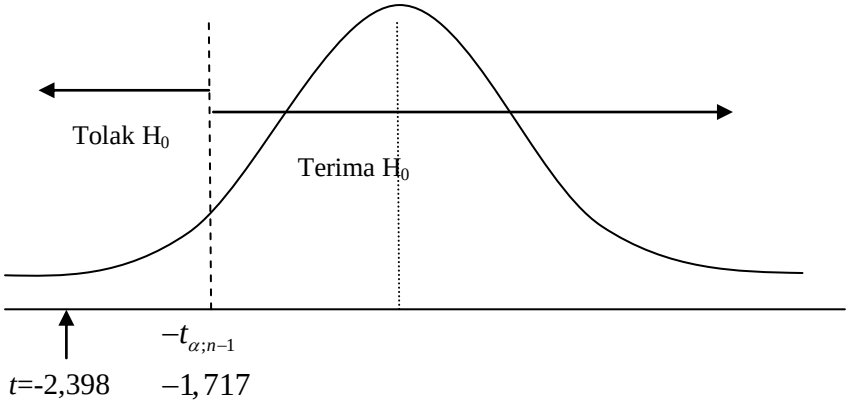
Oleh karena, pengujian hipotesis yang dilakukan adalah pengujian hipotesis pihak kiri, maka nilai kritis yang diperoleh dari Tabel Distribusi *t-Student* adalah :

$t_{\alpha;n-1} = t_{0{,}05;22} = 1{,}717$

Kriteria Uji :

Terima H_O jika t >-$t_{\alpha;n-1}$, Tolak dalam hal lainnya

Untuk memudahkan dalam mengambil kesimpulan, gambarkan kurva daerah penolakan dan penerimaan H_0 sebagai berikut :

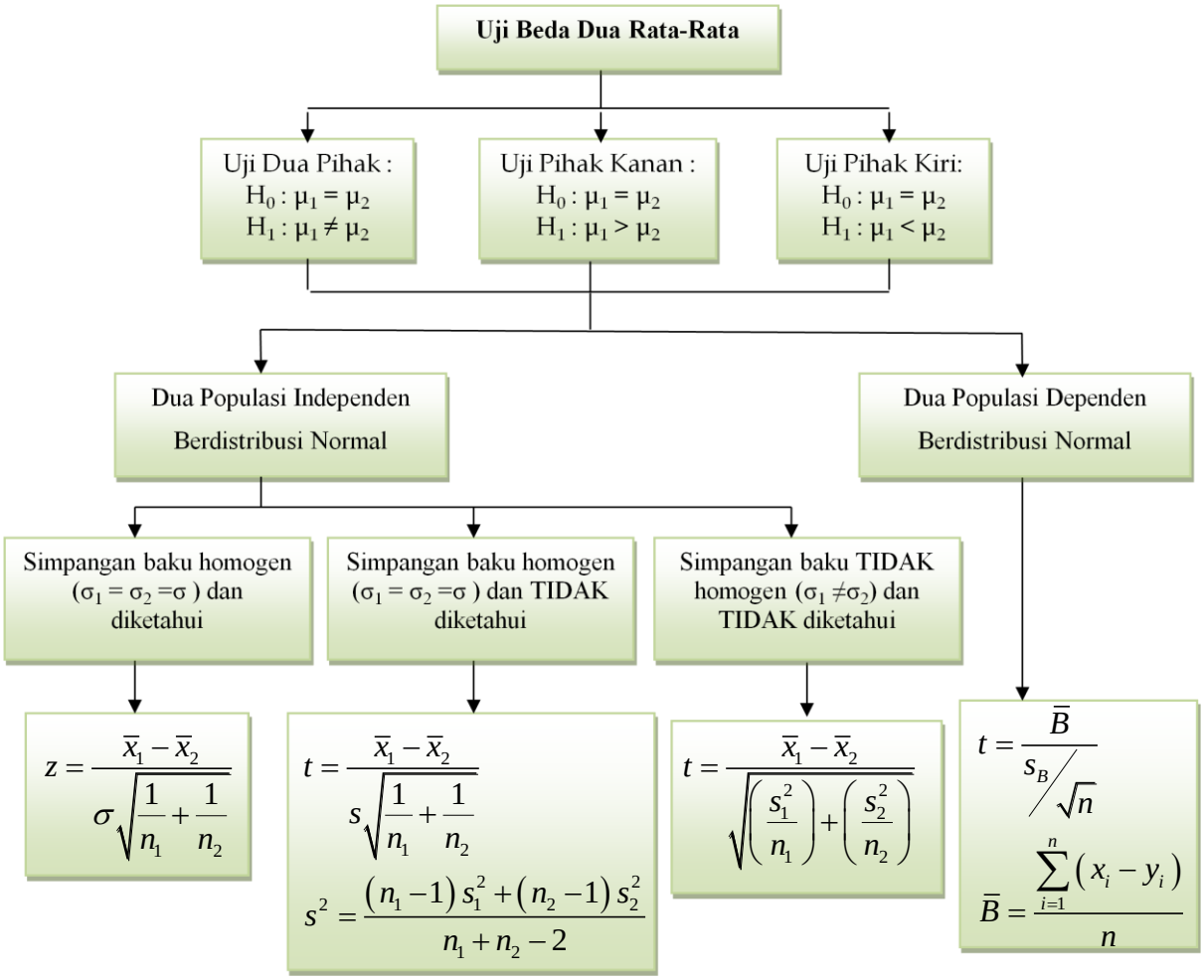


Berdasarkan nilai statistik uji yang diperoleh dan kriteria uji yang digunakan, dapat dilihat pada kurva bahwa nilai statistik uji (t) yang bernilai -2,398 berada di daerah penolakan H_0 sehingga dapat disimpulkan bahwa H_0 ditolak. Dengan kata lain, hasil penelitian dapat digunakan sebagai bahan pendukung untuk menguatkan keluhan masyarakat bahwa isi bersih makanan dalam kaleng sudah berkurang daripada yang tertera pada etiket.

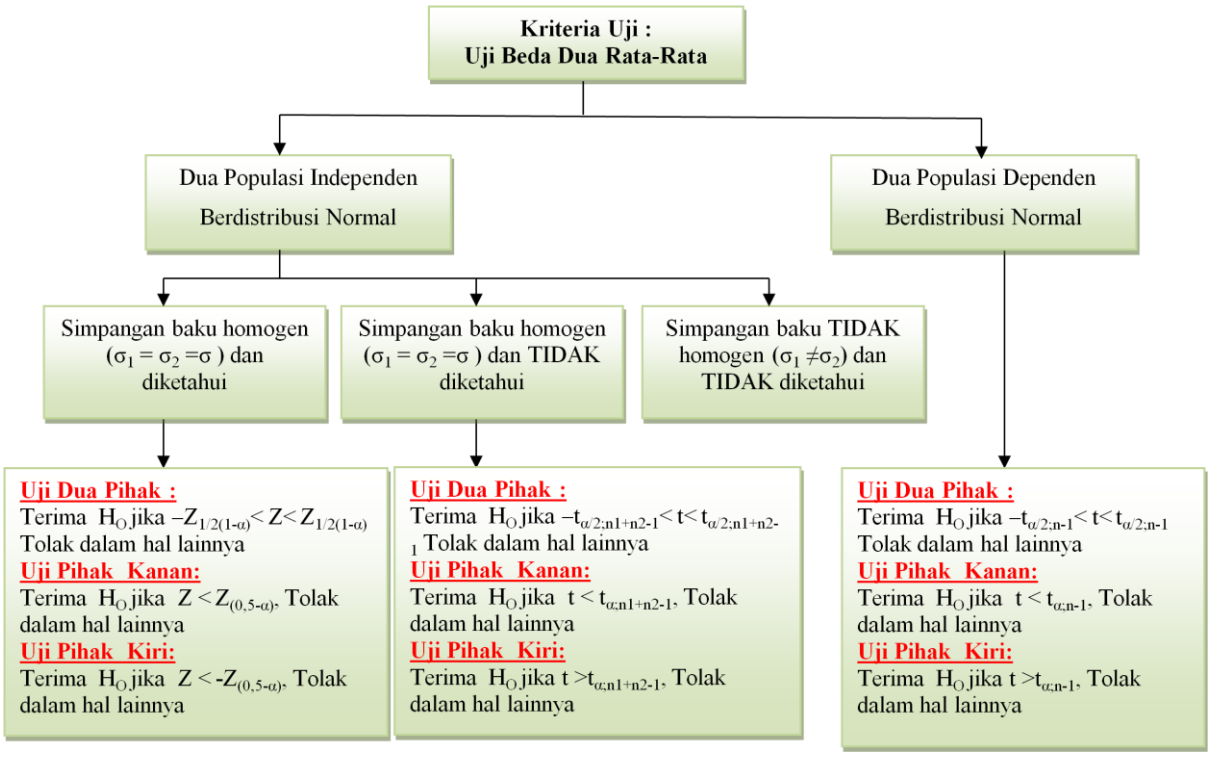
b) Uji Rata-Rata Dua Sampel

Sebuah penelitian tidak selalu hanya melibatkan pengujian hipotesis untuk satu sampel,akan tetapi, bisa dua atau lebih. Sama seperti pengujian hipotesis rata-rata satu

sampel, statistik uji pada pengujian rata-rata dua sampel dibedakan berdasarkan asumsi yang harus dipenuhi, sebagai berikut :



Kriteria uji yang digunakan untuk masing-masing pengujian hipotesis adalah sebagai berikut:



Sedangkan untuk asumsi Simpangan baku tidak homogen ($\sigma_1 \neq \sigma_2$) dan tidak diketahui, kriteria uji yang digunakan yaitu :

Uji Dua Pihak :

Terima H_O jika

$$-\frac{\left[\left(s_1^2/n_1\right)t_{\alpha/2,(n_1-1)} + \left(s_2^2/n_2\right)t_{\alpha/2,(n_2-1)}\right]}{\left(s_1^2/n_1\right)+\left(s_2^2/n_2\right)} < t < \frac{\left(s_1^2/n_1\right)t_{\alpha/2,(n_1-1)} + \left(s_2^2/n_2\right)t_{\alpha/2,(n_2-1)}}{\left(s_1^2/n_1\right)+\left(s_2^2/n_2\right)}$$

Terima dalam hal lainnya.

Uji Pihak Kanan:

Tolak H_O jika

$$t \geq \frac{\left(s_1^2/n_1\right)t_{\alpha,(n_1-1)} + \left(s_2^2/n_2\right)t_{\alpha,(n_2-1)}}{\left(s_1^2/n_1\right)+\left(s_2^2/n_2\right)}$$

Terima dalam hal lainnya.

Uji Pihak Kiri:

Tolak H_O jika

$$t \leq -\frac{\left[\left(s_1^2/n_1\right)t_{\alpha,(n_1-1)} + \left(s_2^2/n_2\right)t_{\alpha,(n_2-1)}\right]}{\left(s_1^2/n_1\right)+\left(s_2^2/n_2\right)}$$

Terima dalam hal lainnya.

Perhatikan contoh kasus berikut, semacam barang dihasilkan dengan menggunakan dua proses. Ingin diketahui apakah kedua proses itu menghasilkan hal yang

sama atau tidak terhadap kualitas barang itu ditinjau dari rata-rata daya tekannya. Untuk itu diadakan percobaan sebanyak 20 hari dari hasil proses kesatu dan 20 pula dari hasil proses kedua. Rata-rata dan simpangan baku hasil dari proses satu adalah 9,25 kg dan 2,24 kg. Sedangkan rata-rata dan simpangan baku dari hasil proses kedua adalah 10,40 kg dan 3,12 kg. Dengan menggunakan taraf signifikansi 5% bagaimanakah hasilnya?

Hipotesis yang diuji adalah sebagai berikut :

$H_0 : \mu_1 = \mu_2$ (kedua proses menghasilkan barang dengan rata-rata dan daya tekan yang sama)

$H_1 : \mu_1 \neq \mu_2$ (kedua proses menghasilkan barang dengan rata-rata dan daya tekan yang berbeda)

$\alpha = 0{,}05$

Dari uraian kasus diperoleh informasi sebagai berikut :

- Proses satu :

  Ukuran sampel (n_1) = 20

  Rata-rata ($\bar{x}_1$) = 9,25

  Simpangan baku (s_1) = 2,24

- Proses dua:

  Ukuran sampel (n_2) = 20

  Rata-rata ($\bar{x}_2$) = 10,40

  Simpangan baku (s_2) = 3,12

Dengan demikian, statistik ujinya dapat dihitung sebagai berikut :

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{s_1^2}{n_1}\right) + \left(\frac{s_2^2}{n_2}\right)}}$$

$$= \frac{9,25 - 10,40}{\sqrt{\left(\frac{2,24^2}{20}\right) + \left(\frac{3,12^2}{20}\right)}} = 1,339$$

Oleh karena, pengujian hipotesis yang dilakukan adalah pengujian hipotesis dua pihak, maka nilai kritis nya dihitung sebagai berikut :

$$\frac{\left(s_1^2/n_1\right)t_{\alpha/2,(n_1-1)}+\left(s_2^2/n_2\right)t_{\alpha/2,(n_2-1)}}{\left(s_1^2/n_1\right)+\left(s_2^2/n_2\right)}=\frac{\left(\frac{2,24^2}{20}\right)t_{0,05/2,(20-1)}+\left(\frac{3,12^2}{20}\right)t_{0,05/2,(20-1)}}{\left(\frac{2,24^2}{20}\right)+\left(\frac{3,12^2}{20}\right)}$$

$$=\frac{\left(\frac{2,24^2}{20}\right)2,09+\left(\frac{3,12^2}{20}\right)2,09}{\left(\frac{2,24^2}{20}\right)+\left(\frac{3,12^2}{20}\right)}$$

$$=2,09$$

Sehingga wilayah penerimaan H_0 adalah :

$$-\frac{\left[\left(s_1^2/n_1\right)t_{\alpha/2,(n_1-1)}+\left(s_2^2/n_2\right)t_{\alpha/2,(n_2-1)}\right]}{\left(s_1^2/n_1\right)+\left(s_2^2/n_2\right)}<t<\frac{\left(s_1^2/n_1\right)t_{\alpha/2,(n_1-1)}+\left(s_2^2/n_2\right)t_{\alpha/2,(n_2-1)}}{\left(s_1^2/n_1\right)+\left(s_2^2/n_2\right)}$$

$$-2,09<t<2,09$$

Kriteria Uji :

Terima H_O $-2,09<t<2,09$, Tolak dalam hal lainnya

Berdasarkan nilai statistik uji yang diperoleh dan kriteria uji yang digunakan, statistik uji *t* yang diperoleh adalah 1,339 dan berada di daerah penerimaan H_0 sehingga dapat disimpulkan bahwa H_0 diterima. Dengan kata lain, kedua proses menghasilkan rata-rata dan daya tekan yang sama.

Perhatikan contoh kasus kedua berikut, seorang ibu percaya bahwa les privat dapat meningkatkan nilai ujian anaknya. Untuk itu, dilakukan pengamatan terhadap 10 mata pelajaran yang diikuti oleh anaknya sebelum dan sesudah mengikuti les privat tersebut. Hasilnya sebagai berikut :

Nilai Sebelum	30	21	21	27	20	25	27	22	28	18
Nilai Sesudah	31	22	37	24	30	15	25	42	19	38

Apakah yang dapat disimpulkan dari hasil ujian?

Hipotesis yang diuji adalah sebagai berikut :

H_0 : μ_B =0 (les privat tidak dapat meningkatkan nilai ujian)

H_1 : μ_B <0 (les privat dapat meningkatkan nilai ujian)

$\alpha = 0{,}05$

Untuk dapat menghitung nilai statistik uji, terlebih dahulu dilakukan perhitungan nilai selisih antara nilai sebelum dan nilai sesudah mengikuti les privat sebagai berikut :

Nilai Sebelum	30	21	21	27	20	25	27	22	28	18
Nilai Sesudah	31	22	37	24	30	15	25	42	19	38
B	-1	-1	-16	3	-10	10	2	-20	9	-20

Rata-rata dan simpangan baku dari nilai selisih yang diperoleh adalah sebagai berikut :

$$\bar{B} = \frac{\sum_{i=1}^{n}(x_i - y_i)}{n}$$
$$= \frac{(-1)+(-1)+(-16)+3+(-10)+2+(-20)+9+(-20)}{10}$$
$$= -4,4$$
$$S_B = \sqrt{\frac{\sum_{i=1}^{n}\left((x_i - y_i)-\bar{B}\right)^2}{n-1}} = 11,35$$

Dengan demikian statistik ujinya adalah :

$$t = \frac{\bar{B}}{s_B/\sqrt{n}} = \frac{-4,4}{11,35/\sqrt{10}} = -1,227$$

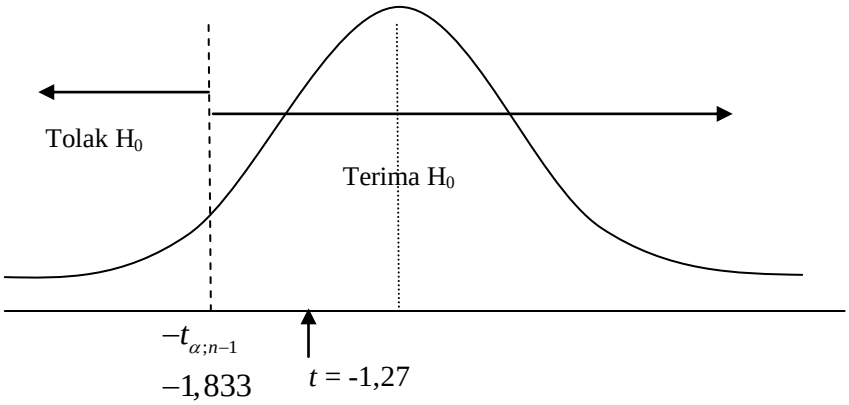
Pengujian hipotesis yang dilakukan adalah pengujian dua sampel dependen pihak kiri, sehingga nilai kritisnya diperoleh melalui tabel distribusi *t-Student* sebagai berikut :

$t_{\alpha;n-1} = t_{0,05;9} = 1,833$

Kriteria uji :

Terima H_O jika $t > t_{\alpha;n-1}$, Tolak dalam hal lainnya

Untuk memudahkan dalam mengambil kesimpulan, gambarkan kurva daerah penolakan dan penerimaan H_0 sebagai berikut :

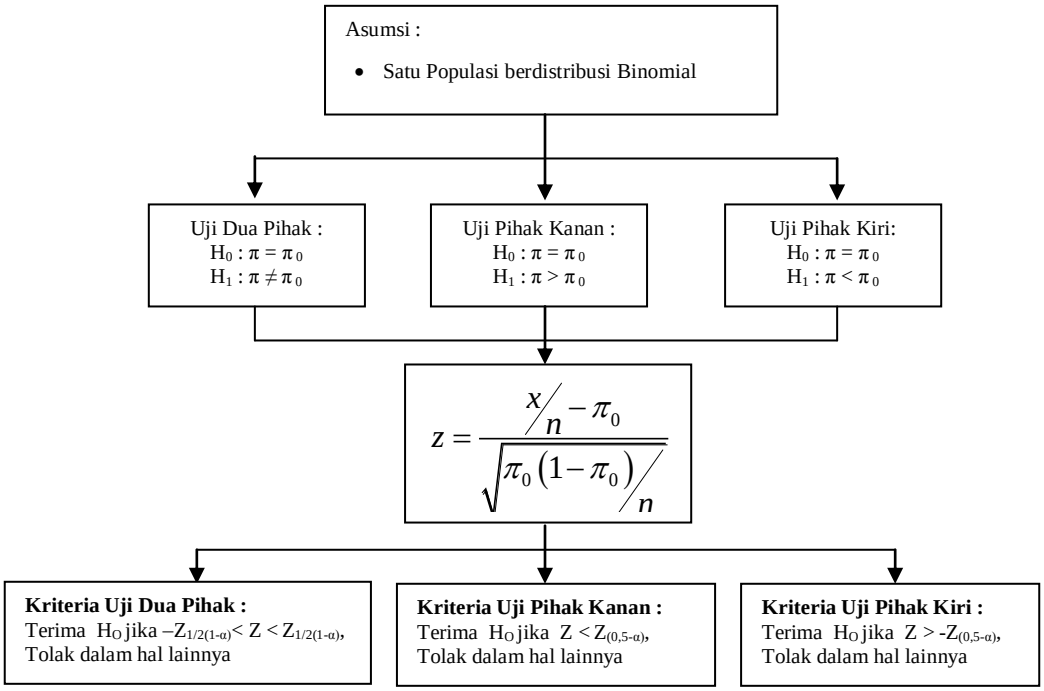


Berdasarkan nilai statistik uji yang diperoleh dan kriteria uji yang digunakan, dapat dilihat pada kurva bahwa nilai statistik uji (t) yang bernilai -1,27 berada di daerah penerimaan H_0 sehingga dapat disimpulkan bahwa H_0 diterima. Dengan kata lain, les privat tidak dapat meningkatkan nilai ujian, artinya nilai rata-rata nilai ujian sebelum dan sesudah mengikuti les privat relatif sama.

6.4. Uji Proporsi

Proporsi, persentase dan rasio pada dasarnya sama yaitu merupakan hasil bagi dua bilangan, akan tetapi pengertiannya berbeda. Proporsi adalah pecahan yang pembilangnya merupakan bagian dari penyebutnya. Proporsi artinya jumlah/frekuensi dari suatu sifat tertentu dibandingkan dengan seluruh populasi dimana sifat tersebut didapatkan. Proporsi digunakan untuk melihat komposisi suatu variabel dalam populasi. Bentuk ini sering dinyatakan dalam persen, yaitu dengan mengalikan pecahan ini dengan 100%. Sedangkan rasio adalah perbandingan antara dua besaran atau lebih. Dalam menghitung rasio harus menggunakan satuan yang sama, apabila terdapat perbedaan maka harus dilakukan penyamaan satuan terlebih dahulu. Rasio dinotasikan sebagai a/b atau $a : b$, dimana $b \neq 0$.

Dalam statisika, proporsi sebuah variabel dalam sebuah populasi dinotasikan sebagai π dan berdistribusi binomial. Statistik uji yang digunakan dalam pengujian hipotesis proporsi untuk satu populasi adalah sebagai berikut :



Perhatikan contoh kasus berikut, misalkan sebuah penelitian ingin menguji bahwa distribusi jenis kelamin laki-laki dan jenis kelamin perempuan adalah sama. Sebuah sampel acak berukutan 4800 orang terdiri atas 2458 orang laki-laki. Dalam taraf nyata 5% , betulkan distribusi kedua jenis kelamin itu sama?

Hipotesis yang diuji adalah sebagai berikut :

$H_0 : \pi = 0,5$

$H_1 : \pi \neq 0,5$

$\alpha = 0,05$

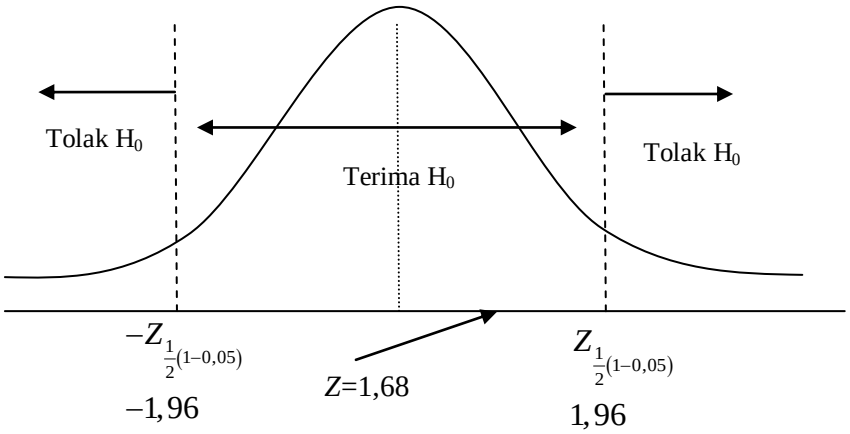
Statistik Uji :

$$z = \frac{x/n - \pi_0}{\sqrt{\pi_0(1-\pi_0)/n}}$$

$$z = \frac{2458/4800 - 0,5}{\sqrt{0,5(1-0,5)/4800}} = 1,68$$

Nilai kritis :

$Z_{1/2(1-0,05)} = 1,96$



Nilai *Z* yang diperoleh berada dalam wilayah penerimaan H_0 sehingga dapat disimpulkan bahwa H_0 diterima, artinya proporsi antara laki-laki dan perempuan sama besar.

Contoh berikutnya adalah seorang pejabat mengatakan bahwa paling banyak 60% anggota masyarakat termasuk golongan A. Sebuah sampel acak telah diambil yang terdiri atas 8500 orang ternyata 5426 termasuk golongan A. Apabila taraf signifikansi yang digunakan 1%, dapat diterimakah pernyataan tersebut?.

Hipotesis yang diuji yaitu :

$H_0 : \pi \leq 0,6$

$H_1 : \pi > 0,6$

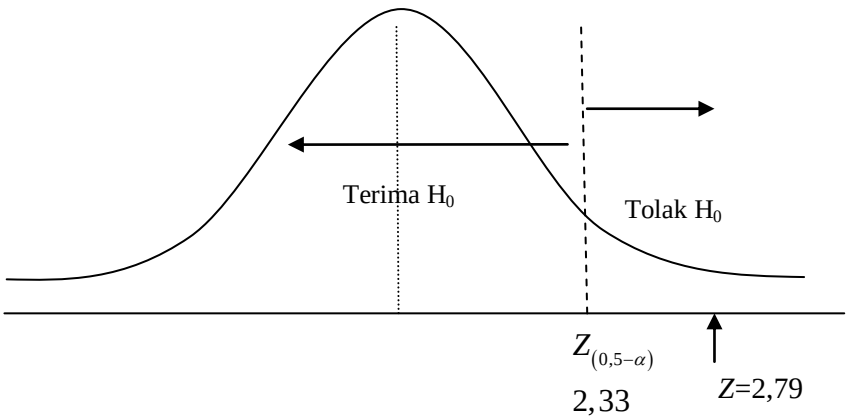
$\alpha = 0,05$

Statistik Uji :

$$z = \frac{5426/8500 - 0,6}{\sqrt{0,6(1-0,6)/8500}} = 2,79$$

Nilai Kritis :

$Z_{(0,05-0,01)} = 2,33$



Nilai *Z* yang diperoleh berada dalam wilayah penolakan H_0 sehingga dapat disimpulkan bahwa H_0 ditolak, artinya persentase anggota masyarakat yang termasuk golongan A lebih besar dari 60%.

Contoh berikut adalah kasus untuk pengujian pihak kiri. Seorang prodeusen menyatakan bahwa paling sedikit 95% dari barang yang dihasilkan tergolong pada kualitas yang memuaskan. Pemeriksaan terhadap 200 barang yang ia hasilkan, ternyata 18 diantaranya tidak disenangi oleh konsumen karena cacat. Selidikilah pernyataan produsen tadi dengan taraf nyata 1%.

Hipotesis yang diuji adalah sebagai berikut :

$H_0 : \pi \geq 0,95$

$H_1 : \pi < 0,95$

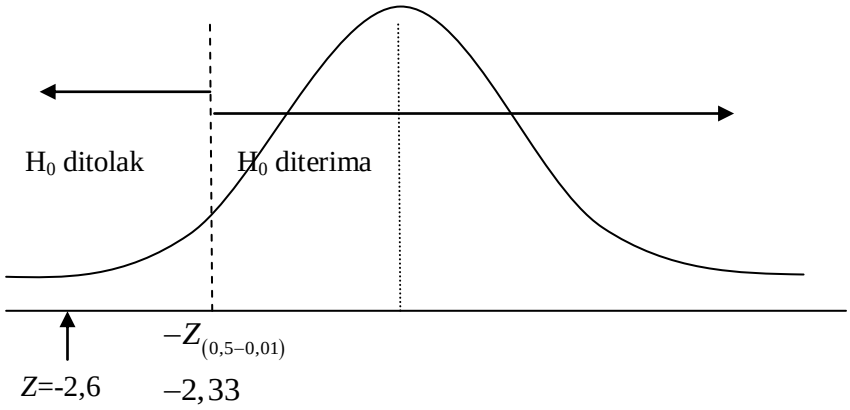
$\alpha = 0,01$

Statistik Uji :

$$z = \frac{182/200 - 0,95}{\sqrt{0,95(1-0,95)/200}} = -2,6$$

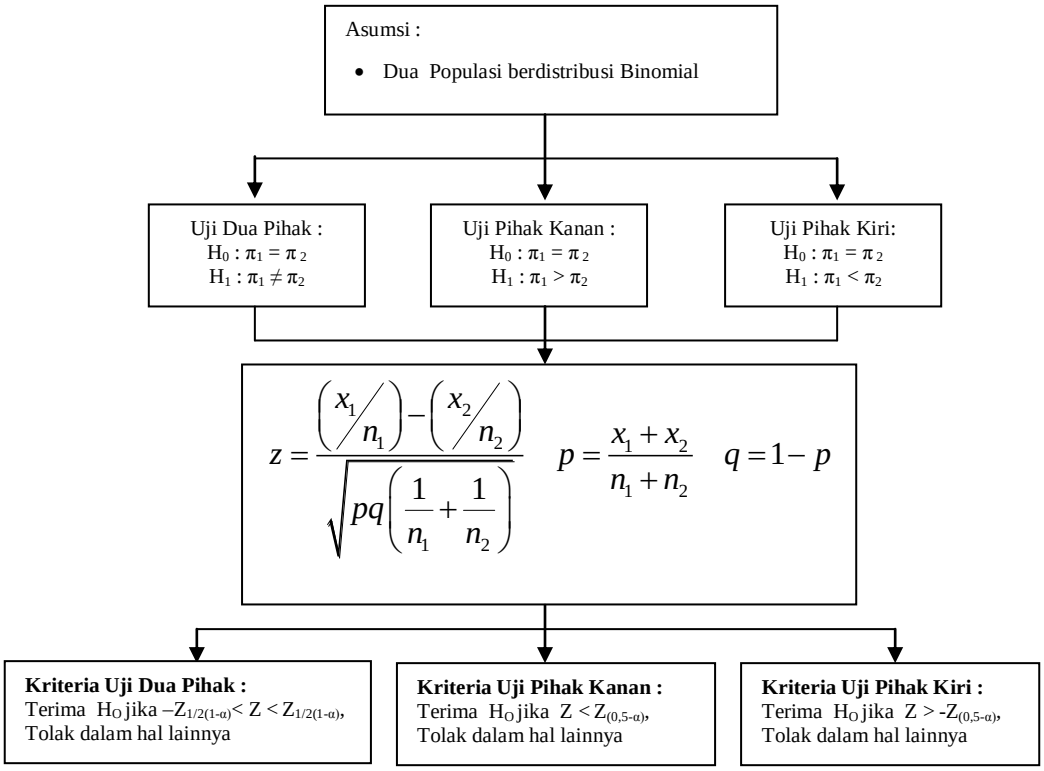
Nilai Kritis :

$Z_{(0,5-0,01)} = 2,33$



Nilai *Z* yang diperoleh berada dalam wilayah penolakan H_0 sehingga dapat disimpulkan bahwa H_0 ditolak, artinya kurang dari 95% dari barang yang dihasilkan tergolong pada kualitas yang memuaskan.

Sama halnya dengan uji rata-rata, dalam pengujian proporsi juga terkadang melibatkan dua populasi. Dalam hal ini, kedua populasi diasumsikan berdistribusi Binomial, sebagai berikut :



Perhatikan contoh kasus berikut, suatu penelitian dilakukan di daerah A terhadao 250 pemilih. Ternyata 150 pemilih menyatakan akan memilih calon C. Di daerah B penelitian dilakukan terhadap 300 pemilih dan terdapat 162 yang akan memilih calon C. Adakah perbedaan yang nyata mengenai pemilihan calon C di antara kedua daerah itu?

Hipotesis yang diuji :

$H_0 : \pi_1 = \pi_2$

$H_1 : \pi_1 \neq \pi_2$

$\alpha = 0{,}05$

Statistik Uji :

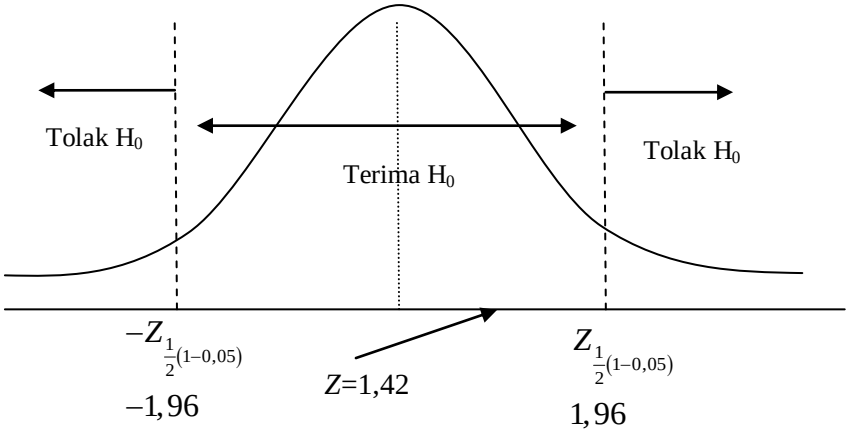
$$p = \frac{150+162}{250+300} = 0{,}5673 \quad q = 1 - 0{,}5673 = 0{,}4327$$

$$z = \frac{\left(150/250\right) - \left(162/300\right)}{\sqrt{(0{,}5673 \times 0{,}4327)\left(\frac{1}{250} + \frac{1}{300}\right)}} = 1{,}42$$

$$z = \frac{\left(x_1/n_1\right) - \left(x_2/n_2\right)}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad p = \frac{x_1 + x_2}{n_1 + n_2} \quad q = 1 - p$$

Nilai Kritis :

$Z_{1/2(1-0,05)} = 1,96$



Nilai *Z* yang diperoleh berada dalam wilayah penerimaan H_0 sehingga dapat disimpulkan bahwa H_0 diterima, artinya tidak terdapat perbedaan yang nyata antara kedua daerah itu dalam pemilihan calon C.

Selanjutnya, misalkan terdapat dua kelompok yaitu kelompok A dan kelompok B. Masing-masing terdiri dari 100 pasien yang menderita semacam penyakit. Kepada kelompok A diberikan serum tertentu tetapi tidak kepada kelompok B. Setelah jangka waktu tertentu, terdapat 80 yang sembuh dari kelompok A dan 68 dari kelompok B. Apakah penelitian ini memperlihatkan serum ikut membantu menyembuhkan penyakit?

Hipotesis yang diuji :

$H_0 : \pi_A = \pi_B$

$H_1 : \pi_A > \pi_B$

$\alpha = 0,05$

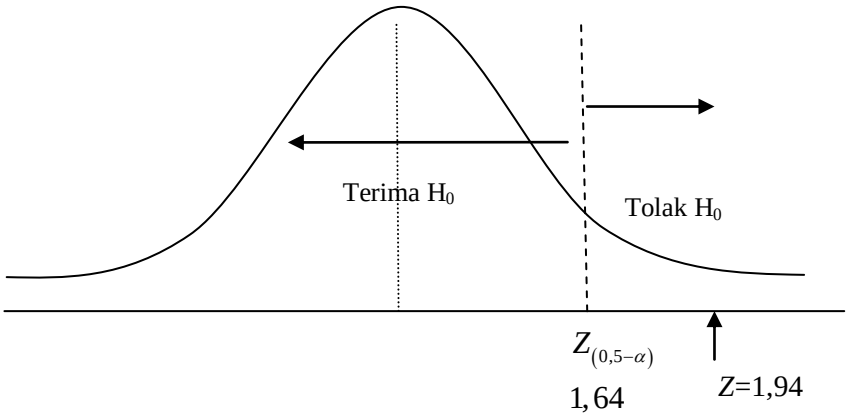
Statistik Uji :

$$p = \frac{80+68}{100+100} = 0,74 \quad q = 1-0,74 = 0,26$$

$$z = \frac{\left(80/100\right) - \left(68/100\right)}{\sqrt{\left(0,74 \times 0,26\right)\left(\frac{1}{100}+\frac{1}{100}\right)}} = 1,94$$

Nilai Kritis :

$Z_{(0,5-0,05)} = 1,64$



Nilai Z yang diperoleh berada dalam wilayah penolakan H_0 sehingga dapat disimpulkan bahwa H_0 ditolak, artinya pemberian serum dapat dikatakan membantu menyembuhkan penyakit.

LATIHAN

1. Berikanlah contoh untuk memperlihatkan bahwa menguji sebuah pernyataan memang diperlukan!
2. Uraikanlah secara singkat dan jelas, apa yang dimaksud dengan :
   a. Hipotesis
   b. Hipotesis nol
   c. Hipotesis alternatif
   d. Eror jenis I
   e. Eror Jenis II
   f. Taraf Signifikansi
   g. Nilai Kritis
3. Seorang pemilik pabrik rokok mempunyai anggapan bahwa rata-rata nikotin yang dikandung oleh setiap batang rokok adalah sebesar 20 *mg*, dengan alternatif lebih kecil dari itu. Dari 10 batang rokok yang dipilih secara acak, diperoleh hasil sebagai berikut :
   20 23 18 24 25 17 16 17 21 18
   Dengan menggunakan taraf signifikansi 5%, ujilah pendapat tersebut!
4. Seorang produsen menyatakanbahwa paling sedikit 95% dari barang yang dihasilkan tergolong pada kualitas yang memuaskan. pemeriksaan terhadap 200 barang yang ia hasilkan, ternyata 18 di antaranya tidak disenangi oleh konsumen karena cacat. Selidikilah pernyataan produsen tersebut dengan taraf signifikansi 1%!
5. Di sebuah pabrik telah dilakukan penyelesaian semacam proyek. Mula-mula dikerjakan 40 orang buruh dengan menggunakan metode A, kemudian dengan metode B telah dikerjakan oleh sebanyak 50 orang buruh. Waktu yang diperlukan dalam penyelesaian proyek tersebut, oleh metode A rata-rata setiap orang memerlukan waktu 55 menit, sedangkan apabila metode B yang digunakan rata-ratanya memerlukan waktu 58 menit. Jika simpangan bakunya masing-masing 5,5 menit dan 8 menit, tentukan apakah kedua metode tadi mempunyai perbedaan yang nyata ataukah tidak untuk penyelesaian proyek itu?
6. Suatu riset pemasaran dilakukan di Jakarta dan Surabaya. Tujuan dari riset ini adalah untuk mengetahui apakah perbedaan yang nyata antara ibu rumah tangga yang senang akan Rinso dibandingkan dengan Dino. Di Jakarta dari 100 orang ibu rumah tangga yang ditanya, terdapat 68 orang yang mengatakan lebih senang Rinso daripada Dino. Sementara di Surabaya, di antara 300 orang yang ditanya, terdapat 213 yang lebih senang Rinso daripada Dino. Dengan menggunakan taraf signifikansi 1% ujilah pendapat bahwa

proporsi ibu rumah tangga yang lebih senang Rinso daripada Dino di Surabaya dan di Jakarta adalah sama?

7. Ada pendapat bahwa tak ada perbedaan antara gaji bulanan bagi karyawan perusahaan A dan B yaitu sama, dengan alternatif ada perbedaan. Dari hasil wawancara terhadap 100 orang karyawan, (50 dari perusahaan a dan 50 dari perusahaan B), diketahui bahwa rata-rata gaji karyawan perusahaan a adalah Rp. 89.000 dengan simpangan baku rp. 40.000. Sedangkan rata-rata gaji karyawan B adalah Rp. 92.000 dengan simpangan baku Rp.30.000. Dengan menggunakan taraf signifikansi 5% ujilah pendapat tersebut?

Regresi Linear Sederhana dan Korelasi

7.1. Analisis Regresi

7.1.1. Definisi Analisis Regresi

Istilah regresi pertama kali diperkenalkan oleh Francis Galton (1886). Dalam sebuah makalah yang terkenal, Galton menemukan bahwa, walaupun terdapat sebuah kecenderungan bagi orang tua dengan tinggi badan yang tinggi memiliki anak-anak yang tinggi dan bagi orang tua dengan tinggi badan pendek memiliki anak-anak yang pendek, terdapat sebuah kecenderungan bahwa rata-rata tinggi badan dari anak-anak dari orang tua dengan tinggi badan tertentu bergerak atau *mundur* (*regress*) ke arah tinggi badan rata-rata populasi secara keseluruhan.

Hukum Regresi Semesta (*law of universal regression*) yang dikemukakan oleh Galton diperkuat oleh Karl Pearson, temannya. Pearson mengumpulkan lebih dari seribu catatan tinggi badan anggota kelompok keluarga. Dia menemukan bahwa rata-rata tinggi badan anak laki-laki pada kelompok ayah yang tinggi, lebih rendah atau kurang dari tinggi badan ayahnya. Sedangkan, rata-rata tinggi anak laki-laki pada kelompok ayah yang pendek, lebih besar atau lebih tinggi daripada tinggi badan ayahnya. Dengan demikian *mundurnya* (*regressing*) anak laki-laki yang tinggi maupun pendek serupa ke arah rata-rata tinggi badan semua laki-laki. Dengan kata lain, Galton menyebut hal tersebut sebagai "*regression to mediocrity*" atau "*kemunduran ke arah sedang*".

Definisi dari analisis regresi sendiri telah diungkapkan oleh beberapa penulis sebagai berikut :

- Menurut Gujarati (2004), Analisis regresi berkenaan dengan studi ketergantungan satu variable, *variable tak bebas*, pada satu atau lebih variable lain, *variabel yang menjelaskan (explanatory variables)*, dengan maksud menaksir dan atau meramalkan nilai-rata-rata hitung (*mean*) atau rata-rata (populasi) variabel tak bebas, dipandang dari segi nilai yang diketahui atau tetap (dalam pengambilan sampel berulang) variabel yang menjelaskan (yang belakangan).
- Menurut Sanford Wisberg (2005), Analisis regresi merupakan sebuah analisis yang digunakan untuk menjawab pertanyaan mengenai kebergantungan variabel respon (variabel tak bebas) pada satu atau lebih variabel predictor (variabel bebas/variabel menjelaskan/Explanatory Variable), termasuk prediksi nilai dari variabel respon,

menemukan variabel-variabel predictor yang penting dan memperkirakan dampak perubahan nilai dari sebuah variabel predictor atau sebuah perlakuan terhadap nilai variabel respon.

- Menurut Rudolf J. Freund (2003), Analisis regresi merupakan sebuah metode statistika yang digunakan untuk menganalisis hubungan antara dua atau lebih variabel sedemikian rupa sehingga salah satu variabel dapat diprediksi atau dijelaskan dengan menggunakan informasi dari variabel lainnya.
- Menurut Lukas Setia Atmaja (2009), Analisis regresi adalah suatu proses melakukan estimasi untuk memperoleh suatu hubungan fungsional antara variabel acak Y dengan variabel X

Dari keempat definisi di atas, jelas bahwa sebuah analisis regresi bertujuan untuk menaksir atau menduga nilai variabel *y* (variabel respon/variabel tak-bebas/varianel dependen.

7.1.2. Analisis Regresi dan Pengaruh

Pada sebuah penelitian sosial, sering terjadi kesalahfahaman dalam menerapkan analisis regresi. Dimana analisis regresi sering digunakan untuk menetapkan sebuah hubungan sebab-akibat. Suatu hubungan statistik, bagaimanapun kuat dan sugestif, tidak pernah dapat menetapkan hubungan sebab-akibat, gagasan kita mengenai sebab-akibat harus datang dari luar statistik, pada akhirnya dari beberapa teori atau lainnya (Kendall dan Stuart, dalam Gujarati, 2004).

Syarat untuk dapat berbicara pengaruh atau pola kausalitas atau hubungan sebab-akibat dalam sebuah analisis regresi, yaitu :

1. *Temporal order*
2. *Association*
3. *Eliminating alternatives*
4. Dukungan teori

Syarat pertama yang harus dipenuhi untuk dapat berbicara pengaruh dalam sebuah analisis regresi adalah adanya urutan kejadian yaitu terjadinya sebab mendahului terjadinya akibat yang disebut sebagai ***temporal order*.** Dengan kata lain, variabel tak bebas (respon/dependen) yang dinotasikan sebagai *Y* terjadi setelah variabel bebas (independen/penjelas/*explanatory*) yang dinotasikan sebagai variabel *X*, terjadi.

Misalnya, sebuah kepala keluarga ingin menaksir atau menduga besar pengeluaran rumah tangga per bulan berdasarkan penghasilan yang diterimanya. Dalam hal ini, pengeluaran rumah tangga adalah sebuah variabel tak bebas yaitu variabel *Y* jelas terjadi setelah kepala keluarga tersebut menerima penghasilan yaitu variabel bebas atau variabel *X*. Dengan demikian, terdapat urutan kejadian yaitu penghasilan terjadi mendahului terjadinya pengeluaran.

Selain adanya urutan kejadian, syarat kedua yang harus dipenuhi untuk dapat berbicara pengaruh dalam sebuah analisis regresi adalah adanya hubungan antara variabel *Y* dan variabel *X*. Salah satu cara untuk mengukur asosiasi antara variabel-variabel adalah menggunakan **koefisien korelasi**. Dikaitkan dengan kasus sebelumnya, maka untuk dapat berbicara pengaruh antara penghasilan dan pengeluaran, harus dipastikan terlebih dahulu bahwa terdapat hubungan antara penghasilan dan pengeluaran. Dengan kata lain, harus dihitung nilai koefisien korelasi antara penghasilan dan pengeluaran dan dilakukan pengujian signifikansi koefisien korelasi tersebut.

Syarat ketiga yang harus dipenuhi yaitu variabel akibat (variabel *Y*) disebabkan oleh variabel sebab (variabel *X*) dan bukan oleh variabel lain. Masih berkaitan dengan kasus sebelumnya, variabel pengeluaran terjadi disebabkan oleh variabel adanya variabel penghasilan bukan oleh variabel lainnya.

Hal paling penting untuk dapat berbicara pengaruh dalam sebuah analisis regresi adalah syarat keempat, yaitu dukungan teori. Artinya, untuk dapat mengatakan bahwa variabel *Y* dipengaruhi oleh variabel *X* harus didasarkan pada teori pendukung atau referensi yang menyatakan bahwa variabel *X* mempengaruhi variabel *Y*. Masih dikaitkan dengan kasus yang sama, untuk dapat mengatakan bahwa besar penghasilan mempengaruhi besar pengeluaran maka harus didasarkan pada referensei yang menyatakan bahwa variabel penghasilan (*X*) mempengaruhi variabel pengeluaran (*Y*), yaitu bisa diperoleh dalam ilmu ekonomi.

7.1.3. Analisis Regresi dan Korelasi

Berbicara regresi atau analisis regresi tidak dapat lepas dari korelasi. Akan tetapi terdapat perbedaan mendasar dari kedua analisis ini. Dalam sebuah analisis regresi, hubungan yang diamati adalah sebuah hubungan asimetris antara variabel *X* terhadap variabel *Y*. Dengan kata lain, hubungan yang diamati adalah hubungan satu arah, yaitu

variabel *X* mempengaruhi variabel *Y* (*X*→*Y*). Sementara dalam sebuah analisis korelasi, hubungan yang diamati adalah sebuah hubungan simetris antara variabel *X* dan variabel *Y*.

Sebagai contoh, digunakan kasus yang sama pada sub bab sebelumnya. Analisis regresi berbicara pengaruh dari variabel penghasilan terhadap variabel pengeluaran (satu arah), sedangkan analisis korelasi berbicara keeratan hubungan antara dua variabel tersebut, tidak berbicara apakah variabel penghasilan mempengaruhi variabel pengeluaran maupun sebaliknya. Dengan kata lain, dalam analisis regresi variabel diperlakukan berbeda yaitu variabel tak bebas (*Y*) dan variabel bebas (*X*), sedangkan dalam analisis korelasi, kedua variabel diperlakukan sama atau setara.

7.1.4. Penaksiran Parameter Regresi

Hasil akhir dari sebuah analisis regresi adalah membentuk persamaan regresi yang nantinya akan digunakan untuk menaksir atau menduga rata-rata variabel tak-bebas (*Y*) berdasarkan variabel bebas (*X*). Persamaan regresi populasi didefinisikan sebagai :

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

Seperti yang sudah dijelaskan sebelumnya, bahwa penelitian tidak dapat dilakukan melalui sebuah sensus. Oleh karena, jika pengumpulan data dilakukan secara sensus, maka waktu, tenaga dan uang yang dibuthkan akan sangat besar. Untuk mengurangi ketiga resiko tersebut, maka penelitian dilakukan melalui teknik sampling. Dengan demikian, persamaan regresi yang diperoleh merupakan persamaan regresi sampel yang didefinisikan sebagai :

$$Y_i = a + bX_i + e_i$$

Tujuan analisis regresi adalah menaksir atau menduga rata-rata variabel tak bebas (*Y*) berdasarkan nilai variabel bebas (*X*) yang diketahui. Untuk itu, perlu dilakukan penaksiran parameter *a* dan *b* dalam persamaan regresi sampel di atas. Sehingga akan diperoleh persamaan regresi taksiran yang didefinisikan sebagai :

$$\hat{Y}_i = a + bX_i$$

Dengan : "^" dibaca "*hat*" atau "*topi*"

$\hat{Y}_i$ = penaksir (estimator) $E(Y_i|X_i)$

a = penaksir α

b = penaksir β

ε_i = *Error* (kekeliruan)

e_i = residual (Sisa)

Terdapat dua teknik penaksiran yang dapat digunakan dalam penaksiran parameter *a* dan *b*, yaitu metode *ordinary least square* (OLS) dan metode maksimum likelihood. Taksiran parameter regresi yang diperoleh melalui metode OLS adalah sebagai berikut :

$$b = \frac{\left(\sum_{i-1}^{n} x_i y_i\right) - \frac{\left(\sum_{i-1}^{n} x_i\right)\left(\sum_{i-1}^{n} y_i\right)}{n}}{\left(\sum_{i-1}^{n} x_i^2\right) - \frac{\left(\sum_{i-1}^{n} x_i\right)^2}{n}}$$

$$a = \frac{\sum_{i-1}^{n} y_i}{n} - b\frac{\sum_{i-1}^{n} x_i}{n} = \bar{Y} - b\bar{X}$$

Metode OLS pertama kali dikemukakan oleh Carl Friedrich Gauss kebangsaan Jerman. Gauss membuat asumsi yang dikenal sebagai asumsi klasik regresi yaitu :

1. Nilai harapan kekeliruan untuk setiap nilai *X* adalah nol atau ditulis : $E(\varepsilon_i|X_i) = 0$
2. Gangguan ε_i dan ε_j tidak berkolerasi atau tidak terjadi korelasi berurutan atau tidak terjadi autokorelasi (non-autokorelasi). Atau ditulis : $Cov(\varepsilon_i, \varepsilon_j)=0$
3. Varians gangguan ε_i untuk setiap X_i adalah konstan yang sama dengan dengan σ^2 (homoskedastisitas) Atau ditulis : $var(\varepsilon_i|X_i)= \sigma^2$
4. Gangguan ε_I tidak berkolerasi dengan X_i. Atau ditulis $Cov(\varepsilon_i, X_i)=0$

Dari keempat asumsi klasik di atas, terdapat tiga asumsi yang membentuk sebuah asumsi utama dalam sebuah analisis regresi, yaitu asumsi (1),(2) dan (3) yang secara ringkas ditulis sebagai $\varepsilon_i \sim N(0, \sigma^2)$. Artinya, *error term* dalam model regresi populasi harus mengikuti distribusi normal dengan rata-rata *0* dan simpangan baku σ.

7.1.5. Langkah-Langkah Analisis Regresi

Untuk dapat menaksir atau menduga nilai variabel tak bebas (*Y*) berdasarkan nilai variabel bebas (*X*) yang diketahui, melalui sebuah persamaan regresi, maka langkah-

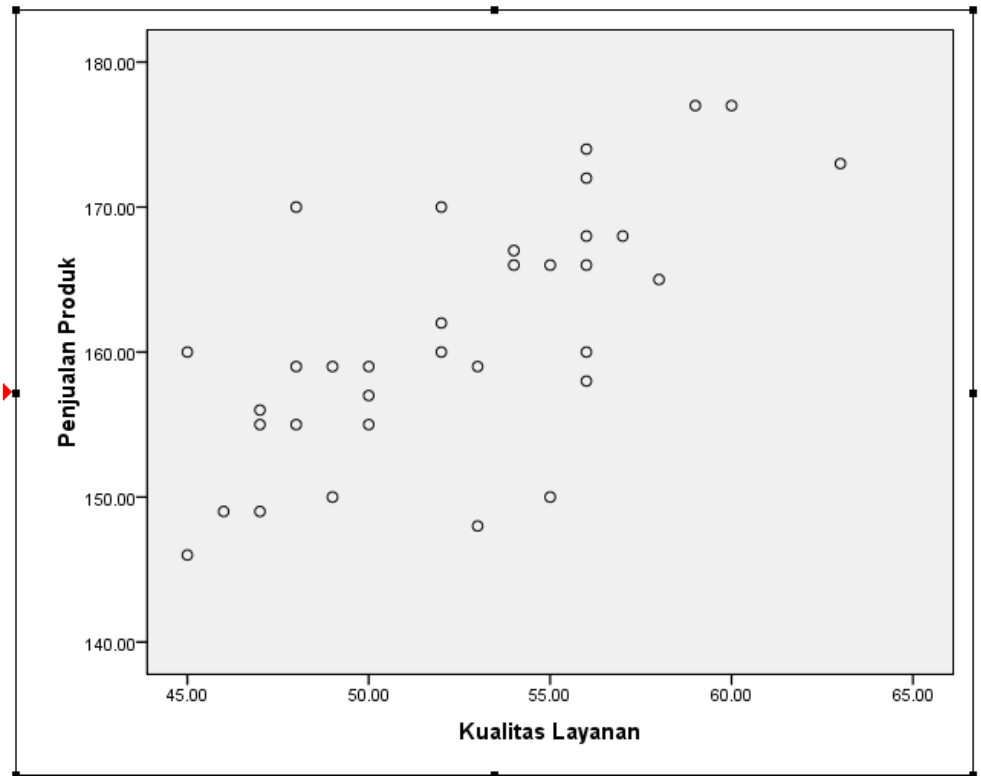
langkah yang harus dilakukan dalam membentuk persamaan regresi tersebut adalah sebagai berikut :

1. Telaah hubungan linear antara variabel *Y* dan variabel *X*
2. Uji asumsi klasik
3. Uji model secara overall
4. Uji model secara parsial
5. Koefisien determinasi

7.1.5.1. Telaah Hubungan Linear

Hubungan linear antara variabel *X* dan variabel *Y* dapat ditelaah melalui sebuah plot atau *scatter plot*. Jika sebuah plot membentuk sebuah trend ke atas atau trend menurun ke bawah maka hubungan antara variabel *X* dan variabel *Y* adalah linear. Sebaliknya jika plot membentuk sebuah pola musiman (gelombang) dan pola acak maka hubungan antara variabel *X* dan variabel *Y* bukanlah linear.

Perhatikan *Scatter plot* berikut :



plon t di atas, jelas menunjukkan bahwa semakin tinggi nilai kualitas layanan (variabel *X*) yaitu titik-titik nilainya semakin ke kanan, semakin tinggi juga nilai penjualan produk (variabel *Y*) yaitu titik-titik nilainya semakin ke atas. Dengan demikian titik-

titik nilai dari pasangan (*x*,*y*) membentuk sebuah pola menaik atau membentuk trend linear ke atas, artinya hubungan antara variabel kualitas layanan dan variabel penjualan produk adalah linear.

7.1.5.2. Uji Asumsi Klasik

Seperti yang sudah dijelaskan sebelumnya, dalam membentuk sebuah persamaan regresi, terdapat beberapa asumsi yang harus dipenuhi yaitu normalitas, non-autokorelasi, homoskedastisitas dan non-multikolinieritas (khusus untuk analisis regresi berganda)

Asumsi normalitas yang dimaksud adalah bahwa residual (*error terms*) dari model regresi harus mengikuti pola distribusi normal. Terdapat dua cara dalam menguji asumsi ini, yaitu metode grafik melalui plot normal kuantil dan uji hipotesisi melalui statistik Kolmogorov-Smirnov). Hipotesis yang diuji adalah sebagai berikut :

H_0 : Distribusi residual (*error term*) mengikuti pola distribusi normal

H_1 : Distribusi residual (*error term*) tidak mengikuti pola distribusi normal

Asumsi kedua yang harus dipenuhi dalam sebuah analisis regresi adalah non-autokorelasi. Asumsi non-autokorelasi memberikan pengertian bahwa tidak terdapat korelasi di antara dua residual (*error terms*) yang beruntun. Sama halnya dengan asumsi normalitas, terdapat dua cara dalam menguji asumsi ini yaitu melalui metode grafik dan uji hipotesis. Metode grafik yang digunakan adalah dengan memplotkan nilai *e* (residual) terhadapat *t* (waktu), dimana jika secara visual terdapat pola yang sistematis maka *e* (residual) cenderung berautokorelasi. Oleh karena uji asumsi melalui metode grafik selalu subjektif menurut penelitinya dan sering ditemukan adanya kekeliruan dalam pengambilan kesimpulan maka diperlukan sebuah uji hipotesis untuk memperkuat hasil pengujian melalui metode grafik. Uji hipotesis dilakukan dengan menggunakan statistik Durbin-Watson dengan hipotesis yang diuji adalah sebagai berikut :

$H_0 : \rho = 0$ Tidak terdapat autokorelasi

$H_1 : \rho \neq 0$ Terdapat autokorelasi

Statistik Durbin-Watson didefinisikan sebagai :

$$d_H = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

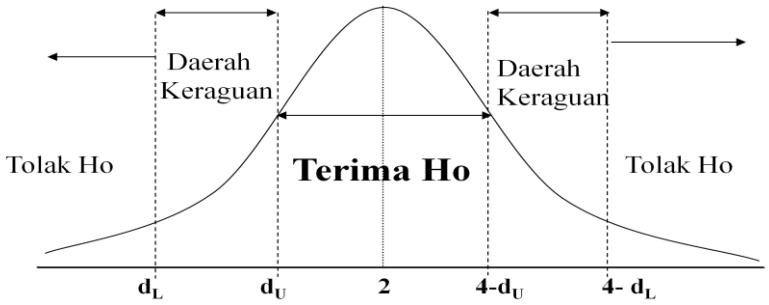
Kriteria Uji

Bandingkan d_H dengan d_{Tabel}

d_{Tabel} dari tabel Durbin Watson

$d_L=d_B$: Batas Bawah

$d_U=d_A$: Batas Atas



Asumsi klasik lain dalam sebuah analisis regresi adalah homoskedastisitas, artinya varians dari residual (*error terms*) untuk setiap nilai *X* adalah sama atau konstan. Untuk menguji asumsi ini dilakukan dengan dua cara, yaitu metode grafik menggunakan *scatter plot* antara nilai taksiran variabel *Y* dengan nilai residual kuadrat atau nilai variabel *X* dengan nilai residual kuadrat.

$$\hat{Y} \; dengan \; e^2 \; atau \; X \; dengan \; e^2$$

Jika plot tidak menunjukkan pola tertentu, maka homoskedastisitas dapat dikatakan terpenuhi. Sama seperti dua asumsi sebelumnya, untuk memperkuat hasil pengujian melalui metode grafik digunakan uji hipotesis menggunakan statistik *rank Spearman*. Prosedur pengujiannya adalah sebagai berikut :

1. Uji Hipotesis :

   H_0 :ρ=0 Tidak Terdapat Heteroskedastisitas

   H_1 : ρ≠0 Terdapat Heteroskedastisitas
2. Dengan MKTB hitung e_i untuk ∀i =1,2,..,n
3. Tentukan Rank $|e_i|$ dengan X_i tanpa memperhatikan tanda – atau +
4. Hitung Statistik Uji :

$$t = \frac{r_s\sqrt{n-2}}{\sqrt{1-r_s^2}} \sim t_{n-2} \; dengan \; r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

d_i : Selisih rank X_i dengan $|e_i|$

5. Kriteria uji :

   H_0 diterima jika $-t_{\alpha/2;n-2} \leq t \leq t_{\alpha/2;n-2}$ dan ditolak jika $t > t_{\alpha/2;n-2}$ atau $t < -t_{\alpha/2;n-2}$

Jika variabel bebas (X) yang dilibatkan dalam analisis regresi lebih dari satu maka analisis regresi yang dlakukan dinamakan sebagai analisis regresi berganda. Dalam membentuk persamaan regresi pada analisis regresi berganda, selain ketiga asumsi yang sudah dijelaskan sebelumnya, terdapat asumsi lain yang harus dipenuhi, yaitu asumsi non-Multikolinieritas. Asumsi ini memberikan pengertian bahwa tidak terdapat korelasi di antara variabel-variabel bebas (X_i) dengan (X_j). Pemeriksaan asumsi ini dilakukan menggunakan statistik *Variance Inflation Factor* (VIF) dengan kriteria bahwa nilai VIF yang lebih besar dari 10 mengindikasikan adanya kolinieritas di antara variabel bebas.

7.1.5.3. Uji Model Secara *Over All*

Setelah asumsi klasik dipenuhi, langkah selanjutnya dalam sebuah analisis regresi adalah menguji model secara *over all*. Tujuan dari uji ini adalah untuk menguji apakah persamaan regresi yang dibentuk signifikan dapat digunakan untuk menaksir atau menduga nilai variabel *Y* berdasarkan nilai variabel *X* yang diketahui. Hipotesis yang diuji adalah sebagai berikut :

H_0 : b = 0 (Tidak terdapat pengaruh dari variabel *X* terhadap Variabel *Y*)

H_1 : b ≠ 0 (Terdapat pengaruh dari variabel *X* terhadap Variabel *Y*)

Taraf Signifikansi α

Statistik Uji :

$$F_{hitung} = \frac{KT\left(\text{Re } gresi\right)}{KT\left(Galat\right)}$$

Kriteria Uji :

Tolak H_0 jika $F_{hitung} > F_{\alpha,v1,v2}$

Terima H_0 jika $F_{hitung} < F_{\alpha,v1,v2}$

$v_1 = k$ banyaknya variabel independen

$v_2 = n\text{-}k\text{-}1$

7.1.5.4. Uji Model Secara Parsial

Uji model secara parsial dilakukan untuk menguji apakah variabel *X* mempengaruhi variabel *Y* secara signifikan. Hipotesis yang diuji sama seperti pada uji model secara *over all*, yaitu :

H_0 : $b = 0$ (Tidak terdapat pengaruh dari variabel *X* terhadap Variabel *Y*)

H_1 : $b \neq 0$ (Terdapat pengaruh dari variabel *X* terhadap Variabel *Y*)

Taraf Signifikansi α

Statistik Uji :

$$s(b_1) = \sqrt{\frac{KTG}{Sxx}} \qquad t_{hitung} = \frac{b_1}{s(b_1)}$$

Kriteria Uji :

Terima H_0 jika $-t_{\alpha/2,n-2} \leq t_{hitung} \leq t_{\alpha/2,n-2}$, tolak dalam hal lainnya.

7.1.5.5. Koefisien Determinasi

Koefisien determinasi menunjukkan persentase fluktuasi atau variasi pada suatu variabel *Y* dapat dijelaskan atau disebabkan oleh variabel lain (*X*). Koefisien determinasi merupakan koefisien korelasi yang dikuadratkan. Koefisien determinasi didefinisikan sebagai :

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y} - \bar{y})^2}{\sum_{i=1}^{n}(y - \bar{y})^2} = \frac{JK(\text{Re}\,gresi)}{JK(Total)}$$

Hubungan antara koefisien determinasi dengan *standard error of estimate* didefinisikan sebagai :

$$R^2 = \frac{JK(\text{Re}\,gresi)}{JK(Total)} = 1 - \frac{JK(\text{Re}\,sidu)}{JK(Total)}$$

$$Std.Error = \sqrt{\frac{JK(\text{Re}\,sidu)}{n-2}}$$

7.2. Korelasi

7.2.1. Koefisien Korelasi Sederhana

Koefisien korelasi menyatakan besar keeratan hubungan antara variabel *X* dan variabel *Y* yang didefinisiikan sebagai :

$$r = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n}}{\sqrt{\left[\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}\right]\left[\sum_{i=1}^{n} Y_i^2 - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n}\right]}} = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$$

Sebelum kita menyatakan besar keeratan hubungan antara kedua variabel tersebut, terlebih dahulu harus dilakukan pengujian hipotesis apakah terdapat hubungan yang signifikan antara kedua variabel. Hipotesis yang diuji yaitu :

$H_0 : \rho = 0$

$H_1 : \rho \neq 0$

Taraf Signifikansi α

Statistik Uji :

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

Kriteria Uji :

Terima H_0 jika $-t_{\alpha/2,n-2} \leq t_{hitung} \leq t_{\alpha/2,n-2}$, tolak dalam hal lainnya.

Setelah terbukti melalui hasil pengujian bahwa terdapat hubungan yang signifikan antara kedua variabel baru kita dapat menafsirkan arti dari koefisien korelasi yang telah diperoleh. Arti dari koefisien korelasi menurut Guilford (Mindra Jaya, 2010) sebagai berikut :

Nilai Koefisien Korelasi	Tafsiran
0 - < 0,2	Hubungan yang sangat rendah dan dapat diabaikan atau dapat dianggap tidak terdapat korealsi antara variabel
≥0,2 - < 0,4	Hubungan yang rendah atau hubungan tidak erat
≥0,4 - < 0.7	Hubungan yang moderat / sedang
≥0,7 - < 0,9	Hubungan yang erat / kuat
≥0,9 - < 1	Hubungan yang sangat erat / sangat kuat

Nilai koefisien korelasi berkisar pada $-1 \leq r \leq 1$, artinya koefisien korelasi bisa bertanda positif dan negative. Mindra Jaya (2010) menyebutkan bahwa :

- Tanda positif menunjukkan adanya hubungan yang selaras antara variabel bebas dengan variabel tak bebas (dalam arti semakin tinggi nilai dari variabel bebas semakin tinggi pula nilai dari variabel tak bebas)
- Tanda negatif menunjukkan adanya hubungan yang terbalik antara variabel tak bebas dengan variabel bebas (dalam arti semakin tinggi nilai dari variabel bebas semakin kecil nilai dari variabel tak bebas)

7.3. Contoh Kasus

Perhatikan contoh berikut, Sebuah perusahaan melakukan penelitian dengan tujuan ingin mengetahui taksiran penjualan produk berdasarkan kualitas pelayanan yang diberikan.Data yang digunakan adalah sebagai berikut :

No.	Kualitas Layanan (X)	Penjualan Barang (Y)
1	54	167
2	50	155
3	53	148
4	45	146
5	48	170
6	63	173
7	46	149
8	56	166
9	52	170
10	56	174
11	47	156
12	56	158
13	55	150
14	52	160
15	50	157
16	60	177
17	55	166

No.	Kualitas Layanan (X)	Penjualan Barang (Y)
18	45	160
19	47	155
20	53	159
21	49	159
22	56	172
23	57	168
24	50	159
25	49	150
26	58	165
27	48	159
28	52	162
29	56	168
30	54	166
31	59	177
32	47	149
33	48	155
34	56	160

Untuk memperoleh persamaan regresi adalah :

No.	Kualitas Layanan (X)	Penjualan Barang (Y)	XY	X^2	Y^2
1	54	167	9018	2916	27889
2	50	155	7750	2500	24025
3	53	148	7844	2809	21904

No.	Kualitas Layanan (X)	Penjualan Barang (Y)	XY	X^2	Y^2
4	45	146	6570	2025	21316
5	48	170	8160	2304	28900
6	63	173	10899	3969	29929
7	46	149	6854	2116	22201
8	56	166	9296	3136	27556
9	52	170	8840	2704	28900
10	56	174	9744	3136	30276
11	47	156	7332	2209	24336
12	56	158	8848	3136	24964
13	55	150	8250	3025	22500
14	52	160	8320	2704	25600
15	50	157	7850	2500	24649
16	60	177	10620	3600	31329
17	55	166	9130	3025	27556
18	45	160	7200	2025	25600
19	47	155	7285	2209	24025
20	53	159	8427	2809	25281
21	49	159	7791	2401	25281
22	56	172	9632	3136	29584
23	57	168	9576	3249	28224
24	50	159	7950	2500	25281
25	49	150	7350	2401	22500
26	58	165	9570	3364	27225
27	48	159	7632	2304	25281
28	52	162	8424	2704	26244
29	56	168	9408	3136	28224
30	54	166	8964	2916	27556
31	59	177	10443	3481	31329
32	47	149	7003	2209	22201
33	48	155	7440	2304	24025
34	56	160	8960	3136	25600
Total	**1782**	**5485**	**288380**	**94098**	**887291**

$$b = \frac{288380 - \frac{(1782)(5485)}{34}}{94098 - \frac{(1782)^2}{34}} = 1,2874 \quad a = \frac{5485}{34} - \left[(1,2874)\left(\frac{1782}{34}\right)\right] = 93,8495$$

Sehingga diperoleh model regrsi taksirannya adalah sebagai berikut :

$$\hat{y}_i = 93,8495 + 1,2874x_i$$

Artinya, Secara rata-rata penjualan barang pada saat kualitas layanan nol adalah 93,845 (94 *pcs*) dan meningkat sebesar 1,2874 (1 *pcs*) jika kualitas layanan meningkat sebesar satu satuan.

Signifikansi Model secara overall

H_0 : $b = 0$ (Tidak terdapat pengaruh dari kualitas layanan terhadap penjualan produk)

H_1 : $b \neq 0$ (terdapat pengaruh dari kualitas layanan terhadap penjualan produk)

Taraf Signifikansi α =5%

Statistik Uji :

$$F_{hitung} = \frac{KT\left(\text{Re } gresi\right)}{KT\left(Galat\right)}$$

$$\begin{aligned} JK\left(\text{Re } gresi\right) &= b_1 S_{xy} \\ &= b_1 S_{xy} \\ &= 1,2874\left(288380 - \frac{(1782)(5485)}{34}\right) \\ &= 1,2874\left(901,4705\right) \\ &= 1160,5374 \end{aligned}$$

$$\begin{aligned} \text{JK}\left(\text{Total}\right) &= S_{yy} \\ &= \sum_{i=1}^{n} Y_i^2 - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n} \\ &= 887291 - \frac{(5485)^2}{34} \\ &= 2431,4412 \end{aligned}$$

$$\begin{aligned} \text{JK}\left(\text{Residu}\right) &= JK\left(Total\right) - JK\left(\text{Re } gresi\right) \\ &= 2431,4412\text{-}1160,5374 \\ &= 1270,9038 \end{aligned}$$

$$KT\left(\text{Re } gresi\right) = \frac{JK\left(\text{Re } gresi\right)}{v_1} = \frac{1160,5374}{1} = 1160,5374$$

$$KT\left(Galat\right) = \frac{JK\left(\text{Re } sidu\right)}{v_2} = \frac{1270,9038}{34-1-1} = \frac{1270,9038}{32} = 39,7157$$

$$F = \frac{KT\left(\text{Re } gresi\right)}{KT\left(Galat\right)}$$
$$= \frac{1160{,}5374}{39{,}7157}$$
$$= 29{,}2211$$

Tabel Anava nya dapat ditulis sebagai berikut :

Sumber Variasi	DF	JK	KT	F
Regresi	1	1160.5374	1160.537	29.22109
Galat/Residual	32	1270.9038	39.71574	
Total	33	2431.4412		

Kriteria Uji :

Tolak H_0 jika $F_{hitung} > F_{\alpha,v1,v2}$

Terima H_0 jika $F_{hitung} < F_{\alpha,v1,v2}$

v_1 = *k* banyaknya variabel independen

v_2 = *n*-*k*-1

Nilai Kritis:

$F_{0.05;\ 1;32}$ = = FINV(0.05,1,32) = 4,1491

H_0 ditolak karena $F_{hitung} > F_{\alpha,v1,v2}$

Artinya, terdapat pengaruh signifikan dari kualitas layanan terhadap penjualan produk

Signifikansi Model *secara parsial*

H_0 : b = 0 (Tidak terdapat pengaruh dari kualitas layanan terhadap penjualan produk)

H_1 : b ≠ 0 (Terdapat pengaruh dari kualitas layanan terhadap penjualan produk)

Taraf Signifikansi α =5%

Statistik Uji :

$$s(b_1) = \sqrt{\frac{KTG}{Sxx}} = \sqrt{\frac{39{,}7157}{94098 - \frac{(1782)^2}{34}}} = 0{,}2381$$

$$t_{hitung} = \frac{b_1}{s(b_1)} = \frac{1{,}2874}{0{,}2381} = 5{,}4056$$

Kriteria Uji :

Terima H_0 jika $-t_{\alpha/2,n-2} \leq t_{hitung} \leq t_{\alpha/2,n-2}$, tolak dalam hal lainnya.

$t_{\alpha/2,n-2}$ = TINV(0.05,32)= 2,0369

Kesimpulan:

H_0 ditolak karena $t_{hitung} > t_{\alpha/2,n-2}$

Artinya, kualitas layanan berpengaruh signifikan terhadap penjualan produk

Koefisien Determinasi :

$$R^2 = \frac{JK(\text{Re}\,gresi)}{JK(Total)} \times 100\%$$
$$= \frac{1160{,}5374}{2431{,}4412} \times 100\%$$
$$= 47{,}73\%$$

Koefisien determinasi sebesar 47,73% menjelaskan bahwa sebesar 47,73% variasi dari penjualan produk dapat dijelaskan oleh kualitas layanan dalam hubungan yang linear.

Nilai korelasi antara kualitas layanan dan penjualan :

$$r = \frac{\left(288380 - \frac{(1782)(5485)}{34}\right)}{\sqrt{\left[94098 - \frac{(1782)^2}{34}\right]\left[887291 - \frac{(5485)^2}{34}\right]}} = 0{,}690872$$

H_0 : $\rho = 0$ (tidak terdapat korelasi antara kualitas layanan dan penjualan)

$H_1 : \rho \neq 0$ (terdapat korelasi antara kualitas layanan dan penjualan)

Taraf Signifikansi 5%

Statistik Uji : $t = \dfrac{0,690872\sqrt{34-2}}{\sqrt{1-0,690872^2}} = 5,405654$

Kriteria Uji :

Terima H_0 jika $-t_{0,025,32} \leq t_{hitung} \leq t_{0,025,32}$ tolak dalam hal lainnya.

$t_{0,025,32} =$ TINV(0.05,32)= 2,0369

Kesimpulan : $t_{hitung} > 2,0369$ sehingga H_0 ditolak, artinya terdapat korelasi antara kualitas layanan dan penjualan

***Case Study* 1**

Suatu penelitian lingkungan bertujuan untuk mengetahui tingkat pencemaran yang berasal dari mobil. Dalam hal ini diperkirakan bahwa tingkat emisi hydrokarbon (HC) dari mntung dari jaraknya. Dengan demikian, mobil yang masih baru lebih sedikit mengeluarkan HC daripada mobil tua. Untuk itu, sebanyak 10 mobil merek tertentu dipilih secara acak, kemudian diperiksa berapa jarak tempuh (dalam ribuan kilometer) dari mobil tersebut dan diukur tingkat emisi HC-nya (dalam ppm). Hasilnya adalah sebagai berikut :

Jarak (x)	31	38	48	52	63	67	75	84	89	99
Emisi (y)	553	590	608	650	700	680	834	752	845	960

Berdasarkan data di atas :

1. Untuk mengetahui keeratan hubungan antara jarak tempuh dan tingkat emisi, hitung nilai korelasi antara jarak tempuh dengan tingkat emisi (HC), kemudian lakukan uji signifikansi korelasi antara jarak tempuh dan tingkat emisi!
2. Telaah hubungan antara jarak tempuh dengan tingkat emisi dengan membuat *scatter plot* antara kedua variabel tersebut, simpulkan!
3. Hitunglah nilai parameter *a* dan *b*, kemudian tuliskan persamaan regresi antara jarak tempuh dan tingkat emisi (HC), interpretasikan!
4. Lakukan uji signifikansi model regresi pada (3) secara *overall* untuk mengetahui apakah persamaan regresi yang diperoleh dapat digunakan untuk menaksir tingkat emisi (HC), simpulkan!
5. Lakukan uji model regresi pada (3) secara parsial untuk mengetahui apakah jarak tempuh mempengaruhi tingkat emisi (HC) secara signifikan atau tidak, simpulkan!
6. Hitunglah nilai koefisien determinasi untuk mengetahui besar variasi tingkat emisi (HC) yang dapat dijelaskan oleh jarak tempuh, jelaskan!

***Case Study* 2**

PT Cemerlang dalam beberapa bulan gencar mempromosikan sejumlah peralatan elektronik dengan membuka outlet-outlet di berbagai daerah. Berikut data mengenai penjualan dan biaya promosi yang dikeluarkan.

Daerah	Sales (juta rupiah)	Promosi (juta rupiah)
Jakarta	205	26
Tangerang	206	28
Bekasi	254	35
Bogor	246	31
Bandung	201	21
Semarang	291	49
Solo	234	30
Yogya	209	30
Surabaya	204	24
Purwokerto	216	31
Madiun	245	32
Tuban	286	47
Malang	312	54
Kudus	265	40
Pekalongan	322	42

Perusahaan tersebut, ingin mengetahui apakah promosi yang dilakukan berpengaruh secara signifikan terhadap tingkat penjualan (*sales*) dan persamaan regresi yang dapat digunakan untuk memprediksi penjualan (*sales*) berdasarkan biaya promosi yang dikeluarkan. Berdasarkan data di atas :

1. Untuk mengetahui keeratan hubungan antara promosi dan tingkat penjualan, hitung nilai korelasi antara kedua variabel tersebut kemudian lakukan uji signifikansi korelasi antara kedua variabel tersebut!
2. Telaah hubungan antara promosi dan tingkat penjualan dengan membuat *scatter plot* antara kedua variabel tersebut, simpulkan!
3. Hitunglah nilai parameter *a* dan *b*, kemudian tuliskan persamaan regresi antara promosi dan tingkat penjualan, interpretasikan!

4. Lakukan uji signifikansi model regresi pada (3) secara *overall* untuk mengetahui apakah persamaan regresi yang diperoleh dapat digunakan untuk memprediksi penjualan berdasarkan biaya promosi yang dikeluarkan, simpulkan!
5. Lakukan uji model regresi pada (3) secara parsial untuk mengetahui apakah tingkat penjualan dipengaruhi oleh promosi secara signifikan, simpulkan!
6. Hitunglah nilai koefisien determinasi untuk mengetahui besar variasi tingkat penjualan yang dapat dijelaskan oleh biaya promosi yang dikeluarkan, jelaskan!

***Case Study* 3**

Bank Indonesia (BI) sebagai Bank Sentral saat ini tidak berniat menaikkan tingkat suku bunga untuk dapat menekan lajunya dollar terhadap rupiah, karena jika kenaikkan suku bunga ini dilakukan oleh BI, hal tersebut tentunya akan memperburuk rekapitulasi perbankan, yang pada akhirnya sektor dunia usaha lainnya kembali terpuruk.
BI menduga bahwa menguatnya kurs dollar yang terjadi pada semua mata uang (termasuk rupiah) dewasa ini lebih disebabkan oleh perilaku *Federal Reserve* (Bank Sentral Amerika) yang menaikkan suku bunganya, hal seperti ini diprediksi hanya bersifat temporer dan situsional, yang diharapkan tidak akan berlangsung lama.
Untuk membuktikan dugaanya, BI berniat untuk melakukan penelitian untuk mengetahui seberapa besar pengaruh perubahan tingkat suku bunga oleh *Federal Reserve* terhadap perubahan kurs rupiah. Untuk itu diamati 10 data perubahan tingkat suku bunga yang diimbangi oleh perubahan kurs, hasilnya sebagai berikut :

Perubahan Tingkat Suku Bunga (%)	-0.5	0.1	0.3	-0.1	0.5	-0.5	-0.4	0.2	-0.3	0.4
Perubahan Kurs Rupiah (%)	200	50	-200	50	-300	300	250	-100	350	-250

Berdasarkan data tersebut :

(i) Tentukan persamaan regresi linear antara perubahan tingkat suku bunga dengan perubahan kurs rupiah, interpretasikan!

(ii) Tentukan besar keeratan hubungan (korelasi) antara perubahan tingkat suku bunga dengan perubahan kurs rupiah, kemudian lakukan uji signifikansinya dan simpulkan!

(iii) Lakukan uji model secara *over all* dan parsial untuk menentukan apakah dapat dikatakan dengan semakin turunnya tingkat suku bunga akan menyebabkan semakin kuatnya nilai rupiah?

(iv) Hitung nilai koefisien determinasi untuk mengetahui seberapa besar perubahan tingkat suku bunga memberikan kontribusi terhadap perubahan kurs rupiah?

Analisis *Time Series*

8.1. Definisi *Time Series*

Time series atau data deret waktu adalah sekumpulan hasil observasi yang diatur dan didapat menurut urutan kronologis, biasanya dalam interval waktu yang sama (Sudjana, 2005). *Time series* juga sering disebut data berkala yaitu data yang dikumpulkan dari waktu ke waktu untuk menggambarkan perkembangan suatu kegiatan, misalnya perkembangan produksi, harga, hasil penjualan, jumlah personil, penduduk, jumlah kecelakaan, jumlah kejahatan, jumlah peserta KB, dan lain sebagainya (J.Supranto, 2008). Selain dikenal sebagai data deret waktu dan data berkala, *time series* dikenal juga sebagai deret berkala yaitu suatu rangkaian seri dari nilai-nilai suatu variabel yang dicatat dalam jangka waktu yang berurutan (Lukas Setia Atmaja, 2009).

Berdasarkan ketiga definisi tersebut, jelas bahwa sebuah *times series* adalah data yang dikumpulkan menurut urutan waktu. Sebagai contoh, data mengenai penjualan mingguan sebuah produk di toko, data produksi bulanan di sebuah perusahaan industri dan data produksi tahunan bijih besi di Indonesia. Ketiga contoh data tersebut adalah *time series* dimana urutan waktunya adalah per minggu, per bulan dan atau per tahun.

8.2. Komponen *Time Series*

Suatu *time series* memiliki empat komponen yaitu Trend (T), variasi musiman (V), variasi siklis (S) dan *irreguler* atau random (R). Nilai-nilai suatu *time series* adalah kombinasi dari keempat komponen tersebut. Kombinasi yang dimaksud bersifat *multiplicative* atau perkalian yaitu :

Time Series = T x V x S x R

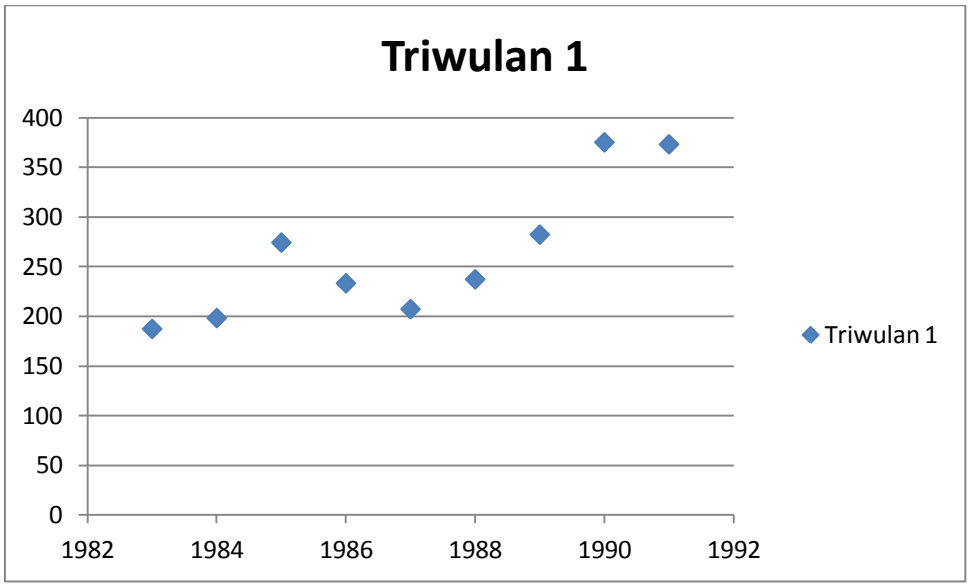
8.2.1. Trend (T)

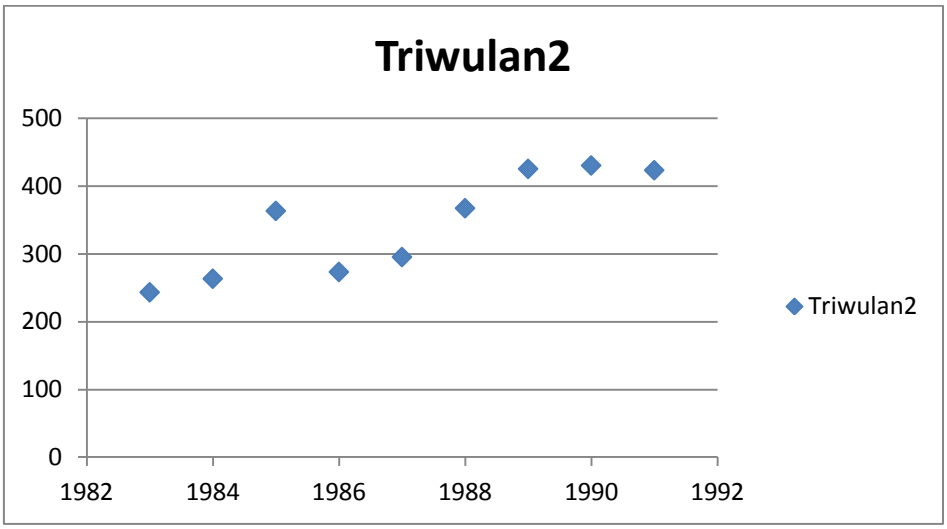
Trend melukiskan gerak data deret waktu selama jangka waktu yang panjang dan cukup lama (Sudjana, 2005). Menurut J. Supranto (2008), gerakan trend jangka panjang yaitu suatu gerakan yang menunjukkan arah perkembangan secara umum (kecenderungan menaik/menurun). Hal yang sama diungkapkan oleh Lukas Setia Atmaja (2009), trend merupakan gerakan jangka panjang yang memiliki kecenderungan menuju pada satu arah yaitu arah naik atau turun.

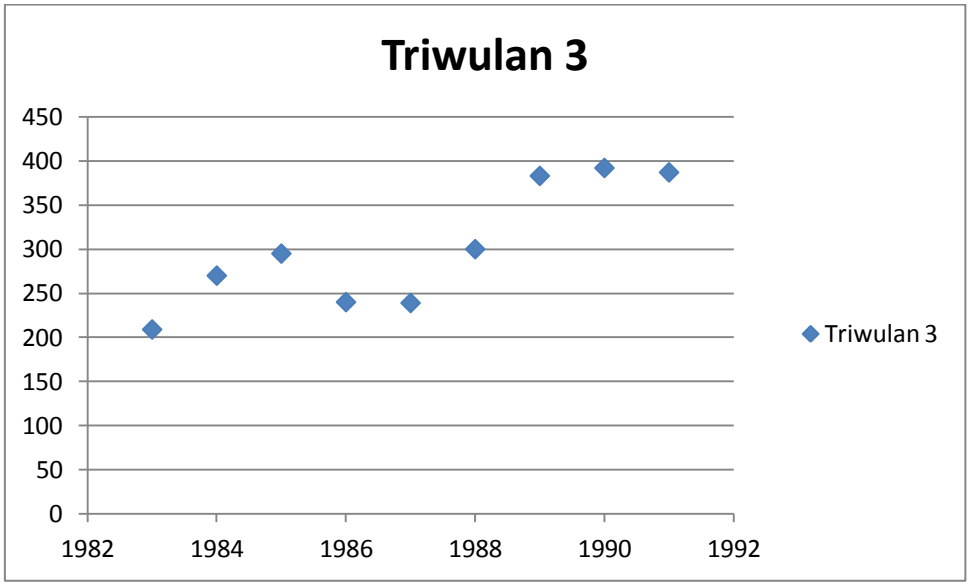
Perhatikan kasus berikut, misalkan terdapat sebuah data mengenai penjualan mobil di sebuah perusahaan sebagai berikut :

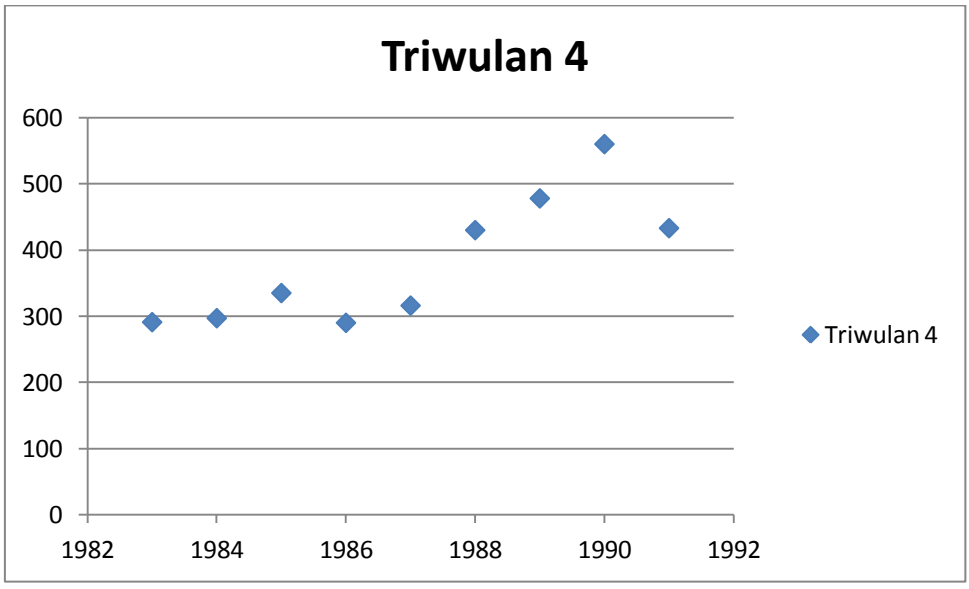
Tahun	Triwulan 1	Triwulan2	Triwulan 3	Triwulan 4
1983	187	243	209	291
1984	198	263	270	297
1985	274	363	295	335
1986	233	273	240	290
1987	207	295	239	316
1988	237	367	300	430
1989	282	425	383	478
1990	375	430	392	560
1991	373	423	387	433

Perusahaan tersebut ingin melihat gerakan penjualan mobil tersebut pada masing-masing triwulan. Untuk melihat gerakan tersebut dibuat *scatter plot* sebagai berikut :




Triwulan2
500
400
300
200
100
0
1982
1984
1986
1988
1990
1992
Triwulan2


Triwulan 3
450
400
350
300
250
200
150
100
50
0
1982
1984
1986
1988
1990
1992
Triwulan 3


Triwulan 4
600
500
400
300
200
100
0
1982
1984
1986
1988
1990
1992
Triwulan 4

Berdasarkan keempat *scatter plot* jelas terlihat bahwa gerakan perkembangan penjualan mobil di persuhaan tersebut cenderung ke arah naik. Artinya, dari tahun ke tahun mengalami kenaikan.

8.2.2. Variasi Siklis (S)

Variasi siklis merupakan perkembangan perekonomian yang turun naik sekitar trend. Variasi ini dapat dilukiskan dalam empat fase, untung, mundur, depresi dan pemulihan (Sudjana, 2005). Sama halnya dengan yang diungkapkan oleh J.Supranto (2008), variasi siklis adalah gerakan/variasi jangka panjang di sekitar garis trend (berlalu untuk data tahuan). Contohnya adalah *business cycle* yaitu gerakan siklis yang menunjukkan jangka waktu terjadinya kemakmuran (*prosperity*), kemunduran (*recession*), depresi (*depression*) dan pemulihan (*recovery*). Pengertian lain dari variasi siklis adalah suatu gerakan jangka panjang yang memiliki unsur siklus, yaitu perluasan (*expansion*), puncak (*peak*), kemunduran (*contraction*) dan depresi (*trough*).

8.2.3. Variasi Musiman (V)

Gerakan musiman terjadi lebih teratur bila dibandingkan dengan gerakan siklis dan bersifat lengkap, biasanya, selama satu tahun kalender (Sudjana, 2005). Faktor yang menyebabkan terjadinya gerak musiman adalah iklim dan kebiasaan. Jelas bahwa, variasi musiman adalah gerakan yang mempunyai pola tetap dari waktu ke waktu (J.Supranto, 2008). Sementara, menurut Lukas S.A. (2009), variasi musim adalah gerakan jangka pendek, kurang dari satu tahun, yang berluang secara teratur dari tahun ke tahun. Sebagai contoh adalah, harga bahan pokok naek menjelang bulan Ramdhan.

8.2.4. *Irreguler* atau Random (R)

Gerak *irreguler* bersifat tidak teratur dan sukar dikuasai (Sudjana, 2005). Gerak *rreguler* dikenal juga sebagai *random movements* adalah gerakan atau variasi yang sifatnya sporadis, misalnya naik turunnya produksi akibat banjir yang datangnya tidak teratur, gempa bumi, dll (J.Supranto, 2009).

8.3. Trend Linear dengan Metode *Least Square*

Persamaan garis lurus suatu trend, didefinisikan sebagai :

$$\hat{y} = a + bx$$

$\hat{y}$: nilai proyeksi variabel Y untuk suatu nilai X

a : Konstanta

b : Slope

Y : Observasi

x : Waktu (dalam *coding*)

Berdasarkan metode *Least Square*, taksiran parameter *a* dan *b* dalam sebuah persamaan trend didefinisikan sebagai :

$$a = \frac{\sum_{i=1}^{n} y_i}{n} \qquad b = \frac{\sum_{i=1}^{n} y_i \cdot x_i}{\sum_{i=1}^{n} x_i}$$

Untuk dapat membentuk sebuah persamaan trend, periode waktu *t* (hari, minggu, bulan, tahun, dll.) ditransformasi ke dalam sebuah koding waktu dengan ketentuan sebagai berikut :

8.3.1. Koding Waktu Data Ganjil

Jika data yang dikumpulkan banyaknya adalah ganjil, maka periode waktu yang berada di tengah (Median) dijadikan sebagai tahun dasar dengan koding waktunya (*x*) adalah nol (*0*). Koding waktu untuk periode sebelum tahun dasar adalah negatifnya yaitu -1,-2,-3,... sedangakan untuk periode waktu setelahnya adalah positifnya yaitu 1,2,3,...

Untuk lebih jelasnya perhatikan contoh berikut :

Tahun	Koding
1980	-2
1981	-1
1982	**0**
1983	1
1984	2

8.3.2. Koding Waktu Data Genap

Jika data yang dikumpulkan banyaknya adalah ganjil, maka tahun dasarnya berada di antara dua periode waktu yang berada di tengah (Median). Koding waktu untuk periode sebelum tahun dasar adalah negatifnya yaitu -1,-3,-5,... sedangkan untuk periode waktu setelahnya adalah positifnya yaitu 1,3,5,...

Untuk lebih jelasnya, perhatikan contoh berikut :

Tahun	Koding
1980	-5
1981	-3
1982	-1
1983	1
1984	3
1985	5

Perhatikan contoh kasus berikut, data berikut adalah data penjualan PT. Mendota selama 5 tahun. Data tersebut disajikan pada tabel berikut :

Tahun	Penjualan
1985	7
1986	10
1987	9
1988	11
1989	13

Perusahaan tersebut ingin membentuk sebuah persamaan trend yang dapat digunakan untuk menaksir atau menduga penjualan pada masa yang akan datang. Hasil penaksiran ini dapat digunakan untuk rencana produksi untuk tahun-tahun selanjutnya.

Tahun	**Penjualan**	***X***	X^2	***X* x *Y***
1985	7	-2	4	-14
1986	10	-1	1	-10
1987	9	0	0	0
1988	11	1	1	11
1989	13	2	4	26
Jumlah	**50**	**0**	**10**	**13**

$$a = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{50}{5} = 10 \qquad b = \frac{\sum_{i=1}^{n} y_i \cdot x_i}{\sum_{i=1}^{n} x_i} = \frac{13}{10} = 1,3$$

Dengan demikian, persamaan trend yang diperoleh dapat dituliskan sebagai :

$$\hat{y} = 10 + 1,3x$$

Berdasarkan persamaan trend yang diperoleh, selain dapat menaksir atau menduga nilai penjualan tahun-tahun yang akan datang, dapat juga digunakan untuk menaksir atau menduga penjualan di masa lalu, sebagai berikut :

Tahun	Penjualan	X	Penjualan Menurut Persamaan
1985	7	-2	10 + 1,3 (-2) = 7,4
1986	10	-1	10 + 1,3 (-1) = 8,7
1987	9	0	10 + 1,3 (0) = 10
1988	11	1	10 + 1,3 (1) = 11,3
1989	13	2	10 + 1,3 (2) = 12.6

- Prediksi penjualan tengah tahun 1990

  Koding tahun 1990 = 3

  $\hat{y} = 10 + 1,3(3) = 13,9$

  Artinya, taksiran penjualan pada pertengahan tahun 1990 adalah 13,9 satuan penjualan (bisa unit atau bisa dalam mata uang rupiah)

- Prediksi penjualan awal tahun 1991

  Koding awal tahun 1991 = 3 + 0,5 = 3,5

  $\hat{y} = 10 + 1,3(3,5) = 14,55$

  Artinya, taksiran penjualan pada awal tahun 1991 adalah 14,55 satuan penjualan (bisa unit atau bisa dalam mata uang rupiah)

8.4. Hubungan Tahun Dasar dengan Trend Tahunan

Jika dari sebuah data telah diperoleh persamaan trend dengan tahun dasar t_0 maka dengan seiring bertambahnya waktu tahun dasar tersebut menjadi sebuah tahun di masa lalu yang jaraknya cukup jauh dengan data yang dimiliki pada saat ini *t*. Agar persamaan trend yang sudah diperoleh tetap dapat digunakan untuk menaksir atau menduga nilai variabel pada masa

yang akan datang, dengan kondisi data yang lebih *up to date* maka tahun dasar trend dapat dirubah. Apabila tahun dasar dirubah, maka :

1. Konstanta (*a*) persamaan trend berubah

   a persamaan trend dengan tahun dasar baru sama dengan nilai trend pada periode dasar yang baru
2. Slope (*b*) persamaan trend tetap

Sebagai contoh perhatikan kasus berikut. Misalkan dari sebuah data diperoleh sebuah persamaan trend $\hat{Y} = 780 + 42X$ dengan tahun dasar awal 1983 dan ingin dirubah tahun dasarnya menjadi awal 1985 sebagai berikut :

Tahun	Coding (1983)	Coding (1985)
78	-9	-14
79	-7	-12
80	-5	-10
81	-3	-8
82	-1	-6
83	1	-4
84	3	-2
85	5	0
86	7	2
87	9	4

Berdasarkan tabel di atas, maka :

$b_{\text{baru}} = b$ lama $= 42$ $\qquad$ $a_{\text{baru}} = 780 + 42\,(5) = 990$

dengan demikian, diperoleh persamaan trend baru dengan tahun dasar awal 1985 yaitu $\hat{Y} = 990 + 42X$

Sebuah persamaan trend tahunan, selain dapat dirubah tahun dasarnya, juga dapat dirubah menjadi trend rata-rata bulanan, tren rata-rata kwartalan, trend bulanan dan tren kwartalan sebagai berikut :

- Trend Rata-Rata Bulanan

  $$\hat{Y} = \frac{a}{12} + \frac{b}{12} X$$

- Trend Rata-Rata Kwartalan

  $$\hat{Y} = \frac{a}{4} + \frac{b}{4} X$$

- Trend Bulanan

$$\hat{Y} = \frac{a}{12} + \frac{b}{(12)^2}U$$

- Trend Kwartalan

$$\hat{Y} = \frac{a}{4} + \frac{b}{(4)^2}U$$

8.5. Trend non-Linear

Pada kenyataanya, data berkala atau *time* series tidak selalu membentuk sebuah trend linear yang cenderung naik atau menurun. Data berkala yang mempunyai jangka waktu atau periode yang pendek cukup baik digambarikan oleh trend linear, akan tetapi data berkala yang mempunyai jangka waktu yang panjang tidak selalu membentuk trend linear akibat perubahan waktu. Sehingga penaksiran persamaan trend data berkala tersebut lebih baik menggunakan trend non-linear. Trend non-linear adalah garis trend yang tidak linear, misalnya trend kuadratik dan trend eksponensial.

1. Trend Kuadratik

Persamaan trend Kuadratik didefinisikan sebagai :

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

Dengan :

$$a = \frac{\left[\left(\sum y\right)\left(\sum x^4\right)\right] - \left[\left(\sum x^2 y\right)\left(\sum x^2\right)\right]}{\left(n\sum x^4\right) - \left(\sum x^2\right)^2}$$

$$b = \frac{\sum xy}{\sum x^2}$$

$$c = \frac{n\left(\sum x^2 y\right) - \left[\left(\sum x^2\right)\left(\sum y\right)\right]}{\left(n\sum x^4\right) - \left(\sum x^2\right)^2}$$

2. Trend Eksponensial

Persamaan trend Kuadratik didefinisikan sebagai :

$$\hat{Y} = a \times b^X$$

$$\ln \hat{Y} = \ln(a) + X \ln(b)$$

Dengan :

$$a = anti \ln \left(\frac{\sum \ln y}{n} \right)$$

$$b = anti \ln \left(\frac{\sum (x \ln y)}{\sum x^2} \right)$$

Permasalahan yang muncul dalam penaksiran sebuah persamaan trend dari data berkala adalah bagaimana kita menentukan trend yang cocok dengan data berkala yang diketahui. Terdapat dua cara dalam memilih trend, sebagi berikut :

- Menganalisis grafik atau *scatter plot*.

  Langkah yang harus dilakukan adalh membuat plot antara data dengan waktu (*t*), kemudian plot tersebut dianalisis apakah membentuk trend linear, kuadratik atau eksponensial,
- Menghitung *Mean Square Error* (MSE)

  Langkah yang harus dilakukan yaitu membuat persamaan trend linear dan non-linear (kuadratik dan Eksponensial). Kemudian tentukan nilai MSE dari masing-masing persamaan trend, bandingkan dan pilihlah persamaan trend yang memiliki nilai MSE terkecil.

8.6. Variasi Musiman

Terdapat tiga alasan mengapa variasi musiman perlu untuk dipelajari :

- Dapat membuat pola perubahan masa lalu
- Dapat memproyeksikan kondisi di masa mendatang dengan mengacu pada pola di masa lalu
- Sekali kita dapat mengetahui pola musiman, maka kita dapat mengubah efeknya dari *time series*.

Variasi musiman dari sebuah data berkala dapat diukur melalui sebuah angka indeks yang disebut sebagai **indeks musim**. Indeks musim adalah angka-angka yang bervariasi dari suatu angka dasar sebesar 100%. Jika suatu periode waktu misalnya dalam bulan tertentu (atau bisa dalam minggu, triwulan dan periode musim lainnya) memiliki nilai indeks musim sebesar 100, artinya pada bulan tersebut tidak terdapat variasi musim.

Indeks musim dapat dihitung menggunakan metode rata-rata dan metode *Ration Moving Average* (RMA). Dalam metode rata-rata, kita akan membandingkan nilai rata-rata musim

dengan suatu nilai rata-rata total. Sedangkan dalam metode RMA, kita akan mencari unsur variasi musim (V).

Perhatikan contoh perhitungan indeks musim melalui metode rata-rata berikut :

Tahun	Triwulan			
	1	2	3	4
	250	290	260	310
	290	270	340	410
	310	480	325	380
	305	360	340	385
	330	440	410	450
Jumlah	1485	1840	1675	1935
Total	6935			
Rata-Rata	297	368	335	387
Rata-Rata Total	346.75			
Indeks Triwulan (%)	85.65	106.13	96.61	111.61
Total Indeks	400.00			

Indeks triwulan diperoleh dari hasil perbandingan antara rata total dengan rata-rata masing-masing triwulan dikalikan dengan 100%. Dimana total indeks dari masing-masing triwulan harus sama dengan 400.

Contoh berikut adalah perhitungan indeks musim melalui metode RMA, dengan menggunakan data yang sama diperoleh :

Tahun	Triwulan	Penjualan	Moving Total	Moving Average	Moving Average yang dipusatkan	Persentase terhadap Moving Average
1988	1	125				
	2	87	442	110.5		
	3	90	447	111.75	111.125	80.9899
	4	140	451	112.75	112.25	124.7216
1989	1	130	456	114	113.375	114.6637
	2	91	462	115.5	114.75	79.3028
	3	95	470	117.5	116.5	81.5451
	4	146	475	118.75	118.125	123.5979
1990	1	138	478	119.5	119.125	115.8447
	2	96	482	120.5	120	80.0000
	3	98	493	123.25	121.875	80.4103
	4	150	502	125.5	124.375	120.6030
1991	1	149	506	126.5	126	118.2540
	2	105	515	128.75	127.625	82.2723

Tahun	Triwulan	Penjualan	Moving Total	Moving Average	Moving Average yang dipusatkan	Persentase terhadap Moving Average
	3	102	529	132.25	130.5	78.1609
	4	159	538	134.5	133.375	119.2127
1992	1	163	548	137	135.75	120.0737
	2	114	575	143.75	140.375	81.2110
	3	112				
	4	186				

perhitungan untuk masing-masing kolom :

- *Moving Total*

  Moving total diperoleh dengan menjumlah penjualan pada empat triwulan yang bergerak dari triwulan pertama sampai dengan triwulan keempat, selanjutnya bergerak dari triwulan kedua sampai dengan triwulan pertama tahun selanjutnya. Contoh :

  442 = 125+87+90+140

  447=87+90+140+130

- *Moving Average*

  Moving average diperoleh dengan membagi *moving total* dengan banyaknya triwulan yang dijumlahkan atau disini banyaknya triwulan yang dijumlahkan adalah 4. Contoh :

  110,5=442/4

  117,75=447.4

- *Moving Average* yang dipusatkan

  Moving average yang dipusatkan diperoleh dengan mencari rata-rata dari dua nilai *moving average*. Contoh :

  111,125=(110,5+111,75)/2

  112,25=(111,75+112,75)/2

- Persentase terhadap *Moving Average*

  Persentase terhadap *Moving Average* diperoleh dari hasil bagi antara *moving average* yang dipusatkan dengan penjualan dikalikan 100%. Contoh :

  80,9899=(111,125/90)x100%

  124,7216=(112,25/140)x100%

Langkah selanjutnya, sajikan persentasi terhadap *moving average* yang diperoleh dari hasil perhitungan pada tabel, menjadi sebagai berikut :

Tahun	Triwulan			
	1	2	3	4
1988	-	-	80.9899	124.7216
1989	114.6637	79.3028	81.5451	123.5979
1990	115.8447	80.0000	80.4103	120.6030
1991	118.2540	82.2723	78.1609	119.2127
1992	120.0737	81.2110	-	-
Rata-rata (V)	**117.2090**	**80.6965**	**80.2765**	**122.0338**
Total V	**400.2159**			
Indeks Musim	**117.1458**	**80.6530**	**80.2332**	**121.9680**

Indeks musim diperoleh dari hasil bagi antara rata-rata (V) dari masing-masing triwulan dengan total V dikalikan 400. Berdasarkan hasil perhitungan indeks pada tabel di atas, diperoleh bahwa pada semua nilai indeks dari keempat triwulan tidaklah 100, sehingga semua triwulan memiliki variasi musim.

8.7. Variasi Siklis

Variasi siklis adalah komponen-komponen deret berkala yang cenderung berada di atas dan di bawah garis trend untuk lebih dari 1 tahun. Identifikasi variasi siklis menggunakan metode residual. Metode residual memperlihatkan gerak siklus di sekitar garis trend. Langkah-langkah metode residual sebagai berikut :

1. Carilah nilai trend untuk masing-masing t_i (i=1,2,..)
2. Bandingkan nilai trend dengan nilai aktual, atau cari nilai persentase trend :

$$Persentase = \frac{Y}{\hat{Y}} \times 100\%$$

3. Jika nilai persentase yang diperoleh <100, artinya variasi siklis berada di bawah garis trend dan jika >100 berada di atas garis trend.

Variasi siklis, juga dapat diidentifikasi melalui nilai siklus relatif yang didefinisikan sebagai :

$$RCR = \frac{\left(Y \times \hat{Y}\right)}{\hat{Y}} \times 100\%$$

LATIHAN

1. Data berikut merupakan data deret waktu mengenai angka kelahiran per 1000 penduduk di suatu negara selama jangka waktu 1915,1920,...,1955.

Tahun	1915	1920	1925	1930	1935	1940	1945	1950	1955
Kelahiran per 1000 penduduk	25	23.7	21.3	18.5	16.9	17.6	19.5	23.6	24

   Buatlah persamaan trend nya dan lakukan peramalan kelahiran penduduk per 1000 penduduk pada 3 tahun selanjutnya!

2. Hasil penjualan barang "*A*" yang dilakukan oleh suatu pabrik selama 1957-1966, dalam jutaan rupiah adalah sebagai berikut :

Tahun	1957	1958	1959	1960	1961	1962	1963	1964	1965	1966
Penjualan (dalam jutaan Rupiah)	265.1	283.1	286.8	339.4	407.6	407.5	457.1	435.7	586.9	604.6

   Jika data tersebut memiliki trend berbentuk eksponensial, maka buatlah persamaan trendnya dan lakukan peramalan penjualan untuk 3 tahun berikutnya!

3. Pesanan semacam barang yang diterima oleh suatu pabrik setiap tahun, dinyatakan dalam jutaan rupiah adalah sebagai berikut :

Tahun	Pesanan Barang
1959	19,91
1960	24,47
1961	23,62
1962	22,69
1963	23,49
1964	28,06
1965	27,61

   Berdasarkan hasil perhitungan, diperoleh bahwa persamaan trend linear dari data pesanan barang tersebut adalah :

$$\hat{Y} = 24,26 + 1,08X$$

   Hitung prediksi rata-rata pesanan barang per bulan pada tahun 1966!

4. Berikut adalah data mengenai banyaknya pengangguran di Negara Indonesia Tahun 1986-2004.

Tahun	Pengangguran (Juta Orang)
1986	1.82
1987	1.82
1988	2.04
1989	2.04
1990	1.91
1991	1.99
1992	2.14
1993	2.2
1994	3.64

Tahun	Pengangguran (Juta Orang)
1996	4.28
1997	4.18
1998	5.05
1999	6.03
2000	5.81
2001	8.01
2002	9.13
2003	9.94
2004	10.25

Tentukan persamaan trend linear kemudian hitung prediksi banyaknya pengangguran untuk tengah tahun 2016!

DAFTAR PUSTAKA

Atmaja, Lukas Setia., Ph.D, Statistika untuk Bisnis dan Ekonomi, Penerbit Andi, Yogyakarta, 2009.

DR. Boediono, Statistika dan Probabilitas, ROSDA, Bandung, 2014.

Freud, Rudolf J. & William J. Wilson. *Statistical Methods, 2nd Edition*. Academic Press, USA, 2003.

Gujarati, Damodar N. *Basic Econometrics, fourth edition*. The McGraw−Hill Companies, 2004.

Jaya, I.G. Mindra. Modul Praktikum Analisis Regresi. Statistika Universitas Padjadjaran, 2010.

Sudjana. Metode Statistika, Edisi 6. Tarsito Bandung. 2005

Supranto,J., *Statistik Teori dan Aplikasi*, Edisi ketujuh Buku 1 dan 2, Penerbit Erlangga Jakarta, 2008.

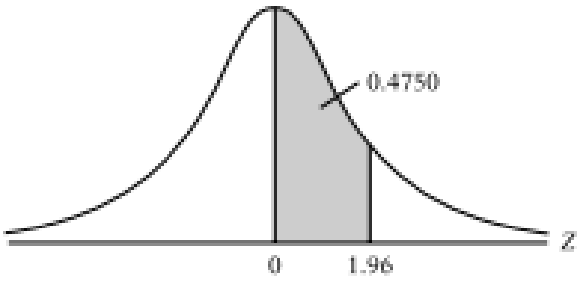
Weisberg, Sanford. *Applied Linear Regression, Third Edition*. John Wiley & Son,Inc. New Jersey. 2005.

TABLE D.1 AREAS UNDER THE STANDARDIZED NORMAL DISTRIBUTION

Example

$\Pr(0 \le Z \le 1.96) = 0.4750$

$\Pr(Z \ge 1.96) = 0.5 - 0.4750 = 0.025$



Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4454	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

Note: This table gives the area in the right-hand tail of the distribution (i.e., $Z \ge 0$). But since the normal distribution is symmetrical about $Z = 0$, the area in the left-hand tail is the same as the area in the corresponding right-hand tail. For example, $P(-1.96 \le Z \le 0) = 0.4750$. Therefore, $P(-1.96 \le Z \le 1.96) = 2(0.4750) = 0.95$.

TABLE D.2 PERCENTAGE POINTS OF THE *t* DISTRIBUTION

Example

$\Pr(t > 2.086) = 0.025$

$\Pr(t > 1.725) = 0.05$ for df = 20

$\Pr(|t| > 1.725) = 0.10$

df \ Pr	0.25 0.50	0.10 0.20	0.05 0.10	0.025 0.05	0.01 0.02	0.005 0.010	0.001 0.002
1	1.000	3.078	6.314	12.706	31.821	63.657	318.31
2	0.816	1.886	2.920	4.303	6.965	9.925	22.327
3	0.765	1.638	2.353	3.182	4.541	5.841	10.214
4	0.741	1.533	2.132	2.776	3.747	4.604	7.173
5	0.727	1.476	2.015	2.571	3.365	4.032	5.893
6	0.718	1.440	1.943	2.447	3.143	3.707	5.208
7	0.711	1.415	1.895	2.365	2.998	3.499	4.785
8	0.706	1.397	1.860	2.306	2.896	3.355	4.501
9	0.703	1.383	1.833	2.262	2.821	3.250	4.297
10	0.700	1.372	1.812	2.228	2.764	3.169	4.144
11	0.697	1.363	1.796	2.201	2.718	3.106	4.025
12	0.695	1.356	1.782	2.179	2.681	3.055	3.930
13	0.694	1.350	1.771	2.160	2.650	3.012	3.852
14	0.692	1.345	1.761	2.145	2.624	2.977	3.787
15	0.691	1.341	1.753	2.131	2.602	2.947	3.733
16	0.690	1.337	1.746	2.120	2.583	2.921	3.686
17	0.689	1.333	1.740	2.110	2.567	2.898	3.646
18	0.688	1.330	1.734	2.101	2.552	2.878	3.610
19	0.688	1.328	1.729	2.093	2.539	2.861	3.579
20	0.687	1.325	1.725	2.086	2.528	2.845	3.552
21	0.686	1.323	1.721	2.080	2.518	2.831	3.527
22	0.686	1.321	1.717	2.074	2.508	2.819	3.505
23	0.685	1.319	1.714	2.069	2.500	2.807	3.485
24	0.685	1.318	1.711	2.064	2.492	2.797	3.467
25	0.684	1.316	1.708	2.060	2.485	2.787	3.450
26	0.684	1.315	1.706	2.056	2.479	2.779	3.435
27	0.684	1.314	1.703	2.052	2.473	2.771	3.421
28	0.683	1.313	1.701	2.048	2.467	2.763	3.408
29	0.683	1.311	1.699	2.045	2.462	2.756	3.396
30	0.683	1.310	1.697	2.042	2.457	2.750	3.385
40	0.681	1.303	1.684	2.021	2.423	2.704	3.307
60	0.679	1.296	1.671	2.000	2.390	2.660	3.232
120	0.677	1.289	1.658	1.980	2.358	2.617	3.160
∞	0.674	1.282	1.645	1.960	2.326	2.576	3.090

Note: The smaller probability shown at the head of each column is the area in one tail; the larger probability is the area in both tails.

Source: From E. S. Pearson and H. O. Hartley, eds., *Biometrika Tables for Statisticians*, vol. 1, 3d ed., table 12, Cambridge University Press, New York, 1966. Reproduced by permission of the editors and trustees of *Biometrika*.