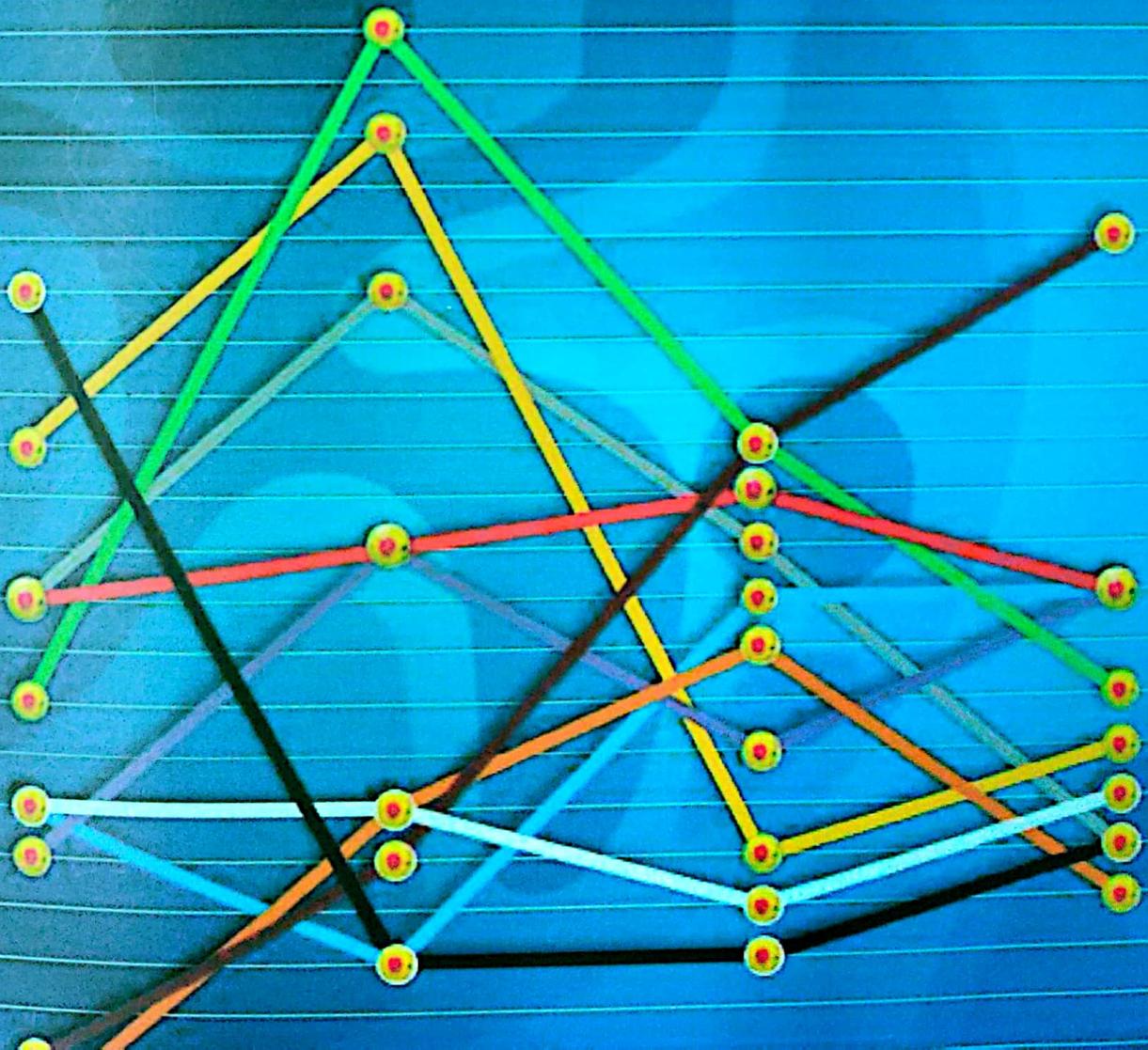


Dr. H. Rohmad, M.Pd.
Supriyanto, M.Si.

Kalimedia

Pengantar STATISTIKA

Panduan Praktis Bagi Pengajar dan Mahasiswa



Pengantar
STATISTIKA

Panduan Praktis Bagi Pengajar dan Mahasiswa

**Dr. H. Rohmad, M. Pd.
Supriyanto, M. Si**

 **Kalimedia**

**PENGANTAR STATISTIKA:
Panduan Praktis Bagi Pengajar dan Mahasiswa**

Penulis:

Dr. H. Rohmad, M. Pd.
Supriyanto, M. Si.

Editor:

Adnan A.R.

**Desain sampul dan Tata letak:
Kukuh Adi Prabowo**

ISBN: 978-602-72122-6-8

Penerbit:

KALIMEDIA
Perum POLRI Gowok Blok D 3 No. 200
Depok Sleman Yogyakarta
e-Mail: kalimediaok@yahoo.com
Telp. 082 220 149 510

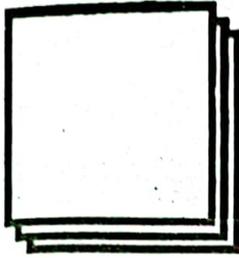
Distributor oleh:

KALIMEDIA
Telp. 0274 486 598
E-mail: marketingkalimedia@yahoo.com

Cetakan, 1 2015

Cetakan, 2 2016

Hak cipta dilindungi oleh undang-undang
Dilarang mengutip atau memperbanyak sebagian
atau seluruh isi buku ini tanpa izin tertulis dari Penerbit.



Kata Pengantar

Buku Statistika ini dipersiapkan oleh penulis agar dapat mengantarkan mahasiswa dan siapa saja yang ingin mempelajari statistika serta pengolahan data kuantitatif dalam penelitian, secara baik, mudah, dan benar. Dua kata terakhir menjadi penting mengingat sebagian orang merasakan dan mengeluh mempelajari statistik termasuk kategori sulit, buku ini diharapkan dapat menjadi pengantar secara "lebih mudah". Begitu juga dengan kata "benar" yakni sesuai dengan kaidah dan prinsip-prinsip statistika. Dalam statistika terdapat sejumlah rumus yang kadang-kadang memiliki ketepatan penggunaan yang berbeda antara satu dengan yang lainnya.

Materi buku ini disusun sedemikian rupa, mulai dari pengantar dan prinsip dasar hingga penggunaan teknik uji secara tepat. Buku ini dapat dipergunakan mahasiswa, dosen, peneliti, guru dengan berbagai disiplin ilmu. Karena secara substantif sebenarnya metode statistika dapat dipergunakan untuk berbagai cabang keilmuan, dan rumus-rumus yang

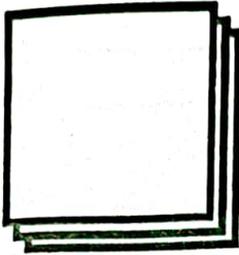
dipakai tidak ada perbedaan antara bidang keilmuan tertentu dengan yang lain.

Penulis mengucapkan terima kasih kepada semua pihak yang telah membantu dan memberikan kritikan dan masukan terhadap buku ini. Penulis menyadari sepenuhnya, sebagai sebuah pengantar buku masih memiliki beberapa kekurangan untuk itu kepada pembaca yang budiman kritik, saran, dan masukan penulis harapkan, agar pada cetakan atau edisi berikutnya lebih baik, sebelumnya diucapkan terima kasih. Semoga buku ini tiada gersang akan manfaat. Amiin.

Penulis

Dr. H. Rohmad, M. Pd.

Supriyanto, M. Si.



Daftar Isi

KATA PENGANTAR	iii
DAFTAR ISI	v
1. KONSEP DASAR STATISTIKA	1
A. Pengertian Statistik dan Statistika	1
B. Pembagian Statistika	3
C. Data Statistik	8
D. Jenis Data	10
E. Konversi Skala Data	17
F. Sumber Data dan Teknik Pengumpulan Data	27
2. PENYAJIAN DATA	37
A. Penyajian Data Kualitatif	37
B. Penyajian Data Kuantitatif	41
C. Penyusunan Tabel Distribusi Frekuensi	43
3. UKURAN PEMUSATAN	55
A. Ukuran Pemusatan	55
B. Ukuran Lokasi	67
C. Ukuran Dispersi	74

4. PELUANG DAN DISTRIBUSINYA	83
A. Peluang	84
B. Distribusi Peluang	92
5. SAMPLING DAN DISTRIBUSINYA	117
A. Sampling	117
B. Distribusi Sampling	132
6. PENGUJIAN HIPOTESIS	145
A. Komponen-komponen dalam Hipotesis Statistik	145
B. Uji Hipotesis untuk Mean Satu Populasi	149
C. Uji Hipotesis untuk Mean Dua Populasi	159
D. Uji Hipotesis untuk Proporsi Satu Populasi	164
E. Uji Hipotesis untuk Proporsi Dua Populasi	168
F. Uji Hipotesis untuk Variansi Satu Populasi	172
G. Uji Hipotesis untuk Variansi Dua Populasi	176
7. REGRESI LINEAR	183
A. Regresi Linier Sederhana	184
B. Regresi Linier Berganda	195
8. ANALISIS KORELASI	211
A. Korelasi	212
B. Uji Hipotesis Koefisien Korelasi	217
C. Uji Korelasi Spearman	220
9. PENELUSURAN TEKNIK ANALISA DATA	231
A. Pertimbangan Utama dalam Pemilihan Teknik Analisis	232
B. Karakteristik Permasalahan Penelitian	233
C. Karakteristik Data Penelitian	244

Daftar Isi

D. Karakteristik Sampel Penelitian	246
E. Karakteristik dan Hubungan Antar Variabel	248
LAMPIRAN	253
DAFTAR PUSTAKA	259
INDEKS	265
BIODATA PENULIS	271

1

Konsep Dasar Statistika

Perkembangan statistika telah mempengaruhi hampir di setiap aspek kehidupan manusia modern. Statistika sangat diperlukan bukan hanya dalam penelitian atau riset saja, tetapi juga dalam bidang pengetahuan lainnya seperti: teknik, industri, ekonomi, astronomi, asuransi, pertanian, bisnis, psikologi, farmasi, ekologi, pengetahuan alam, pengetahuan sosial dan sebagainya. Selanjutnya kita akan membahas metode statistika, bukan statistika teoritik. Disini tidak diuraikan penurunan rumus-rumus ataupun bukti-bukti sifat atau dalil, melainkan uraian teknik statistik untuk penggunaan di dalam bermacam-macam disiplin ilmu.

1.1. PENGERTIAN STATISTIK DAN STATISTIKA

Secara etimologis kata *statistika* berasal dari kata *status* (bahasa Latin) yang mempunyai persamaan arti dengan kata *state* (bahasa Inggris) atau kata *staat* (bahasa Belanda), dan yang dalam bahasa Indonesia diterjemahkan menjadi *negara*. Dalam kamus bahasa Inggris akan dijumpai kata *statistics* dan kata *statistic*. Kedua kata itu mempunyai arti yang berbeda. Kata *statistics* artinya ilmu statistika, sedang kata *statistic* diartikan sebagai “ukuran yang diperoleh atau berasal dari sampel“ yaitu sebagai lawan dari kata “parameter “ yang berarti ukuran yang diperoleh atau berasal dari populasi”. Ditinjau dari segi terminologi, dewasa ini apabila kita membaca atau mendengar istilah *statistik* maka, dalam istilah statistik itu mengandung berbagai macam pengertian, antara lain statistik adalah kumpulan data, bilangan maupun non-bilangan yang disusun dalam tabel dan atau diagram, yang melukiskan atau menggambarkan suatu persoalan. Dari pengertian tersebut, statistik diartikan dalam arti sempit, yaitu keterangan ringkas berbentuk angka-angka. Contoh: Statistik penduduk, yang berarti keterangan mengenai penduduk berupa angka-angka dalam bentuk ringkas, seperti jumlah penduduk dan rata-rata umur penduduk. Selain itu statistik juga mengandung pengertian lain, yakni dipakai untuk menyatakan *ukuran* sebagai wakil dari kumpulan data mengenai sesuatu hal. Contohnya, bila kita meneliti nilai ujian pelajaran

matematika 40 siswa MTs. Diperoleh rata-rata nilainya adalah 75, maka rata-rata 75 ini dinamakan statistik.

Statistika sudah diartikan dalam arti yang luas yaitu merupakan suatu pengetahuan, sekumpulan konsep dan metode yang digunakan untuk mengumpulkan dan menginterpretasi data tentang bidang kegiatan tertentu dan mengambil kesimpulan dalam situasi dimana ada ketidakpastian dan variasi.

Secara garis besar statistika dibedakan menjadi dua yaitu statistika deskriptif dan statistika inferensial. Statistika deskriptif mempelajari metode meringkaskan dan menggambarkan segi-segi yang sangat penting dari data. Sedangkan statistika inferensial mempelajari metode mengevaluasi informasi yang terkandung dalam data dan penafsiran tentang pengetahuan baru yang diperoleh dari informasi itu.

Karena statistika merupakan suatu metodologi ilmiah, yang merupakan cabang dari matematika terapan. Maka metode-metodenya adalah berbagai macam teknik mengumpulkan, mengorganisasikan, mentabulasi, menyajikan, menganalisis, menginterpretasikan, menggambarkan dan menyajikan data dalam bentuk angka-angka. Oleh karena itu, Metode Statistik merupakan prosedur-prosedur yang digunakan dalam mengumpulkan, menyajikan, menganalisis dan menafsirkan data. Metode statistik dapat mengubah informasi yang sederhana sekalipun menjadi informasi yang bermakna.

1.2. PEMBAGIAN STATISTIKA

Berdasarkan kriteria-kriteria tertentu, statistika dapat dibagi atas beberapa macam seperti cara pengolahan data, ruang lingkup penggunaan atau disiplin ilmu yang menggunakannya, dan bentuk parameternya.

A. *Berdasarkan Cara Pengolahan Datanya*

Didasarkan atas cara pengolahan datanya, statistika dapat dibagi dua, yaitu statistika deskriptif dan statistika inferensial.

1. Statistika Deskriptif

Statistika deskriptif atau statistika deduktif adalah statistika yang berkenaan dengan metode atau cara mendeskripsikan, menggambarkan, menjabarkan, atau menguraikan data sehingga mudah dipahami. Statistika deskriptif mengacu pada bagaimana menata atau mengorganisasi data, menyajikan, dan menganalisis data. Menata, menyajikan, dan menganalisis data dapat dilakukan dengan menentukan nilai rata-rata hitung, median, modus, standar deviasi, dan persen/proporsi. Cara lain untuk menggambarkan data adalah dengan membuat tabel, distribusi frekuensi, dan diagram atau grafik.

Contoh:

1. Hasil ujian tengah semester Program Studi Pendidikan Agama Islam mata kuliah Statistika Pendidikan adalah dengan nilai rata-rata 75 dan standar deviasi 15.
2. Sebanyak 50% di antara semua pasien yang menerima suntikan obat tertentu, ternyata kemudian menderita efek samping obat tersebut.

Penarikan kesimpulan pada statistika deskriptif (jika ada) hanya ditujukan pada kumpulan data yang ada. Didasarkan atas ruang lingkup bahasannya, statistika deskriptif mencakup hal-hal berikut.

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

2. *Statistika Inferensial*

Statistika inferensial atau statistika induktif adalah statistika yang berkenaan dengan cara penarikan kesimpulan berdasarkan data yang diperoleh dari sampel untuk menggambarkan karakteristik atau ciri dari suatu populasi. Pada statistika inferensial biasanya dilakukan pengujian hipotesis dan pendugaan mengenai karakteristik atau ciri dari suatu populasi, seperti mean dan standar deviasi. Berikut ini contoh-contoh pernyataan yang termasuk dalam cakupan statistika inferensial.

- a. Akibat penurunan produksi minyak oleh negara-negara penghasil minyak dunia, diramalkan harga minyak akan menjadi dua kali lipat pada tahun yang akan datang.
- b. Dengan mengasumsikan bahwa penurunan jumlah pendaftar mahasiswa baru kurang dari 30% akibat kesalahan system pendaftaran online, maka akan terjadi pengurangan subsidi pendanaan dari Kementerian Agama.

Penarikan kesimpulan pada statistik inferensial ini merupakan generalisasi dari suatu populasi berdasarkan data (sampel) yang ada. Didasarkan atas ruang lingkup bahasannya, maka statistik inferensial mencakup:

- 1) probabilitas atau teori kemungkinan
- 2) distribusi teoritis
- 3) sampling dan distribusi sampling
- 4) pendugaan populasi atau teori populasi
- 5) uji hipotesis
- 6) analisis korelasi dan uji signifikansi, dan

7) analisis regresi untuk peramalan.

Dengan demikian, statistika inferensial sebenarnya merupakan kelanjutan dari statistika deskriptif.

B. Berdasarkan Penggunaannya

Berdasarkan penggunaannya atau disiplin ilmu yang menggunakannya, statistika dapat dibagi atas beberapa macam antara lain:

1. Statistika sosial

Statistika sosial adalah statistika yang digunakan dalam ilmu-ilmu sosial.

2. Statistika pendidikan

Statistika pendidikan adalah statistika yang digunakan dalam ilmu dan bidang pendidikan.

3. Statistika ekonomi

Statistika ekonomi adalah statistika yang digunakan dalam ilmu-ilmu ekonomi.

4. Statistika perusahaan

Statistika perusahaan adalah statistika yang digunakan dalam bidang perusahaan.

5. Statistika pertanian

Statistika pertanian adalah statistika yang digunakan dalam ilmu-ilmu pertanian.

6. Statistika kesehatan

Statistika kesehatan adalah statistika yang digunakan dalam bidang kesehatan.

C. Berdasarkan Bentuk Parameternya

Didasarkan atas bentuk parameternya (data yang sebenarnya), statistika dapat dibagi dua, yaitu statistika parametrik dan statistika nonparametrik.

1. Statistika Parametrik

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

2. Statistika Nonparametrik

Statistika nonparametrik adalah bagian statistika yang parameter dari populasinya tidak mengikuti suatu distribusi tertentu atau memiliki distribusi yang bebas dari persyaratan, dan variansnya tidak perlu homogen. Dengan perkataan lain uji statistik non-*parametrik* adalah uji yang modelnya tidak menetapkan syarat-syarat mengenai parameter-parameter populasi yang merupakan induk sampel penelitiannya.

Perlu disadari bahwa uji nonparametrik selayaknya tidak digunakan apabila uji parametrik dapat diterapkan, karena tingkat kemampuan uji nonparametrik lebih rendah dari pada uji parametrik. Namun anda sebagai

pengambil keputusan atau peneliti jangan salah menafsirkan bahwa derajat kegunaan metode statistik nonparametrik dibawah metode statistik parametrik. Tentu saja tidak demikian, masing-masing metode dibuat dengan spesifikasi khusus sesuai dengan macam data yang digunakan. Peningkatan kemampuan uji nonparametrik harus dengan memperbesar sampel. Namun seperti kita ketahui memperbesar sampel berarti akan menambah biaya, waktu, dll. Penjelasan mengenai apa dan bagaimana menggunakan metode-metode statistik parametrik maupun nonparametrik akan dijelaskan pada Bab yang lain.

1.3. DATA STATISTIK

Data berbentuk jamak, sedang datum berbentuk tunggal. Jadi data sama dengan datum-datum. Data ialah suatu bahan mentah yang jika diolah dengan baik melalui berbagai analisis dapat melahirkan berbagai informasi. Dengan informasi tersebut, kita dapat mengambil suatu keputusan. Apa yang disebut dengan data statistik? Data statistik adalah keterangan atau ilustrasi mengenai sesuatu hal. Data dapat berbentuk kategori, misalnya : rusak, baik, senang, puas, berhasil, gagal, dan sebagainya, atau dapat pula data merupakan suatu kumpulan angka (bilangan). Data jika dipandang dari bentuknya dapat dibedakan menjadi data kuantitatif dan data kualitatif. Data kuantitatif yaitu data yang berupa bilangan dan harganya berubah-ubah atau bersifat variabel. Sedangkan data kualitatif (atribut) yaitu data yang dikategorikan menurut lukisan kualitas obyek yang dipelajari. Dalam statistika dikenal istilah-istilah jenis data, tingkatan data, sumber data, penyajian data, analisis data. Data dianalisis sesuai jenis dan tingkatannya, karena itu masing-masing tingkatan data mempunyai analisis sendiri khususnya dalam analisis korelasi. Data yang baik tentu saja harus mutakhir, cocok dengan masalah penelitian dari sumber yang dapat dipertanggungjawabkan, lengkap, akurat, objektif, dan konsisten. Pengumpulan data sedapat mungkin diperoleh dari tangan pertama. Data yang baik sangat diperlukan dalam penelitian, sebab bagaimanapun canggihnya suatu analisis data jika tidak ditunjang oleh data yang baik, maka hasilnya kurang dapat dipertanggungjawabkan.

Kita sudah mengetahui bahwa data adalah hasil pengukuran suatu variabel. Kita mengambil contoh jenis kelamin, data yang hanya dua kemungkinan, yaitu laki-laki atau perempuan. Data ini berupa *atribut* dan bukan angka. Data semacam ini disebut *data kualitatif*. Selanjutnya, kita memperhatikan variabel umur yang datanya bisa berbentuk angka, misalnya 30 tahun, dan ini disebut *data kuantitatif*.

1.4. JENIS DATA

Walaupun hampir tidak terhitung banyaknya gejala yang dapat diukur dalam sebuah penelitian, pada umumnya hasil pengukuran itu dapat dibedakan ke dalam satu di antara dua kelompok data, yaitu *data kuantitatif* dan *data kualitatif*. Data kuantitatif adalah hasil pengamatan yang diukur dalam skala numerik (bilangan), misalnya rata-rata penjualan buku pelajaran tiap semester, persentase siswa yang tidak naik kelas dalam sebuah sekolah, tingkat kepuasan orang tua terhadap pelayanan sekolah, banyaknya kerusakan alat-alat laboratorium tertentu setiap tahun, dan banyaknya kwh (*kilo watt hour*) listrik terpakai setiap hari.

Selanjutnya, semua data yang bukan kuantitatif disebut *data kualitatif*. Jadi, data kualitatif adalah hasil pengamatan yang berbentuk kategori dan bukan bilangan. Dalam hal ini, setiap pengukuran sebuah data jatuh ke dalam satu dan hanya satu dari dua kategori atau kelompok data tersebut. Beberapa contoh pengelompokan diberikan sebagai berikut.

1. Pekerjaan dari orang tua atau wali murid siswa dari sebuah sekolah. Setiap orang tua hanya mempunyai satu jenis pekerjaan yang sesuai dengan KTPnya.
2. Jenis bahan bakar yang terakhir dibeli oleh seorang pengendara mobil pada hari itu di satu pompa bensin. Lagi, setiap pencatatan hanya jatuh dalam satu dan hanya satu kategori, misalnya, bensin, solar, atau jenis bahan bakar lainnya.
3. Suatu program studi di universitas memiliki peringkat akreditasi. Setiap program studi hanya bisa masuk ke dalam satu kategori A, B, atau C, tidak mungkin masuk dalam lebih dari satu kategori.

Perhatikan bahwa semua contoh di atas memberikan data yang bukan bilangan, melainkan pengukuran kualitatif.

Perbedaan intensitas dari sebuah variabel kualitatif disebut *level*, dan ini bukan bilangan. Level itu dapat digambarkan sebagai berikut.

1. Cara memberi nilai ujian yang terdiri dari lima jenis, yaitu A, B, C, D dan E. Jadi variabel nilai mempunyai lima level.
2. Jenis kelamin adalah variabel kualitatif yang mempunyai dua level, karena hanya dapat dijelaskan dengan atribut *perempuan* dan *laki-laki*.
3. Andaikan kita mengamati urutan jenjang jabatan akademik dosen terdiri dari guru besar, lektor kepala, lektor, dan asisten ahli. Karenanya kita dapat memberikan sebuah ukuran realistis tentang pentingnya setiap posisinya.

Hasil pengukuran data dapat dibedakan atas empat macam skala, yaitu *skala nominal*, *skala ordinal*, *skala interval*, dan *skala rasio*.

A. Skala nominal

Skala pengukuran data yang paling rendah terjadi apabila bilangan atau lambang lain digunakan untuk identitas atau mengklasifikasikan objek, orang atau benda-benda lain. Data hasil klasifikasi tersebut disebut data skala nominal. Sebagai contoh data tentang jenis pekerjaan. Data tersebut misalnya berisi jenis pekerjaan PNS dan bukan PNS. Setiap orang yang bekerja akan termasuk dalam salah satu dari dua himpunan pekerjaan tersebut. *Skala nominal* dapat dinyatakan sebagai ukuran yang tidak sebenarnya. Skor untuk setiap satuan pengamatan, atau individu hanya merupakan tanda atau simbol yang menunjukkan ke dalam kelompok atau kelas mana individu tersebut termasuk. Misalnya, *jenis kelamin* dengan skor yang mungkin "1" untuk laki-laki dan "2" untuk perempuan. Skor 1 dan 2 yang diberikan itu hanya untuk membedakan antara kelompok yang satu dengan yang lainnya. Urutan, selisih, jumlah, dan operasi hitung lainnya terhadap data skala nominal tidak mempunyai arti sehingga tidak bisa dilakukan. Kita tidak bisa mengatakan bahwa 1 lebih besar daripada 0 untuk data jenis kelamin tersebut. Dengan skala nominal, kita dapat mengelompokkan responden atau objek lain ke dalam dua kategori atau lebih menurut variabel yang **diperhatikan**.

B. Skala ordinal

Skala ordinal menunjukkan urutan (peringkat, tingkatan, atau *ranking*) disamping berfungsi sebagai pengelompokan (skala nominal). Misalnya, variabel *tingkatan* dalam suatu rumah susun dengan angka 1, 2, 3,; variabel pendidikan dengan kategori "1" di bawah SD, "2" yang tamat SD, "3" yang tamat SLTP, dan "4" yang tamat SLTA atau di atasnya. Pada skala ordinal, selisih antara dua ukuran, serta operasi hitung lainnya tidak dapat dilakukan karena tidak mempunyai arti, kecuali urutannya yang mempunyai makna. Kita tidak bisa mengatakan bahwa jarak antara 2 dan 3 sama dengan jarak antara 3 dan 4, karena perbedaan antara tamatan SD dan tamatan SLTP tidak sama dengan perbedaan antara tamatan SLTP dan tamatan SLTA ke atas. Skala ordinal ini memungkinkan peneliti untuk mengurutkan respondennya dari tingkatan *paling rendah* ke tingkatan *paling tinggi* atau sebaliknya menurut suatu atribut tertentu. Misalnya, ukuran untuk kelas ekonomi biasanya dipakai ukuran ordinal, yakni kelas ekonomi tingkat atas (skor 3), kelas ekonomi tingkat menengah (skor 2), dan kelas ekonomi tingkat bawah (skor 1). Ukuran ini tidak menunjukkan angka rerata kelas ekonomi, dan tidak memberikan informasi mengenai besar interval atau jarak antara kelas ekonomi rendah dan kelas ekonomi atas. Karena itu, perhitungan statistik yang didasarkan atas perhitungan *rerata* dan *simpangan baku* tidak dapat diterapkan pada data ukuran ordinal. Demikian pula, kita tidak dapat

mengatakan bahwa kelas ekonomi atas (skor 3) tiga kali lebih kaya daripada kelas ekonomi bawah (skor 1), demikian hula tidak dapat dikatakan bahwa kelas ekonomi menengah (skor 2) dua kali lebih kaya daripada kelas ekonomi bawah. Namun, skala ordinal sudah beranjak meninggalkan kelas data kualitatif dan mulai masuk ke kelas data kuantitatif.

C. Skala interval

Skala interval termasuk ukuran yang bersifat numerik, yaitu interval antara dua ukuran yang berbeda mempunyai arti. Misalnya, Amat memperoleh nilai 50 dalam mata pelajaran matematika, sedangkan Badu mendapat nilai 10, maka tidak dapat dinyatakan bahwa pengetahuan amat lima kali pengetahuan Badu dalam mata pelajaran matematika. Walau pun demikian, kita dapat mengatakan bahwa Amat mendapat skor 40 lebihnya dari skor Badu. Jadi, pada skala interval tidak terdapat titik nol mutlak, misalnya Kasim mendapat nilai nol pada ujian matematika tidak berarti ia tidak memiliki pengetahuan sama sekali. Tetapi operasi jumlah dan kurang dapat dilakukan terhadap data skala interval dan mempunyai makna. Dengan demikian, skala interval sudah jelas masuk kelas data kuantitatif yang lebih tinggi tingkatan kuantitatifnya dibandingkan dengan kelas skala ordinal.

Untuk menentukan apakah perbedaan pangkat atau kedudukan sosial sama atau tidak adalah sesuatu yang sulit. Misalnya, apakah perbedaan pangkat antara kapten dan letnan sama dengan perbedaan pangkat antara mayor dan kapten, tentulah tidak mudah kita katakan. Dalam pengukuran skala ordinal masalah “perbedaan dalam jarak” atau interval antara dua titik skala tidak diperhatikan, namun ada skala-skala yang jarak antara dua titik skala memang diketahui. Dalam hal ini dikatakan bahwa pengukuran telah tercapai dalam skala interval. Skala pengukuran interval lebih kuat dibandingkan dengan skala ordinal karena dapat mengetahui jarak (interval) antara dua kelas yang berbeda. Sebagai contoh, skala untuk mengukur temperatur. Kita mengenal dua skala yang banyak digunakan dalam pengukuran temperatur, yaitu Celcius dan Fahrenheit. Unit pengukuran dan titik nol dalam mengukur temperatur adalah sembarang. Namun demikian, kedua skala itu memuat informasi yang sama, sehingga dapat dituliskan dalam perbandingan :

$$F = 9/5 C + 32.$$

Dapat ditunjukkan bahwa perbandingan selisih-selisih temperatur (interval) adalah independen dengan unit pengukuran dan titik nol.

Misalkan kita memiliki perbandingan sebagai berikut :

Celcius	0	10	30	50	100
Fahrenheit	32	50	86	122	212

Perhatikan bahwa perbandingan selisih-selisih antara skala temperatur dalam skala yang satu sama dengan perbandingan antara selisih-selisih yang ekuivalen pada skala yang lain.

Pada skala Celcius, $\frac{(30-10)}{(10-0)} = 2$ dan pada skala Fahrenheit perbandingan

yang ekuivalen adalah $\frac{(86-50)}{(50-32)} = 2$. Dengan kata lain, dalam skala

interval, perbandingan setiap dua interval adalah independen dengan unit pengukuran dan titik nol yang digunakan.

D. Skala rasio

Skala rasio sedikit berbeda dengan skala interval, yakni skala rasio mempunyai titik nol mutlak. Sebagai contoh, variabel umur dalam bulan, tinggi badan dalam meter, luas wilayah dalam kilometer persegi, dan penghasilan dalam rupiah. Jika Ali mempunyai uang Rp 300,00, dan Bakri mempunyai uang Rp 100,00, maka uang Ali sama dengan tiga kali uang Bakri. Kayu yang panjangnya 10 meter adalah dua kali lebih panjang daripada kayu yang panjangnya 5 meter. Sifat ini tidak berlaku pada skala interval, tetapi semua sifat-sifat skala interval juga berlaku untuk skala rasio.

Dari pengetahuan kita tentang skala pengukuran, variabel dapat pula dibedakan atau dikelompokkan menjadi *variabel nominal*, *variabel ordinal*, *variabel interval*, dan *variabel rasio* menurut skala pengukurannya. Variabel nominal dan variabel ordinal biasa disebut *variabel kualitatif* atau *variabel kategori*, sedangkan variabel interval dan variabel rasio disebut *variabel kuantitatif*. Kedua kelompok variabel ini memberikan pengaruh yang berbeda dalam pemilihan teknik analisis data.

Setelah membicarakan data dan skala pengukurannya, kita dapat mengelompokkan variabel menjadi variabel kuantitatif dan variabel kualitatif. Karena itu, kita perlu membahas pengertian kedua jenis variabel tersebut.

Jika suatu skala kecuali memiliki sifat skala interval juga masih memiliki sifat lain, yaitu titik nolnya tertentu, maka skala seperti ini dinamakan skala rasio. Dalam skala rasio, perbandingan dua titik skala sembarang adalah independen dengan unit pengukuran. Contoh skala rasio adalah skala untuk mengukur berat, panjang, isi, dan sebagainya. Skala-skala ini mempunyai titik nol yang berarti. Titik nol yang tidak dipilih

sembarang disebut murni atau asli (absolut). Skala rasio dengan titik nol yang murni dapat menentukan rasio antara dua titik skala dengan jelas.

Dalam skala rasio dapat dikatakan, misalnya “berat badan mahasiswa A dua kali berat badan mahasiswa B”, atau “banyaknya mahasiswa yang hadir hari ini tiga kali lebih banyak daripada hari kemarin “, dan masih banyak contoh yang lain.

1.5. KONVERSI SKALA DATA

Untuk bisa bekerja dengan statistika, semua data dinyatakan dalam bentuk angka atau bilangan, termasuk data kualitatif. Pekerjaan mengkuantitatifkan data adalah suatu proses pengukuran.

Skala interval :merupakan skala pengukuran yang banyak digunakan untuk mengukur fenomena/gejala sosial, dimana pihak responden diminta melakukan rangking terhadap preferensi tertentu sekaligus memberikan nilai (rate) terhadap preferensi tersebut. Jenis skala yang dapat digunakan untuk penelitian sosial, yaitu : Skala Linkert, Skala Guttman, Skala Bogardus, Skala Thurstone, Skala Semantic, Skala Stipel, Skala Paired-Comparison, Skala rank-Order. Kedelapan macam teknik skala tersebut bila digunakan dalam pengukuran, akan mendapatkan data interval, atau rasio. Hal ini tergantung pada bidang yang akan diukur.

Dalam penelitian kuantitatif, peneliti akan menggunakan instrumen untuk mengumpulkan data penelitian. Instrumen penelitian ini digunakan untuk meneliti variabel yang diteliti. Dengan demikian jumlah instrumen yang akan digunakan untuk penelitian tergantung pada jumlah variabel yang diteliti. Instrumen-instrumen penelitian sudah ada yang dibekukan, tapi ada yang harus dibuat peneliti sendiri. Karena instrumen penelitian akan digunakan untuk melakukan pengukuran dengan tujuan menghasilkan data kuantitatif yang akurat, maka setiap instrumen harus mempunyai skala.

A. Skala Likert

Skala Likert ialah skala yang dapat dipergunakan untuk mengukur sikap, pendapat, dan persepsi seseorang atau sekelompok orang tentang suatu gejala atau fenomena pendidikan. Skala Likert adalah suatu skala psikometrik yang umum digunakan dalam kuesioner, dan merupakan skala yang paling banyak digunakan dalam riset berupa survei. Nama skala ini diambil dari nama Rensis Likert, yang menerbitkan suatu laporan yang menjelaskan penggunaannya. Sewaktu menanggapi pertanyaan dalam skala Likert, responden menentukan tingkat persetujuan mereka terhadap suatu pernyataan dengan memilih salah satu dari pilihan yang tersedia. Ada dua bentuk pertanyaan yang menggunakan Likert yaitu pertanyaan

positif untuk mengukur minat positif, dan bentuk pertanyaan negatif untuk mengukur minat negatif. Pertanyaan positif diberi skor 5, 4, 3, 2, dan 1; sedangkan bentuk pertanyaan negatif diberi skor 1, 2, 3, 4, dan 5. Bentuk jawaban skala Likert terdiri dari sangat setuju, setuju, ragu-ragu, tidak setuju, dan sangat tidak setuju.

Penskalaan ini apabila dikaitkan dengan jenis data yang dihasilkan adalah data Ordinal. Selain pilihan dengan lima skala seperti contoh di atas, kadang digunakan juga skala dengan tujuh atau sembilan tingkat. Suatu studi empiris menemukan bahwa beberapa karakteristik statistik hasil kuesioner dengan berbagai jumlah pilihan tersebut ternyata sangat mirip. Skala Likert merupakan metode skala bipolar yang mengukur baik tanggapan positif ataupun negatif terhadap suatu pernyataan. Empat skala pilihan juga kadang digunakan untuk kuesioner skala Likert yang memaksa orang memilih salah satu kutub karena pilihan "netral" tak tersedia. Selain pilihan dengan lima skala seperti contoh di atas, kadang digunakan juga skala dengan tujuh atau sembilan tingkat. Suatu studi empiris menemukan bahwa beberapa karakteristik statistik hasil kuesioner dengan berbagai jumlah pilihan tersebut ternyata sangat mirip. Skala Likert merupakan metode skala bipolar yang mengukur baik tanggapan positif ataupun negatif terhadap suatu pernyataan. Empat skala pilihan juga kadang digunakan untuk kuesioner skala Likert yang memaksa orang memilih salah satu kutub karena pilihan "netral" tak tersedia.

Mengapa data Likert termasuk data ordinal? Sebenarnya alasannya gampang. Sangat setuju pasti lebih tinggi daripada yang setuju. Yang setuju pasti lebih tinggi daripada yang netral, yang netral pasti lebih tinggi daripada yang tidak setuju, sedangkan yang tidak setuju pasti lebih tinggi daripada yang sangat tidak setuju. Namun jarak antara sangat setuju ke setuju dan dari setuju ke netral dan seterusnya tentunya tidak sama, oleh karena itu data yang dihasilkan oleh skala likert adalah data ordinal. Sedangkan cara scoring bahwa Sangat setuju 5, setuju 4, netral 3, tidak setuju 2 dan sangat tidak setuju 1 hanya merupakan kode saja untuk mengetahui mana yang lebih tinggi dan mana yang lebih rendah. Dari cara scoring tersebut kita tidak bisa memaknai bahwa sangat setuju adalah netral ditambah setuju. Tapi permasalahannya sesuai dengan ciri-ciri dari data ordinal, bahwa data ordinal belum bisa dikenai operasi matematis, mengapa pada saat scoring dari skala likert kita menjumlahkan skor di tiap-tiap item padahal jelas-jelas skala data ordinal tidak bisa dijumlahkan?

Prosedur dalam membuat skala Likert adalah sebagai berikut:

1. Peneliti mengumpulkan item-item yang cukup banyak, relevant dengan masalah yang sedang diteliti, dan terdiri dari item yang cukup jelas disukai dan tidak disukai

2. Kemudian item-item itu dicoba kepada sekelompok responden yang cukup representatif dari populasi yang ingin diteliti
3. Responden di atas diminta untuk mengecek tiap item, apakah ia menyenangkan (+) atau tidak menyukainya (-). Respons tersebut dikumpulkan dan jawaban yang memberikan indikasi menyenangkan diberi skor tertinggi. Tidak ada masalah untuk memberikan angka 5 untuk yang tertinggi dan skor 1 untuk yang terendah atau sebaliknya. Yang penting adalah konsistensi dari arah sikap yang diperlihatkan. Demikian juga apakah jawaban '\setuju\' atau '\tidak setuju\' disebut yang disenangi, tergantung dari isi pertanyaan dan isi dari item-item yang disusun
4. Total skor dari masing-masing individu adalah penjumlahan dari skor masing-masing item dari individu tersebut.
5. Respon dianalisis untuk mengetahui item-item mana yang sangat nyata batasan antara skor tinggi dan skor rendah dalam skala total. Misalnya, responden pada upper 25% dan lower 25% dianalisis untuk melihat sampai berapa jauh tiap item dalam kelompok ini berbeda. Item-item yang tidak menunjukkan beda yang nyata, apakah masuk dalam skoring tinggi atau rendah juga dibuang untuk mempertahankan konsistensi internal dari pertanyaan.

Kelemahan skala Likert:

1. Karena ukuran yang digunakan adalah ukuran ordinal, skala Likert hanya dapat mengurutkan individu dalam skala, tetapi tidak dapat membandingkan berapa kali satu individu lebih baik dari individu yang lain.
2. Kadangkala total skor dari individu tidak memberikan arti yang jelas, karena banyak pola respons terhadap beberapa item akan memberikan skor yang sama.

B. Skala Thurstone

Skala Thurstone merupakan skala yang disusun dengan memilih butir yang berbentuk skala interval. Setiap skor memiliki kunci skor dan jika diurut kunci skor menghasilkan nilai yang berjarak sama.

Metode ini mencoba menempatkan sikap seseorang pada rentangan kontinum dari yang sangat unfavorabel hingga sangat fafobel terhadap suatu obyek sikap. Caranya dengan memberikan orang tersebut sejumlah aitem sikap yang telah ditentukan derajat favorabilitasnya. Tahap yang paling kritis dalam menyusun alat ini seleksi awal terhadap pernyataan sikap dan penghitungan ukuran yang mencerminkan derajat favorabilitas

dari masing-masing pernyataan. Derajat (ukuran) favorabilitas ini disebut nilai skala.

Untuk menghitung nilai skala dan memilih pernyataan sikap, pembuat skala perlu membuat sampel pernyataan sikap sekitar lebih 100 buah atau lebih. Pernyataan-pernyataan itu kemudian diberikan kepada beberapa orang penilai (judges). Penilai ini bertugas untuk menentukan derajat favorabilitas masing-masing pernyataan. Favorabilitas penilai itu diekspresikan melalui titik skala rating yang memiliki rentang 1-11. Sangat tidak setuju 1 2 3 4 5 6 7 8 9 10 11 Sangat setuju Tugas penilai ini bukan untuk menyampaikan setuju tidaknya mereka terhadap pernyataan itu. Median atau rerata perbedaan penilaian antar penilai terhadap item ini kemudian dijadikan sebagai nilai skala masing-masing item. Pembuat skala kemudian menyusun item mulai dari item yang memiliki nilai skala terendah hingga tertinggi. Dari item-item tersebut, pembuat skala kemudian memilih item untuk kuesioner skala sikap yang sesungguhnya. Dalam penelitian, skala yang telah dibuat ini kemudian diberikan pada responden. Responden diminta untuk menunjukkan seberapa besar kesetujuan atau ketidaksetujuannya pada masing-masing item sikap tersebut.

Teknik ini disusun oleh Thurstone didasarkan pada asumsi-asumsi: ukuran sikap seseorang itu dapat digambarkan dengan interval skala sama. Perbedaan yang sama pada suatu skala mencerminkan perbedaan yang sama pula dalam sikapnya. Asumsi kedua adalah Nilai skala yang berasal dari rating para penilai tidak dipengaruhi oleh sikap penilai terhadap isu. Penilai melakukan rating terhadap item dalam tataran yang sama terhadap issue tersebut.

Pernyataan yang diajukan kepada responden disarankan oleh Thurstone untuk tidak terlalu banyak, diperkirakan antara 5 sampai 10 butir pertanyaan atau pernyataan. Pembuatan skala Thurstone dapat dilakukan dengan langkah-langkah seperti berikut.

1. Mengumpulkan sejumlah pernyataan yang merepresentasikan secara luas perbedaan tingkat, disenangi, netral, dan tidak disenangi terhadap suatu objek atau subjek yang hendak diteliti.
2. Pernyataan ini diberikan pada sejumlah responden misal 50 orang atau lebih yang cukup mengenal terhadap objek atau subjek agar dapat memilih ke dalam tingkatan kategori. Kategori A terdiri atas pernyataan yang dianggap disenangi atau favorit, E F netral, dan J K merupakan kategori tidak disenangi atau tidak favorit.
3. Klasifikasi pernyataan ke dalam kategori, dengan pertimbangan penilaian terhadap objek atau subjek secara psikologis, tetapi hanya merefleksikan persepsi mereka terhadap kategori pernyataan yang disediakan.

4. Pernyataan yang nilainya menyebar dibuang, dan pernyataan yang mempunyai nilai bersamaan digunakan untuk pembuatan skala.

Skor tinggi pada skala berarti mereka memiliki tingkat prasangka terhadap sifat yang ingin diteliti. Skor terendah berarti responden mempunyai sifat favorit terhadap sifat yang ingin diteliti.

Skala Thurstone tidak terlalu banyak digunakan sebagai instrumen di bidang pendidikan karena model ini mempunyai beberapa kelemahan yang di antaranya seperti berikut.

1. Memerlukan terlalu banyak pekerjaan untuk membuat skala.
2. Nilai pada skala yang telah dibuat memungkinkan pada skor sama mempunyai sikap berbeda.
3. Nilai yang dibuat dipengaruhi oleh sikap para juri atau penilai.
4. Memerlukan tim penilai yang objektif.

C. Skala Guttman

Skala Guttman dikembangkan oleh Louis Guttman. Skala ini mempunyai ciri penting, yaitu merupakan skala kumulatif dan mengukur satu dimensi saja dari satu variabel yang multidimensi. Skala Guttman yaitu skala yang menginginkan jawaban tegas seperti jawaban benar-salah, ya-tidak, pernah-tidak pernah. Untuk jawaban positif seperti setuju, benar, pernah dan semacamnya diberi skor 1; sedangkan untuk jawaban negatif seperti tidak setuju, salah, tidak, tidak pernah, dan semacamnya diberi skor 0. Dengan skala ini, akan diperoleh jawaban yang tegas yaitu Ya - Tidak, Benar - Salah dan lain-lain. Penelitian menggunakan skala Guttman dilakukan bila ingin mendapatkan jawaban yang tegas terhadap suatu permasalahan yang ditanyakan. Ketika membuat skala kumulatif, seorang peneliti harus menentukan, pertama, apakah semua item membentuk skala berdimensi tunggal. Untuk mencapai hal tersebut, perlu dapat menganalisis reproduksi jawaban, yaitu proporsi prediksi kemudian dibuat dengan menggunakan jawaban item-item utama. Kemudian bentuk jawaban yang sebenarnya dipelajari, dan pengukuran dibuat dengan mempertimbangkan respons yang reproduktif terhadap skor total. Skala Guttman mungkin merupakan teknik skala pengukuran yang paling populer dan banyak digunakan pada penelitian sosial.

Skala ini dapat pula dibentuk dalam bentuk checklist atau pilihan ganda. Skor 1 untuk skor tertinggi dan skor 0 untuk terendah. (Analisa seperti pada skala Likert).

Skala Guttman sering pula disebut sebagai teknik kumulatif. Guttman mengembangkan teknik ini guna mengatasi problem yang dihadapi oleh Likert maupun Thurstone. Di samping itu, skala Guttman mempunyai

asumsi bahwa dasar dari fakta di mana beberapa item di bawah pertimbangan yang harus dibuktikan menjadi petunjuk kuat satu variabel dibanding variabel lainnya.

Teknik tersebut dilihat dari sifat-sifatnya sebagai skala yang memiliki dimensi tunggal. Tujuan utama pembuatan skala model ini pada prinsipnya adalah untuk menentukan, jika sikap yang diteliti benar-benar mencakup satu dimensi. Sikap dikatakan berdimensi tunggal bila sikap tersebut menghasilkan skala kumulatif. Sebagai contoh, jika seorang responden yang setuju terhadap item 2, maka ia berarti juga setuju terhadap item nomor 1, sedangkan seorang responden yang setuju dengan item 3 juga berarti ia setuju pada item nomor 2 dan 1 dan seterusnya. Dengan kata lain, seseorang yang setuju pada item tertentu dalam tipe skala akan mempunyai skor yang lebih tinggi pada skala total daripada seseorang yang tidak setuju pada item tersebut.

Responden, sebagai contohnya ditanyakan tentang apakah setuju atau tidak terhadap peran organisasi guru dan orang tua.

- a. Asosiasi guru-orang tua murid mempunyai peran penting dalam perkembangan sekolah.
- b. Asosiasi guru-orang tua murid mempunyai pengaruh kuat terhadap perkembangan sekolah.
- c. Asosiasi guru-orang tua murid merupakan organisasi penting untuk meningkatkan kualitas sekolah.

1.6. SUMBER DATA DAN TEKNIK PENGUMPULAN DATA

Ada beberapa pertanyaan yang sering muncul dalam awal pembelajaran statistika. Bagaimana data diperoleh ? Apa bedanya populasi dengan sampel? Populasi adalah kumpulan individu sejenis yang berada pada wilayah tertentu dan pada waktu yang tertentu pula. Kata populasi (dari bahasa Inggris: Population) dalam bidang statistika berarti sekumpulan data yang menjadi objek inferensi. Populasi adalah keseluruhan pengamatan yang menjadi perhatian kita baik yang berhingga maupun tak berhingga jumlahnya. Nilai-nilai ukuran dalam populasi disebut sebagai parameter. Seringkali tidak praktis mengambil data dari keseluruhan populasi untuk menarik suatu kesimpulan. Untuk itu dilakukan pengambilan sampel yaitu sebagian atau himpunan bagian dari populasi. Sampel merupakan bagian dari populasi yang ingin diteliti; dipandang sebagai suatu pendugaan terhadap populasi, namun bukan populasi itu sendiri. Sampel dianggap sebagai perwakilan dari populasi yang hasilnya mewakili keseluruhan gejala yang diamati. Ukuran dan keragaman sampel menjadi penentu baik tidaknya sample yang diambil. Terdapat dua cara pengambilan sampel, yaitu secara random (random) dan tidak random (non-random). Sampel yang diambil harus dapat

merepresentasikan populasi yang ada. Prosedur pengambilan sampel yang menghasilkan kesimpulan yang konsisten terlalu tinggi atau terlalu rendah mengenai suatu ciri populasi dikatakan *berbias*. Untuk menghindari kemungkinan bias ini perlu dilakukan pengambilan contoh random atau contoh random sederhana. Contoh random sederhana didefinisikan sebagai contoh yang dipilih sedemikian rupa sehingga setiap himpunan bagian yang berukuran n dari populasi mempunyai peluang terpilih yang sama.

A. Teknik Pengumpulan Data Sampel

1. Pengambilan random sederhana (*Simple random sampling*)

Merupakan sistem pengambilan sampel secara random dengan menggunakan undian atau tabel angka random. Tabel angka random merupakan tabel yang dibuat dalam komputer berisi angka-angka yang terdiri dari kolom dan baris, dan cara pemilihannya dilakukan secara bebas. Pengambilan random secara sederhana ini dapat menggunakan prinsip pengambilan sampel dengan pengembalian ataupun pengambilan sampel tanpa pengembalian.

Metode ini memiliki kelebihan mampu mengatasi bias yang muncul dalam pemilihan anggota sampel, dan kemampuan menghitung standard error. Kekurangan yang timbul akibat metode *sampling random* adalah tidak adanya jaminan bahwa setiap sampel yang diambil secara random akan merepresentasikan populasi secara tepat.

2. Pengambilan random secara sistematis

Metode pengambilan secara sistematis (*systematic random sampling*) merupakan sistem pengambilan sampel yang dilakukan dengan menggunakan selang interval tertentu secara berurutan. Misalnya, jika ingin mengambil 1000 sampel dari 5000 populasi secara acak, maka kemungkinan terpilihnya $1/5$. Diambil satu angka dari interval pertama antara angka 1-5, dan dilanjutkan dengan pemilihan angka berikutnya dari interval selanjutnya.

Kelebihan metode sistematis adalah lebih praktis dan hemat dibanding dengan pengambilan random sederhana. Kekurangan dari metode ini adalah tidak bisa digunakan pada penelitian yang heterogen karena tidak mampunya menangkap keragaman populasi heterogen.

3. Pengambilan random berdasar lapisan

Metode pengambilan random berdasarkan lapisan (*stratified random sampling*) merupakan sistem pengambilan sampel yang dibagi menurut lapisan-lapisan tertentu dan masing-masing lapisan memiliki jumlah sampel yang sama.

Kelebihan metode ini adalah lebih tepat dalam menduga populasi karena variasi pada populasi dapat terwakili oleh sampel, sedangkan kekurangannya adalah harus memiliki informasi dan data yang cukup tentang variasi populasi penelitian, kadang-kadang ada perbedaan jumlah yang besar antar masing-masing strata.

4. Pengambilan random berdasar area (*Cluster sampling*)

Metode random berdasar area (*cluster sampling*) merupakan sistem pengambilan sampel yang dibagi berdasarkan areanya. Setiap area memiliki jatah terambil yang sama. Kelebihannya adalah lebih tepat menduga populasi karena variasi dalam populasi dapat terwakili dalam sampel. Sedangkan kekurangannya adalah memerlukan waktu yang lama karena harus membaginya dalam area-area tertentu.

5. Pengambilan menurut tujuan (*Purposive sampling*)

Merupakan pemilihan anggota sampel yang didasarkan atas tujuan dan pertimbangan tertentu dari peneliti. Metode ini memiliki kelebihan yaitu tujuan dari peneliti dapat terpenuhi. Sedangkan kekurangannya adalah belum tentu mewakili keseluruhan variasi yang ada.

B. Metode Pengumpulan Data

Data dapat dikumpulkan langsung oleh peneliti melalui pihak yang disebut sumber primer. Data yang dikumpulkan oleh peneliti melalui pihak kedua atau tangan kedua disebut sumber sekunder, yaitu data yang diperoleh melalui wawancara kepada pihak lain tentang objek dan subjek yang diteliti, dan mempelajari dokumentasi-dokumentasi tentang objek dan subjek yang diteliti. Dari kedua macam sumber data tersebut, tentu saja sumber data primer lebih dapat dipertanggungjawabkan daripada data yang didapat melalui sumber sekunder. Teknik-teknik pengumpulan data dapat dilakukan melalui: wawancara (*interview*), pengamatan (*observation*), angket (*kuesioner*), dan dokumentasi (*documentation*). Wawancara dapat sistematis atau tidak sistematis. Pengamatan dapat tidak langsung atau langsung. Angket dapat tertutup atau terbuka. Peneliti dapat menggunakan salah satu atau gabungan dari tehnik-tehnik pengumpulan data di atas.

Metode yang dipilih untuk setiap variabel tergantung pada berbagai faktor terutama jenis data dan ciri responden. Untuk data historis misalnya tidak bisa ditemukan dengan observasi tetapi dimungkinkan dengan dokumenter atau wawancara. Hal ini tergantung pada karakteristik data variabel, maka metode yang digunakan tidak selalu sama untuk setiap variabel. Berikut ini adalah metode pengumpulan data suatu penelitian.

1. Observasi

Observasi atau pengamatan merupakan salah satu teknik pengumpulan data/fakta yang cukup efektif untuk mempelajari suatu sistem. Observasi adalah pengamatan langsung para pembuat keputusan berikut lingkungan fisiknya dan atau pengamatan langsung suatu kegiatan yang sedang berjalan.

Kelebihan :

- 1) Derajat kepercayaan tinggi
- 2) Konteks sosial yang diamati belum dipengaruhi faktor lain (natural)
- 3) Tidak terbatas hanya pada manusia
- 4) Dapat menggunakan alat bantu

Kelemahan :

- 1) Memerlukan waktu yang lama
- 2) Kurang efektif mengamati gejala pada individu seperti sikap, motivasi, pandangan dan sebagainya
- 3) Tidak dapat mengamati gejala yang peka / rahasia
- 4) Tidak dapat mengamati gejala masa lampau.

2. Wawancara

Menurut pengertiannya wawancara adalah teknik pengumpulan data atau informasi dari “informan” dan atau “Responden” yang sudah ditetapkan, dilakukan dengan cara ”Tanya jawab sepihak tetapi sistematis” atas dasar tujuan penelitian yang hendak di capai. Menurut beberapa ahli, wawancara juga didefinisikan sebagai berikut :

- 1) Wawancara merupakan suatu teknik pengumpulan data dengan jalan mengadakan komunikasi dengan sumber data. Komunikasi tersebut dilakukan dengan dialog (tanya jawab) secara lisan, baik langsung maupun tidak.
- 2) Wawancara adalah salah satu metode untuk mendapatkan data anak atau orang tua dengan mengadakan hubungan secara langsung dengan informan/*face to face relation*
- 3) Wawancara adalah alat untuk memperoleh data atau fakta atau informasi dari seorang murid secara lisan.
- 4) Wawancara informatif adalah suatu alat untuk memperoleh fakta/data informasi dari murid secara lisan. Dengan tujuan mendapatkan data yang diperlukan untuk bimbingan.

Tujuan wawancara.

Ada berbagai tujuan yang dapat dicapai dalam wawancara yaitu :

- 1) Menciptakan hubungan baik diantara dua pihak yang terlibat (subyek wawancara dan pewawancara). Pertemuan itu harus bebas dari segala

- kecemasan dan ketakutan sehingga memungkinkan subyek wawancara menyatakan sikap dan perasaan dengan bebas, tanpa mekanisme pertahanan diri yang kadang-kadang menghambat pernyataannya.
- 2) Meredakan ketegangan yang terdapat dalam subyek wawancara. Oleh karena subyek wawancara pada umumnya membawa berbagai ketegangan emosi ke dalam pertemuan dalam wawancara itu, maka kedua belah pihak harus berusaha meredakan ketegangan di dalam dirinya.
 - 3) Menyediakan informasi yang dibutuhkan. Dalam wawancara kedua belah pihak akan mendapat kesempatan untuk memperoleh informasi yang dibutuhkannya.
 - 4) Mendorong kearah pemahaman diri pada pihak subyek wawancara. Hampir semua subyek wawancara menginginkan pemahaman diri yang lebih baik, dan pada dasarnya memiliki kesanggupan dan bakat yang seringkali tidak dapat berkembang dengan sempurna. Dengan wawancara subyek wawancara akan lebih memahami dirinya.
 - 5) Mendorong ke arah penyusunan kegiatan yang konstruktif pada subyek wawancara

Keuntungan dengan wawancara

Hubungan secara personal, akan memperoleh data secara langsung, cepat dan ekonomis.

- 1) Problem akan langsung mengenai sasaran, penegasan maksud pertanyaan dapat langsung diutarakan.
- 2) Metode ini bersifat fleksibel, mudah menyesuaikan dengan keadaan untuk diarahkan pada relevansi informasi.

Kelemahan dengan wawancara

Jangkauan responden relatif kecil dan memakan waktu lebih lama dari pada angket.

- 1) Biayanya lebih mahal
- 2) Dibutuhkan lebih banyak tenaga pewawancara.

3. Kuesioner

Kuesioner adalah sejumlah pertanyaan tertulis yang digunakan untuk memperoleh informasi dari responden dalam arti laporan tentang pribadinya, atau hal-hal yang diketahui". Kuesioner dipakai untuk menyebutkan metode maupun instrumen. Jadi dalam menggunakan metode angket atau kuesioner instrumen yang dipakai adalah angket atau kuesioner.

Kelebihan kuesioner sebagai berikut:

- 1) Tidak memerlukan hadirnya peneliti.
- 2) Dapat dibagikan secara serentak kepada responden.
- 3) Dapat dijawab oleh responden menurut kecepatannya masing-masing menurut waktu senggang responden.

- 4) Dapat dibuat anonim sehingga responden bebas, jujur dan tidak malu-malu menjawab.
- 5) Dapat dibuat berstandar sehingga semua responden dapat diberi pertanyaan yang benar-benar sama.

Kelemahan kuesioner adalah sebagai berikut:

- 1) Responden sering tidak teliti dalam menjawab sehingga ada pertanyaan yang terlewat tidak terjawab, padahal sukar diulangi diberikan kembali padanya.
- 2) Seringkali sukar dicari validitasnya
- 3) Walaupun dibuat anonim, kadang-kadang responden sengaja memberikan jawaban yang tidak betul atau tidak jujur
- 4) Angket yang dikirim lewat pos pengembaliannya sangat rendah, hanya sekitar 20%. Seringkali tidak dikembalikan terutama jika dikirim lewat pos menurut penelitian
- 5) Waktu pengembaliannya tidak sama-sama, bahkan kadang-kadang ada yang terlalu lama sehingga terlambat

2

Penyajian Data

Dalam bab ini, dibahas bagaimana cara memvisualisasikan bentuk-bentuk penting dari sebuah *data set*, baik untuk data kualitatif maupun kuantitatif. Jika data yang sudah terkumpul, baik melalui pengamatan (observasi), wawancara dan sebagainya, maka data tersebut perlu disusun secara sistematis supaya mudah dimengerti dan dipresentasikan dengan baik.

2.1. PENYAJIAN DATA KUALITATIF

Untuk dapat menganalisis data kualitatif, terlebih dahulu kita perlu menyusun secara sistematis. Data ini dapat disusun menurut beberapa cara. Setelah data terkumpul, biasanya yang ingin diketahui:

- Nilai variabel apa yang telah diukur
- Seberapa sering masing-masing nilai terjadi.

Untuk itu, perlu penyajian data sehingga mudah dibaca atau dimengerti. Penyajian data antara lain dalam bentuk tabel, dan grafik. Tipe grafik tergantung pada tipe data. Jika datanya kualitatif, tabel statistik berupa daftar kategori (kualitas) dan ukuran seberapa sering masing-masing kategori terjadi.

Ukuran yang biasa digunakan:

- **Frekuensi**, banyaknya pengukuran yang terjadi (kejadian) untuk masing-masing kategori.
- **Frekuensi relatif**, proporsi frekuensi masing-masing kategori.
- **Persentase** frekuensi masing-masing kategori.

Selain tabel statistik, dapat juga digunakan *pie chart* atau *bar chart* untuk menunjukkan distribusi data.

Contoh 2.1.

Dalam suatu survei pendidikan umum, 400 sekolah diteliti kualitas pendidikannya. Hasilnya untuk masing-masing sekolah dikategorikan

masuk kualitas A, B, C, atau D. Hasil survey disajikan dalam bentuk tabel statistik, *pie chart*, dan *bar chart* :

Tabel 2.1

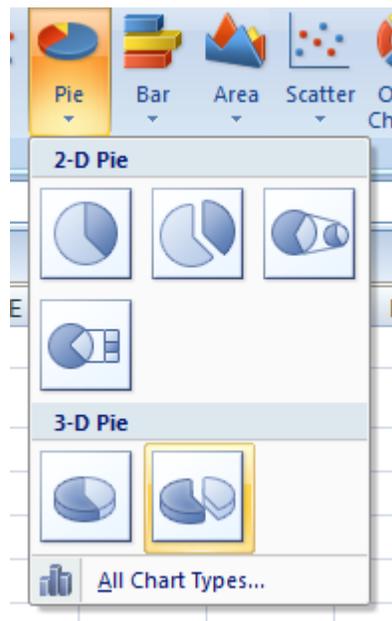
Kategori	Frekuensi	Frekuensi relatif	persen
A	35	$35/400 = 0,09$	9%
B	260	$260/400 = 0,65$	65%
C	93	$93/400 = 0,23$	23%
D	12	$12/400 = 0,03$	3%
Total	400	1	100%

Untuk memperoleh gambar dalam bentuk *pie chart*, dan *bar chart*, kita dapat menggunakan paket Microsoft Office Excel sebagai berikut ini.

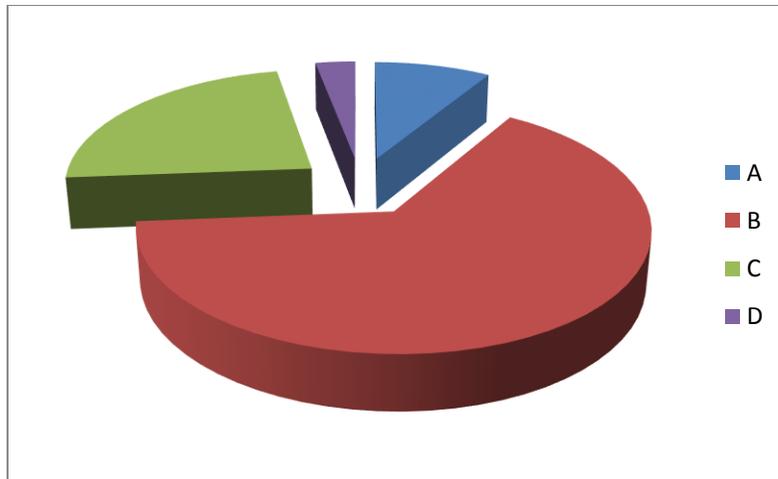
- Tuliskan data dari tabel 2.1 ke dalam shift seperti pada gambar berikut

Kategori	Frekuensi
A	35
B	260
C	93
D	12

- Kita blok cell kategoridan frekuensi, dari A sampai D dan 35 sampai 12.
- Klik insert, pilih pie seperti pada gambar berikut.

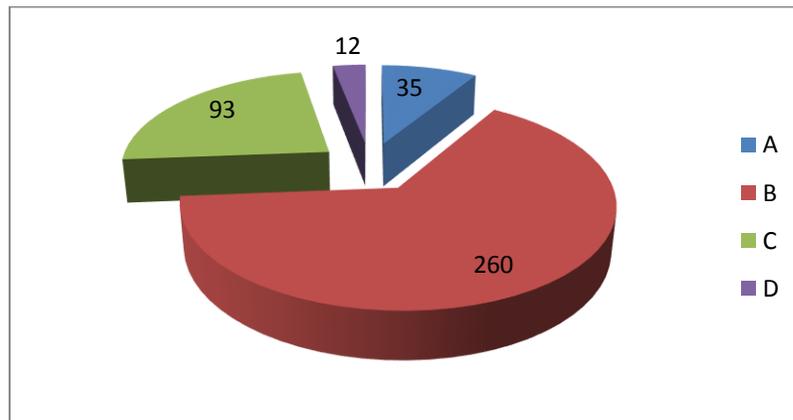


- Pilih sesuai keinginan kita, apakah yang dua dimensi atau tiga dimensi. Misalkan kita memilih 3-D pie. Klik gambar yang kita inginkan. Akan muncul gambar seperti berikut ini.

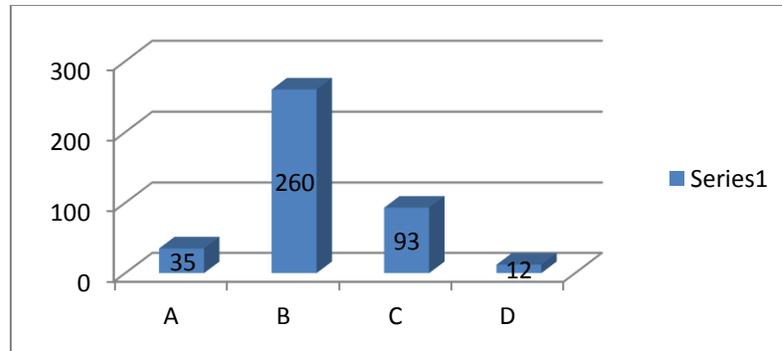


Penyajian data dengan diagram pie ini memiliki keunggulan/cocok dipergunakan jika seseorang ingin mengetahui proporsi dari masing-masing data yang disajikan dengan mencermati perbandingan ukuran gambar.

- Untuk memberikan label nilai masing-masing kategori, klik kanan gambar, kemudian pilih add data labels. Akan dihasilkan gambar yang sudah ada labelnya seperti gambar di bawah ini.



Untuk membuat *bar chart*, lakukan langkah yang sama. Pilihlah bar menu dan sesuaikan dengan keinginan kita. Hasil yang diperoleh seperti gambar di bawah ini.



2.2. PENYAJIAN DATA KUANTITATIF

Kadang-kadang data kuantitatif yang diperoleh dikelompokkan ke dalam segmen atau kategori yang berbeda dalam sebuah populasi. Untuk contoh data tentang rata-rata penghasilan penduduk menurut kelompok umur yang berbeda, jenis kelamin, tempat tinggal, dsb. Dalam hal ini, penghasilan penduduk merupakan data kuantitatif, tetapi dikelompokkan berdasar umur atau jenis kelamin atau tempat tinggal. Dalam kasus ini, dapat digunakan *pie chart* atau *bar chart* untuk mendiskripsikan data. Dengan menggunakan *pie chart* dapat dilihat bagaimana total penghasilan penduduk terdistribusi dalam beberapa kelompok umur. Sedang dari *bar chart* dapat dilihat berapa total penghasilan perkelompok umur, karena tinggi bar menunjukkan jumlah penghasilan per kelompok umur.

Selain *pie chart* dan *bar chart* masih banyak lagi cara yang bisa digunakan untuk menyajikan data kuantitatif. Diantaranya *line chart*.

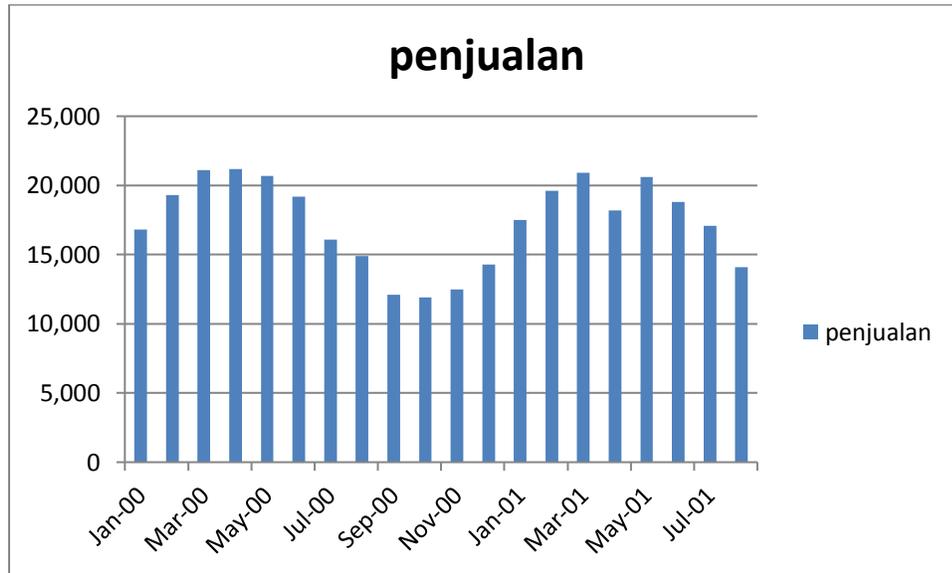
Contoh 2.2 :

Data hasil pengamatan mengenai nilai penjualan suatu perusahaan yang memproduksi suatu barang.

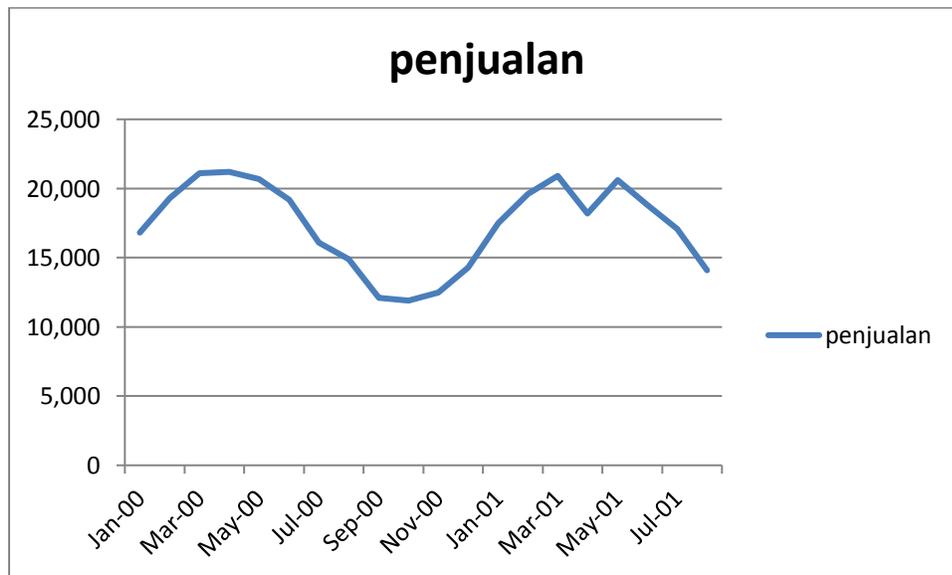
Bulan	penjualan
Januari 2000	16.800
Februari 2000	19.300
Maret 2000	21.100
April 2000	21.200
Mei 2000	20.700
Juni 2000	19.200
Juli 2000	16.100
Agustus 2000	14.900
September 2000	12.100
Oktober 2000	11.900

Bulan	penjualan
November 2000	12.500
Desember 2000	14.300
Januari 2001	17.500
Februari 2001	19.600
Maret 2001	20.900
April 2001	18.200
Mei 2001	20.600
Juni 2001	18.800
Juli 2001	17.100
Agustus 2001	14.100

Untuk melihat perubahan penjualan perbulan secara cepat, maka dibuat grafik sebagai berikut:



Kadang-kadang kita juga membutuhkan diagram polygon untuk melihat pola penjualannya.



2.3. PENYUSUNAN TABEL DISTRIBUSI FREKUENSI

A. Langkah-langkah Menyusun tabel Distribusi Frekuensi

Jika data yang ada adalah kuantitatif dan banyak sekali jumlahnya, maka untuk memudahkan dalam analisa data perlu dibuat distribusi

frekuensi atau tabel frekuensi. Berikut ini akan disajikan bagaimana cara menyusun distribusi kuantitatif, yaitu :

1. Tentukan rentang data, yaitu selisih data terbesar dengan data terkecil.
2. Tentukan banyaknya kelas interval. Untuk menentukan banyaknya kelas interval, hal yang sering digunakan adalah dengan menggunakan aturan Sturges, yaitu dengan rumus :

$$b = 1 + 3,3 \log(n)$$

dimana n adalah banyaknya data. Hasil perhitungan dari rumus Sturges menghasilkan bilangan pecahan, sehingga perlu dilakukan pembulatan. Sebagai contoh, jika banyaknya data, $n = 60$.

$$b = 1 + 3,3 \log(60)$$

$$1 + 3,3 \times 1,778 = 6,92$$

Kita dapat membulatkan banyaknya kelas interval sebesar 6 atau 7.

3. Tentukan lebar atau panjang kelas interval. Kita gunakan rumus :

$$p = \frac{r}{b}$$

dengan r adalah rentang dan b adalah banyaknya kelas interval. Nilai p hasil perhitungan juga berupa bilangan pecahan. Pembulatan yang dilakukan adalah menyesuaikan dengan decimal dari data. Jika data berupa bilangan bulat, nilai p juga harus bulat. Demikian seterusnya, jika data berupa satu angka di belakang koma, maka p juga harus satu angka di belakang koma.

4. Tentukan ujung bawah kelas interval yang pertama. Ujung bawah kelas interval yang pertama dapat berupa nilai data terkecil atau suatu nilai yang kurang dari data terkecil tetapi jaraknya tidak boleh terlalu jauh dari data terkecil.
5. Masukkan masing-masing frekuensi berdasarkan kelas interval yang sesuai.

Contoh 2.3.

Di bawah ini adalah data nilai ulangan harian mata pelajaran matematika 52 siswa kelas 6 dari suatu sekolah dasar.

75	74	72	88	89	81	73	67	70
65	89	69	67	84	67	89	78	75
52	84	60	71	90	65	80	75	80
89	66	67	70	93	56	70	53	90
97	55	78	75	90	54	68	83	
78	54	92	68	95	76	75	88	

Langkah-langkah yang kita lakukan adalah sebagai berikut:

1. Tentukan rentang, yaitu selisih data terbesar dengan data terkecil. Data terbesar adalah 97 dan terkecil adalah 52

$$r = 97 - 52 = 45$$

2. Tentukan banyaknya kelas interval

$$b = 1 + 3,3 \log(52) = 6,66$$

Nilai b hasil perhitungan adalah 6,66 sehingga banyaknya kelas interval dapat kita ambil 6 atau 7. Dalam contoh ini, misalkan kita ambil 7.

3. Tentukan lebar kelas.

$$p = \frac{45}{7} = 6,43$$

Karena data dalam bentuk bilangan bulat, maka lebar kelas juga harus berupa bilangan bulat. Kita dapat mengambil nilai 6 atau 7. Untuk kenyamanan kita gunakan pembulatan ke atas, yaitu 7.

4. Tentukan ujung bawah kelas interval yang pertama. Nilai ujung bawah kelas pertama boleh menggunakan data terkecil, 52 atau bilangan yang kurang dari 52, misalnya 50 atau 51. Kita pilih 52 sebagai ujung bawah kelas pertama.
5. Masukkan frekuensi sesuai dengan kelas masing-masing. Tabel yang diperoleh adalah seperti di bawah ini.

Tabel 2.1. Tabel Distribusi Frekuensi Nilai Ujian Siswa

Nilai		Frekuensi
52	58	6
59	65	3
66	72	13
73	79	11
80	86	6
87	93	11
94	100	2

Jika kita akan menyusun data diatas dengan menggunakan Ms Office Excel, ikuti langkah-langkah berikut ini.

- 1) Tuliskan data asli ke dalam cell A1 sampai A52.
- 2) Tuliskan batas bawah pada cell C1:C7 dan batas atas dalam D1:D7.

- 3) Tuliskan di cell E1 rumus sebagai berikut :
=frequency(A1:A52;D1:D7)
- 4) Blok cell E2:E7
- 5) Tekan F2
- 6) Tekan secara bersama tombol **Ctrl**, **Shift**, dan **Enter**

Distribusi frekuensi atau tabel frekuensi dapat pula dilengkapi dengan kolom-kolom frekuensi kumulatif (frekuensi kumulatif kurang dari) frekuensi relatif atau frekuensi persentase. Frekuensi kumulatif lebih dari batas atas adalah banyaknya data yang tercakup dalam kisaran mulai dari batas atas sampai batas bawah terkecil.

Kadang-kadang kita tidak hanya ingin mengetahui berapa banyak siswa yang memiliki nilai tertentu, melainkan kita juga ingin mengetahui berapa banyak siswa yang memiliki nilai kurang dari angka tertentu, ataupun lebih dari angka tertentu.

Berikut ini contoh tabel distribusi kumulatif, distribusi kumulatif kurang dari dan distribusi kumulatif lebih dari yang diambil dari Contoh 2.3.

Tabel 2.2. Tabel Distribusi Frekuensi Kumulatif

Nilai		Frekuensi	Frekuensi Kumulatif
52	58	6	6
59	65	3	9
66	72	13	22
73	79	11	33
80	86	6	39
87	93	11	50
94	100	2	52

Tabel 2.3. Tabel Distribusi Frekuensi Kumulatif “kurang dari”

Nilai	Banyaknya siswa
kurang dari 52	0
kurang dari 59	6
kurang dari 66	9
kurang dari 73	22
kurang dari 80	33
kurang dari 87	39
kurang dari 94	50
kurang dari 101	52

Tabel 2.4. Tabel Distribusi Frekuensi Kumulatif “lebih dari”

Nilai	Banyaknya siswa
lebih dari 52	52
lebih dari 59	46
lebih dari 66	43
lebih dari 73	30
lebih dari 80	19
lebih dari 87	13
lebih dari 94	2
lebih dari 101	0

Penyusunan distribusi frekuensi yang telah kita bahas hanya merupakan pedoman umum, karena setiap peneliti dapat menentukan ketentuan-ketentuan lain sesuai dengan masalah tujuan dan pengalamannya.

B. Menggambar Grafik Distribusi Frekuensi

Informasi yang terkandung di dalam sebaran frekuensi yang berbentuk tabel akan lebih mudah dipahami apabila disajikan dalam bentuk gambar grafik distribusi frekuensi.

Dalam menggambar grafik suatu distribusi frekuensi interval kelas distribusi kita letakkan pada sumbu X sedangkan frekuensi kita letakkan pada sumbu Y. Selanjutnya dengan alas tiap-tiap interval kita buat persegi panjang-persegi panjang dengan tinggi sama dengan frekuensi masing-masing interval itu. Maka diperoleh grafik yang dinamakan **histogram** sebagai contoh kita gambar grafik histogram distribusi tersebut diatas.

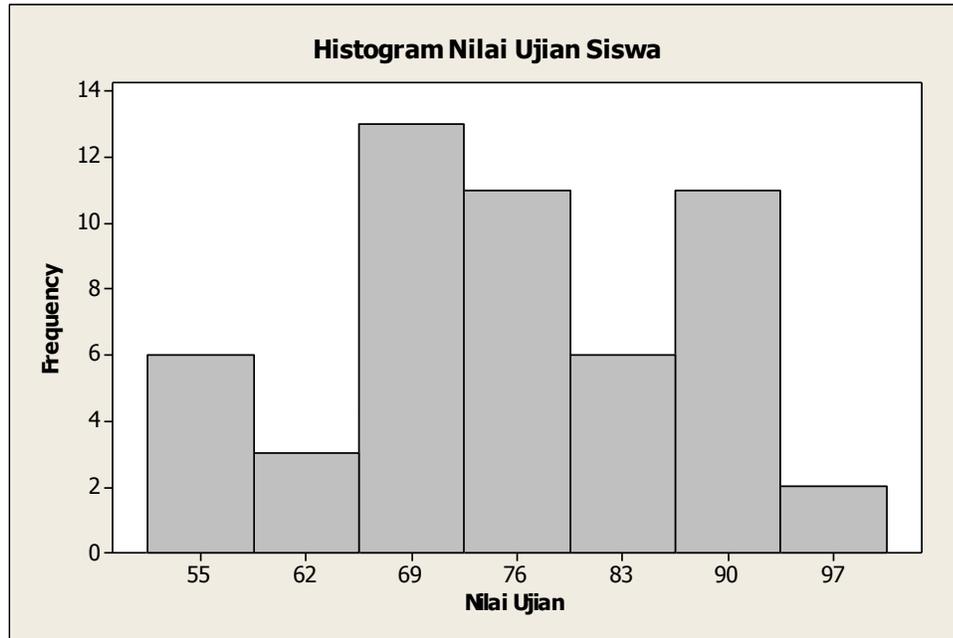
Contoh 2.3.

Pandang data pada Contoh 2.2. Dari menggunakan bantuan paket Minitab kita dapat membuat tabel distribusi frekuensi dan menggambar grafik histogramnya dengan langkah sebagai berikut :

- 1) Masukkan data pada kolom C1
- 2) Pilih Graph >> Histogram...
- 3) Pilih Simple.... Klik OK
- 4) Masukkan C1 ke dalam graph variable, klik OK
- 5) Setelah muncul gambar, klik kanan pada gambar diagram tersebut.
- 6) Pilih Binning
- 7) Pilih pada interval definition >> number of intervals : 7, karena kita hitung banyaknya kelas 7.
- 8) Masukkan nilai tengah masing-masing kelas ke dalam midpoint/cutpoint position
- 9) Klik OK

- 10) Untuk mengedit tampilan judul, klik kanan histogram of C1
- 11) Pilih edit title
- 12) Gantilah dengan histogram nilai ujian siswa

Gambar yang diperoleh adalah seperti di bawah ini.



Distribusi frekuensi dapat juga digambar dengan grafik frekuensi polygon. Cara menggambar grafik ini adalah dengan meletakkan interval distribusi frekuensi pada sumbu X dan frekuensinya pada sumbu Y seperti pada menggambar histogram diatas tadi. Selanjutnya kita tentukan titik tengah tiap intervalnya dan dari titik tengah-titik tengah itu kita ukurkan ke atas sejauh frekuensi masing-masing maka diperoleh grafik batang. Apabila ujung-ujung batang yang saling berdekatan dihubungkan dengan garis lurus maka diperoleh grafik frekuensi polygon.

Penyajian data dalam bentuk gambar grafik di maksudkan untuk menunjukkan cirri-ciri penting dari suatu distribusi frekuensi secara keseluruhan walaupun cara penyajiannya itu memberikan informasi yang lebih sedikit dibandingkan table. Namun bentuk grafik biasanya lebih menarik bagi pembaca. Penyajian-penyajian grafik yang lain seperti pie chart, pictogram dan sebagainya bisa anda pelajari sendiri.

C. Penyusunan Distribusi Frekuensi Kualitatif

Cara penyusunan distribusi kuantitatif pada dasarnya sama dengan penyusunan frekuensi distribusi kuantitatif. Pertama-tama harus kita tentukan banyaknya kelas (kategori) yang akan kita gunakan. Banyaknya kelas yang kita gunakan sangat tergantung dari keadaan data dan juga

keinginan/maksud penelitian. Dalam menyusun kelas perlu diperhatikan jangan sampai ada data yang tidak termuat, juga harus dihindari kelas-kelas yang tumpang tindih (overlap).

Contoh 2.4

Hasil ujian akhir Statistika Pendidikan di Jurusan Tarbiyah pada Tahun Akademik 2012/2013 seperti pada tabel berikut ini.

No	NIM	NAMA	PARTPS.	TUGAS	SISIPAN	UTAMA	ANGKA	HURUF
1	072331179	YOUSEPS MAHENDRA	100	100	40	52	56,80	C
2	082333006	ASRIYANI NH	80	100	40	50	55,00	D
3	082333008	DANI TAMAM ADI	100	100	68,75	50	67,50	B
4	082333036	KHOTIBUL UMAM	100	100	68,75	77	78,30	B
5	092331101	THOFIQR ROHMAN	90	100	88,75	29	66,60	B
6	092331106	MUHAMMAD ISA NURUL	100	100	70	90	84,00	A
7	092331110	NUR AFIAH	90	100	72,5	60	72,50	B
8	092331123	FAHRI AS'ARI	80	100	55		41,00	E
9	092331136	DIAN PURWANTI	80	100	55		41,00	E
10	092331142	BAHRUN ULUM	90	100	68,75	50	67,00	B
11	102331097	FERY HELMI KURNIAWAN	100	100	75	67	76,80	B
12	102331101	IDATUL FARIHAH	80	80	75	90	82,00	A
13	102331104	AFIDATUN KHASANAH	100	100	81,25	70	80,50	A
14	102331109	LATHIFATUL MARHAMAH	80	100	75	87	83,80	A
15	102331110	LUTFIANA LAELA	100	100	75	75	80,00	A
16	102331113	NOVI SUBAH AHYANI	100	100	81,25	100	92,50	A
17	102331114	MUTTAMIMATUL KHIKMAH	100	100	78,75	42	68,30	B
18	102331115	ANIS ZULIA A'LIMATUN N	100	100	81,25	95	90,50	A
19	102331116	SRI RAHMAYANTI	100	100	87,5	77	85,80	A
20	102331118	ALFIYAN NAUFARY	90	100	72,5	72	77,30	B
21	102331120	SATRIA WILOPO	90	100	68,75	60	71,00	B
22	102331122	NURUL HIDAYANTI	100	100	63,75	92	82,30	A
23	102331127	OOM KOMALASARI	90	100	87,5	77	85,30	A
24	102331129	MUCHAMAD SAMSU RIJAL	90	100	81,25	72	80,80	A
25	102331130	ZIAN FEBRIANA	100	100	66,25	95	84,50	A
26	102331136	SITI KHOLISOH	100	100	75	92	86,80	A
27	102331137	HENDAR RAMADANI	100	100	88,75	92	92,30	A
28	102331141	SEHUN MUHTAROM	100	100	82,5	57	75,80	B
29	102331142	FATHURROCHMAN	100	100	75	92	86,80	A
30	102331143	AHMAD NAZAL HADI F	100	100	88,75	92	92,30	A
31	102331145	PUJI LESTARI	100	100	75	100	90,00	A

32	102331147	WASMAWATI	100	100	93,75	83	90,70	A
33	102331148	MOHAMMAD FAHMI	100	100	85	88	89,20	A
34	102331149	SOLIHUN	100	100	75	77	80,80	A
35	102331151	DIANA PUTRI SARI	100	100	91,25	100	96,50	A
36	102331152	SITI RAHMAH	100	100	100	96	98,40	A
37	102331153	SUROTUL YASIN	100	100	87,5	85	89,00	A
38	102331154	TAUHID SUROHMAT	100	100	75	80	82,00	A
40	102331157	KHASAN	100	100	91,25	67	83,30	A
41	102331158	MASRURI	100	100	80	37	66,80	B
42	102331160	BAETY FITRIANA	100	100	71,25	98	87,70	A
43	102331162	TAUFIK QUROHMAN	100	100	60	100	84,00	A
44	102331164	DWI SETIAWAN	100	100	88,75	82	88,30	A
45	102331165	HERI PURWANTO	100	100	68,75	43	64,70	C
46	102331166	AHMAD FAJRI AMIRUDIN	100	100	71,25	67	75,30	B
47	102331168	ANDI AROHMAN	100	100	83,75	87	88,30	A
48	102331169	TITIS EMA NURMAYA	100	100	93,75	87	92,30	A
49	102331172	LUPI LINDA ISNANI	90	100	93,75	92	93,80	A
50	102331174	DURROTUL KHAKIMAH	90	100	50	95	77,50	B
51	102331177	WASTIYANI	100	100	100	100	100,00	A
52	102338005	MUKH SIHABUDIN	100	100	70	85	82,00	A
53	102338006	AFRILIA PURWANINGSIH	90	100	87,5	95	92,50	A
54	102338010	SUBEKTI RIDHO	90	100	81,25	60	76,00	B
55	102338020	RINI HIDAYATI	90	100	87,5	95	92,50	A
56	102338032	AJI SANTOSO	100	100	81,25	58	75,70	B
57	102338048	MAKMUN MUROD E Z	100	100	76,25	80	82,50	A

Jika kita hanya akan melihat distribusi perolehan nilai dari mahasiswa tadi, kita lebih suka menggunakan tabel distribusi frekuensi seperti berikut ini.

Tabel distribusi frekuensi data nilai ujian Statistika pendidikan di Jurusan Tarbiyah

No	Nilai	Frekuensi
1	A	20
2	B	15
3	C	10
4	D	3
5	E	2

Untuk memperoleh tabel distribusi frekuensi di atas, kita dapat menggunakan MS Excel dengan cara berikut ini.

- 1) Misalkan di cell B66 akan memuat "No.", C66 untuk "Nilai" dan D66 untuk "Frekuensi"
- 2) Untuk cell B67 sampai B71 tuliskan angka 1 sampai 5.
- 3) Untuk cell C67 sampai C71 tuliskan nilai huruf A, B, C, D, dan E.
- 4) Untuk cell D67, tuliskan =COUNTIF(I9:I64;"A"), karena data nilai kita dimulai dari baris ke 9 sampai baris ke 64.
- 5) Berturut-turut, untuk kategori nilai B, C, D, dan E, dituliskan pada cell D68 sampai D71 dengan menuliskan =COUNTIF(I9:I64;"B"), =COUNTIF(I9:I64;"C"), =COUNTIF(I9:I64;"D") dan =COUNTIF(I9:I64;"E") .
- 6) Jika kita akan menjumlahkan semua perolehan nilai tadi, misalkan di cell D72, tuliskan : =SUM(D67..D71).

3

Ukuran Pemusatan

3.1. UKURAN PEMUSATAN

Ukuran pemusatan merupakan suatu ukuran yang menetapkan letak titik pemusatan dimana terdapat kecenderungan bagi setiap variabel untuk mengarah kepadanya. Suatu ukuran pemusatan merupakan suatu bilangan tunggal yang dipergunakan untuk mewakili suatu kelompok data. Karena kelompok-kelompok data yang berbeda-beda memiliki sifat-sifat numerik yang berlainan, maka suatu ukuran pemusatan dapat lebih baik dalam menggambarkan sekelompok data tertentu dari yang lain.

Berikut ini akan diuraikan tentang tiga buah ukuran dasar dari ukuran pemusatan, yaitu rata-rata hitung (mean, rata-rata), median, dan modus.

A. Rata-rata

Rata-rata hitung (atau sering disebut dengan rata-rata atau mean) merupakan suatu bilangan tunggal yang digunakan untuk mewakili nilai pusat dari sebuah distribusi. Dalam pemakaian sehari-hari orang awam lebih mempergunakan istilah rata-rata dari istilah rata-rata hitung. Bagi sekelompok data, rata-rata adalah nilai rata-rata dari data itu. Secara teknis dapat dikatakan bahwa rata-rata dari sekelompok variabel adalah jumlah nilai pengamatan dibagi dengan banyaknya pengamatan.

Rata-rata hitung adalah jumlah semua nilai data yang diamati dibagi banyaknya data yang diamati. Secara rumus dituliskan sebagai:

$$\bar{X} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

Dalam notasi sigma, rumus di atas biasanya dituliskan dalam bentuk:

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

Contoh 3.1.

Tentukan nilai rata-rata berat badan dari lima mahasiswa yang berturut-turut beratnya adalah sebagai berikut:
45, 52, 57, 65 dan 49.

Dari soal diketahui banyaknya data, $n = 5$. Rumus yang digunakan adalah:

$$\bar{X} = \frac{\sum_{i=1}^5 X_i}{n}$$
$$\bar{X} = \frac{45 + 52 + 57 + 65 + 49}{5} = \frac{268}{5} = 53,6$$

Rumus rata-rata di atas digunakan untuk nilai rata-rata data yang tidak terbobot atau data tunggal. Tetapi jika data diketahui seperti pada Tabel 3.1, rumus untuk rata-ratanya akan sedikit berbeda seperti di bawah ini.

Tabel 3.1. Data terbobot

X_i	Frekuensi (f_i)
X_1	n_1
X_2	n_2
X_3	n_3
.....
X_k	n_k
Total	$n = n_1 + n_2 + n_3 + \dots + n_k$

$$\bar{X} = \frac{\sum_{i=1}^k X_i \cdot f_i}{n}$$

Contoh 3.2.

Diketahui tabel perolehan nilai (dalam bilangan bulat) ulangan harian matematika di suatu kelas sebagai berikut.

Nilai	banyaknya siswa
5	4
6	14
7	24
8	18
9	7
10	5
Total	72

Tentukan rata-rata nilai harian dari kelas tersebut.

Untuk mengerjakan contoh di atas, sebaiknya kita siapkan dahulu tabel seperti berikut ini:

X_i	f_i	$X_i \cdot f_i$
5	4	20
6	14	84
7	24	168
8	18	144
9	7	63
10	5	50
Total	72	529

Dari tabel terlihat bahwa $\sum_{i=1}^6 f_i = 72$ dan $\sum_{i=1}^6 X_i \cdot f_i = 529$. Kita dapat menghitung rata-ratanya

$$\bar{X} = \frac{529}{72} = 7,347.$$

Rumus-rumus di atas umumnya digunakan untuk data yang tidak terlalu banyak jumlahnya. Jika n cukup besar, kita akan banyak menemui kesulitan perhitungan. Oleh karena itu ada cara lain yang disebut dengan cara rata-rata data berkelompok.

Misalkan data sudah tersusun dalam suatu tabel distribusi frekuensi.

Nilai	Tanda Kelas	Frekuensi
$a_1 - b_1$	$X_1 = \frac{(a_1 + b_1)}{2}$	f_1
$a_2 - b_2$	$X_2 = \frac{(a_2 + b_2)}{2}$	f_2
$a_3 - b_3$	$X_3 = \frac{(a_3 + b_3)}{2}$	f_3
.....
$a_k - b_k$	$X_k = \frac{(a_k + b_k)}{2}$	f_k
Total		$n = \sum_{i=1}^k f_i$

Untuk menghitung rata-rata, kita dapat menggunakan rumus data berkelompok dengan cara mencari jumlah perkalian tanda kelas dengan frekuensi dibagi dengan total jumlah frekuensinya. Cara ini kadang-kadang sangat menyulitkan dalam perhitungan, sehingga kita menggunakan bentuk rumus yang lain, dengan menggunakan nilai rata-rata dugaan.

$$\bar{X} = AM + p \frac{\sum_{i=1}^k f_i \cdot d_i}{\sum_{i=1}^k f_i}$$

dengan AM adalah suatu nilai tanda kelas yang dipilih, p adalah panjang kelas interval, dan d_i adalah koding untuk kelas ke- i .

Contoh 3.3.

Hitunglah rata-rata dari data yang tersaji dalam tabel distribusi frekuensi berikut ini.

Nilai		Frekuensi
52	58	6
59	65	3
66	72	13
73	79	11
80	86	6
87	93	11
94	100	2

Untuk menjawab soal diatas, lakukan langkah-langkah berikut ini.

- 1) Susunlah tabel di atas dengan menambahkan kolom tanda kelas (nilai tengah interval) seperti berikut ini.

Nilai		Nilai Tengah	Frekuensi
52	58	55	6
59	65	62	3
66	72	69	13
73	79	76	11
80	86	83	6
87	93	90	11
94	100	97	2

- 2) Tentukan nilai AM dengan mengambil sembarang nilai tengah. Misalkan kita ambil 76. Baris yang sesuai untuk koding kelas 73- 79 adalah 0. Kelas di atasnya berturut-turut mengecil dan kelas di bawahnya membesar, seperti pada tabel berikut.

Nilai		Nilai Tengah	Frekuensi	Koding (di)
52	58	55	6	-3
59	65	62	3	-2
66	72	69	13	-1
73	79	76	11	0
80	86	83	6	1
87	93	90	11	2
94	100	97	2	3

- 3) Kalikan frekuensi dengan koding yang sesuai, seperti tabel berikut dan jumlahkan hasilnya.

Nilai		Nilai Tengah	Frekuensi	Koding (di)	fi*di
52	58	55	6	-3	-18
59	65	62	3	-2	-6
66	72	69	13	-1	-13
73	79	76	11	0	0
80	86	83	6	1	6
87	93	90	11	2	22
94	100	97	2	3	6
Total			52		-3

- 4) Masukkan ke dalam rumus, dengan $p = 7$, $AM = 76$, $\sum f_i = 52$ dan $\sum f_i d_i = -3$,

$$\bar{X} = AM + p \frac{\sum_{i=1}^7 f_i \cdot d_i}{\sum_{i=1}^7 f_i}$$

$$\bar{X} = 76 + 7 \frac{-3}{52} = 75,6$$

B. Median

Median merupakan nilai yang membagi serangkaian nilai variabel (data) sedemikian rupa sehingga setengah dari rangkaian itu mempunyai nilai yang lebih kecil dari atau sama dengan nilai media. Sedangkan

setengahnya lagi memiliki nilai yang sama dengan atau lebih besar dari nilai median. Median dapat juga disebut rata-rata karena yang menjadi dasar adalah letak variabel bukan nilainya.

Cara menentukan median untuk data yang belum dikelompokkan:

$$Me = \begin{cases} \frac{X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}}{2}, & \text{jika } n \text{ genap} \\ X_{(\frac{n+1}{2})} & \text{jika } n \text{ ganjil} \end{cases}$$

Contoh 3.4.

Hitung median dari data berikut ini : 7, 5, 6, 8, 4, 9, 10, 12, 9, 10, 13

Pertama yang kita lakukan adalah mengurutkan data dari terkecil sampai yang terbesar.

4, 5, 6, 7, 8, 9, 9, 10, 10, 12, 13

Dari data tersebut tadi, karena banyaknya data, $n = 11$, maka Median data tersebut adalah data tersebut yang ke $\frac{11+1}{2} = \frac{12}{2} = 6$.

Sehingga, $Me = X_{(6)} = 9$

Jika data sudah tersaji dalam tabel distribusi frekuensi, median dapat dihitung dengan langkah-langkah sebagai berikut.

- 1) Carilah setengah dari total frekuensi ($\frac{n}{2}$);
- 2) Buatlah kolom frekuensi kumulatif, dengan cara jumlahkan frekuensi mulai dari interval kelas pertama dan seterusnya.
- 3) Tentukan F, yaitu frekuensi kumulatif sebelum kelas yang memuat data ke- $\frac{n}{2}$.
- 4) Kelas yang memuat data ke- $\frac{n}{2}$ disebut kelas median. Tentukan batas bawah kelas mediannya (BB_{md})
- 5) Tentukan frekuensi kelas mediannya, f.
- 6) Tentukan lebar kelas (panjang kelas interval), p
- 7) Hitung dengan rumus

$$Me = BB_{Md} + p \frac{\left(\frac{n}{2} - F\right)}{f_{Md}}$$

Contoh 3.5.

Tentukan median dari data yang tersaji dalam tabel distribusi frekuensi berikut ini.

Nilai		Frekuensi
52	58	6
59	65	3
66	72	13
73	79	11
80	86	6
87	93	11
94	100	2

Langkah-langkah penyelesaian contoh di atas sebagai berikut:

1) Tambahkan kolom frekuensi kumulatif.

Nilai		Frekuensi	Frekuensi Kumulatif
52	58	6	6
59	65	3	9
66	72	13	22
73	79	11	33
80	86	6	39
87	93	11	50
94	100	2	52

2) Tentukan kelas median.

Diketahui banyaknya data, $n = 52$, maka “calon median” adalah data

terurut yang ke- $\frac{n}{2} = \frac{52}{2} = 26$

Data ke- 26 terletak pada kelas 73 – 79, sehingga kelas ini disebut sebagai kelas median. Batas bawah kelas median,

$$BB_{md} = 73 - 0,5 = 72,5$$

3) Nilai frekuensi kumulatif sebelum kelas median, F , yaitu $F = 22$ dan frekuensi kelas median, $f_{md} = 11$.

4) Panjang kelas interval, $p = 7$.

5) Masukkan semua nilai tadi ke dalam rumus :

$$Me = BB_{Md} + p \frac{\left(\frac{n}{2} - F\right)}{f_{Md}}$$

$$Me = 72,5 + 7 \frac{\left(\frac{52}{2} - 22\right)}{11}$$

$$Me = 75,045$$

C. Modus

Modus adalah nilai data yang memiliki frekuensi terbanyak. Untuk data tunggal (yang belum dikelompokkan) sangat mudah untuk menentukan modus, yakni dengan mencari nilai mana yang sering muncul atau frekuensinya paling tinggi.

Contoh 3.6.

Sumbangan PMI warga Purwanegara berturut-turut sebagai berikut: Rp. 7500 8000 9000 8000 3000 5000 8000. Besarnya sumbangan yang paling sering dilakukan warga adalah Rp. 8000, sehingga kita dapat mengatakan bahwa modus : Rp. 8000

Ukuran modus dapat terjadi data dengan beberapa modus (multi-modus) dan bahkan dapat pula terjadi data tanpa modus

Contoh 3.7.

- a. Berat 5 orang bayi : 3.6 3.5 2.9 3.1 3.0 (Tidak Ada Modus)
- b. Umur Mahasiswa : 19 18 19 18 23 21 19 21 18 20 22 17
Modus : 18 dan 19

Modus dari distribusi frekuensi adalah

$$Mo = BB_{\text{mod}} + p \frac{b_1}{b_1 + b_2}$$

dengan BB_{mod} adalah batas bawah kelas yang memuat modus, p panjang kelas interval, b_1 adalah selisih frekuensi kelas modus dengan frekuensi sebelum kelas modus, dan b_2 adalah selisih frekuensi kelas modus dengan kelas sesudah kelas modus.

Contoh 3.8.

Tentukan modus dari data di bawah ini.

Nilai		Frekuensi
52	58	6
59	65	3
66	72	13
73	79	11
80	86	6
87	93	11
94	100	2

- 1) Untuk menentukan kelas modus, tentukan kelas yang memiliki frekuensi terbanyak. Kelas modulusnya adalah 66 – 72, karena kelas ini memiliki frekuensi tertinggi.

2) Batas bawah kelas modusnya, $BB_{\text{mod}} = 66 - 0,5 = 65,5$

3) Panjang kelas interval, $p = 7$

4) Nilai b_1 adalah

$$b_1 = 13 - 3 = 10 \text{ dan } b_2 = 13 - 11 = 2$$

5) Hitung dengan rumus :

$$Mo = BB_{\text{mod}} + p \frac{b_1}{b_1 + b_2}$$

$$Mo = 65,5 + 7 \frac{10}{10 + 2} = 71,333$$

3.2. UKURAN LOKASI

Salah satu ukuran numerik yang penting adalah ukuran lokasi, yaitu suatu ukuran sepanjang garis horizontal yang letaknya ditengah distribusi data. Ukuran lokasi sekumpulan data adalah nilai yang representatif bagi keseluruhan nilai data atau dapat menggambarkan distribusi data itu, khususnya dalam hal letaknya (lokasinya). Nilai tersebut dihitung dari keseluruhan data bersangkutan sehingga cenderung terletak di urutan tengah setelah data diurutkan menurut besarnya. Beberapa ukuran lokasi yang akan dibicarakan adalah kuartil, desil, dan persentil.

A. Kuartil

K_1 atau kuartil 1 adalah suatu nilai yang membagi data sedemikian sehingga sekitar 25% dari data tersebut berada di bawahnya. Jadi kuartil 1 adalah suatu nilai yang berada pada posisi $\frac{1}{4}$ dari banyaknya data setelah data tersebut diurutkan.

Kuartil 2 (Median) adalah suatu nilai yang membagi data sedemikian sehingga kira-kira 50% dari data tersebut berada di bawahnya dan 50% berada di atasnya. Jadi Kuartil 2 (Median) adalah suatu nilai yang berada pada posisi $\frac{1}{2}$ dari banyaknya data setelah data tersebut diurutkan.

K_3 atau kuartil 3 adalah suatu nilai yang membagi data sedemikian sehingga sekitar 25% dari data tersebut berada di atasnya. Jadi Kuartil 3 berada pada posisi $\frac{3}{4}$ dari banyaknya data setelah data tersebut diurutkan.

Untuk menentukan nilai K_1 , K_2 , dan K_3 , kita harus tentukan dahulu lokasi dari ukuran tersebut.

Lokasi K_i , LK_i , untuk $i = 1,2,3$, dirumuskan sebagai berikut :

$LK_i = \frac{i(n+1)}{4}$, sehingga Kuartil ke- i dapat ditentukan dengan:

$$K_i = \text{data ke } \frac{i(n+1)}{4}$$

Contoh 3.9.

Tentukan K_1 , K_2 , dan K_3 dari data berikut ini : 11 6 17 9 12 4 4 14
20 10 15

Jawab:

Terlebih dahulu data asal diurutkan dari kecil ke besar :

4 4 6 9 10 11 12 14 15 17 20

Banyaknya data, $n = 11$.

Lokasi atau letak data ke-1,

$$LK_1 = \frac{1(n+1)}{4} = \frac{1(11+1)}{4} = \frac{12}{4} = 3$$

Jadi, $K_1 = X_{(3)} = 6$

Dengan cara yang sama,

$$LK_2 = \frac{2(n+1)}{4} = \frac{2(11+1)}{4} = \frac{24}{4} = 6$$

Jadi, $K_2 = X_{(6)} = 11$

$$LK_3 = \frac{3(n+1)}{4} = \frac{3(11+1)}{4} = \frac{36}{4} = 9$$

Jadi, $K_3 = X_{(9)} = 15$

Contoh 3.10.

Tentukan K_1 , K_2 dan K_3 dari data berikut ini :

2 5 3 8 9 6 5 4 11 10 12 10 12 15 9 7

Diketahui sebanyak 16 data, $n = 16$. Setelah diurutkan, data menjadi seperti urutan berikut ini.

2 3 4 5 5 6 7 8 9 9 10 10 11 12 12 15

$$LK_1 = \frac{1(n+1)}{4} = \frac{1(16+1)}{4} = \frac{17}{4} = 4,25$$

Jadi, $K_1 = X_{(4)} + 0,25(X_{(5)} - X_{(4)}) = 5 + 0,25(5 - 5) = 5$

$$LK_2 = \frac{2(16+1)}{4} = \frac{2(16+1)}{4} = \frac{34}{4} = 8,5$$

Jadi, $K_2 = X_{(8)} + 0,5(X_{(9)} - X_{(8)}) = 9 + 0,5(9 - 9) = 9$

$$LK_3 = \frac{3(n+1)}{4} = \frac{3(16+1)}{4} = \frac{51}{4} = 12,75$$

Jadi, $K_3 = X_{(12)} + 0,75(X_{(13)} - X_{(12)}) = 10 + 0,75(11 - 10) = 10,75$

Untuk data berkelompok, langkah-langkah yang dilakukan masih sama, seperti menentukan lokasi. Selanjutnya digunakan rumus seperti berikut ini.

$$K_i = BB_{K_i} + p \frac{\left(i \cdot \frac{n}{4} - F_{K_i} \right)}{f_{K_i}}$$

Contoh 3.11.

Tentukan K_1 , K_2 , dan K_3 data di bawah ini.

Nilai		Frekuensi	Frekuensi Kumulatif
52	58	6	6
59	65	3	9
66	72	13	22
73	79	11	33
80	86	8	41
87	93	9	50
94	100	2	52

1) Tentukan letak dari K_1 , K_2 , dan K_3 dengan rumus yang sama pada data yang tidak terkelompok.

$$LK_1 = \frac{1(n+1)}{4} = \frac{1(52+1)}{4} = \frac{53}{4} = 13,25$$

$$LK_2 = \frac{2(n+1)}{4} = \frac{2(52+1)}{4} = \frac{106}{4} = 26,5$$

$$LK_3 = \frac{3(n+1)}{4} = \frac{3(52+1)}{4} = \frac{159}{4} = 39,75$$

2) Tentukan batas bawah masing-masing kuartil.

Batas bawah K_1 adalah, $BB_{K_1} = 65,5$ karena data yang ke 13,25 terletak pada kelas interval 66 – 72. Dengan cara yang sama, berturut-turut $BB_{K_2} = 72,5$ dan $BB_{K_3} = 79,5$.

3) Hitung nilai K_1 , K_2 dan K_3 .

$$K_1 = BB_{K_1} + p \frac{\left(1 \cdot \frac{52}{4} - F_{K_1} \right)}{f_{K_1}} = 65,5 + 7 \frac{(13-9)}{13} = 67,65$$

$$K_2 = BB_{K_2} + p \frac{\left(2 \cdot \frac{52}{4} - F_{K_2} \right)}{f_{K_2}} = 72,5 + 7 \frac{(26-22)}{11} = 75,045$$

$$K_3 = BB_{K_3} + p \frac{\left(3 \cdot \frac{52}{4} - F_{K_3} \right)}{f_{K_3}} = 79,5 + 7 \frac{(39-33)}{8} = 84,75$$

B. Desil

Desil membagi data menjadi sepuluh bagian yang sama besar, sehingga letak dari D_i (desil ke- i) diringkas

$$LD_i = \frac{i(n+1)}{10}$$

Contoh 3.12.

Diketahui sampel dengan data : 42; 46; 47; 50; 54; 56; 60; 65; 72; 76; 82; dan 84. Desil ke tujuh (D_7) dari data diatas adalah

Penyelesaian :

$$\text{Letak desil ke tujuh } D_7 = \frac{7}{10}(12+1) = 9,1$$

Nilai desil ke tujuh adalah

$$D_7 = 72 + 0,1(76 - 72) = 72 + 0,1 \times 4 = 72,4$$

Contoh 3.13.

Data terkelompok berikut, nilai desil ke-9 (D_9) adalah

Kelas interval	Frekuensi
50 – 59	2
60 – 69	4
70 – 79	3

Penyelesaian :

Data pada tabel diatas dapat dituliskan kembali dalam bentuk seperti ini.

Kelas interval	Frekuensi	F kumulatif
50 – 59	2	2
60 – 69	4	6
70 – 79	3	9

$$\text{Letak desil ke tujuh } D_9 = \frac{9}{10}(9+1) = 9$$

Kelas desil 9 adalah (70 – 79).

$$BB = 69,5$$

$$p = 10$$

$$n = 9$$

$$F = 6, \text{ dan } fd = 3$$

$$D_9 = 69,5 + 10 \left[\frac{\frac{9}{10} \times 9 - 6}{3} \right]$$

$$D_9 = 69,5 + 10 \left[\frac{8,1 - 6}{3} \right] = 69,5 + 7 = 76,5$$

C. Persentil

Jika data dibagi menjadi 100 bagian yang sama, maka ukuran itu disebut persentil. Letak persentil dirumuskan dengan:

$$LP_i = \frac{i}{100}(n+1) \text{ dengan } i=1,2,3,\dots,99$$

Contoh 3.14.

Tentukan P_{35} dan P_{90} untuk data berikut ini : 51, 55, 56, 59, 63, 65, 69, 74, 74, 85, 91, 93, 96.

Banyaknya data, $n = 13$.

$$LP_{35} = \frac{35}{100}(13+1) = \frac{35 \times 14}{100} = 4,9 \text{ dan } LP_{90} = \frac{90}{100}(13+1) = \frac{90 \times 14}{100} = 12,6$$

$$P_{35} = X_{(4)} + 0,9(X_{(5)} - X_{(4)}) = 59 + 0,9(53 - 59) = 59 + 3,6 = 62,6$$

dan

$$P_{90} = X_{(12)} + 0,6(X_{(13)} - X_{(12)}) = 93 + 0,6(96 - 93) = 93 + 1,8 = 94,8$$

Contoh 3.15.

Diketahui data sebagai berikut :

Berat badan	f (Kg)
31 – 39	2
40 – 48	6
49 – 57	14
58 – 66	18
67 – 75	10

Dari data di atas, maka persentil ke sepuluh (P_{10}) adalah

Penyelesaian :

Tabel di atas dapat dituliskan lagi sebagai

Berat badan	f (Kg)	F kumulatif
31 – 39	2	2
40 – 48	6	8
49 – 57	14	22

58 – 66	18	40
67 – 75	10	50

Lokasi P10.

$$LP_{10} = \frac{10}{100}(50+1) = \frac{510}{100} = 5,1$$

Data yang ke 5,1 terletak pada kelas interval 40 – 48, sehingga kelas tersebut disebut sebagai kelas persentil ke 10.

$$BB = 39,5$$

$$p = 9$$

$$n = 50$$

$$F = 2$$

$$fp = 6$$

$$P_{10} = 39,5 + 9 \left(\frac{\frac{10}{100} \cdot 50 - 2}{6} \right) = 39,5 + 9 \cdot \left(\frac{3}{6} \right) = 44$$

3.3. UKURAN DISPERSI

Ukuran penyebaran ini menunjukkan suatu variasi dari suatu distribusi data. Dengan mengetahui variasi suatu data maka kita bisa mengambil kesimpulan secara lebih tepat tentang distribusi suatu data. Pada ukuran pemusatan, beberapa distribusi angka dengan rata-rata yang sama bisa memiliki ukuran variasi yang berbeda.

Lihat ilustrasi berikut ini, Tabel Distribusi nilai mahasiswa

Matakuliah	Nilai mahasiswa					Rata-rata
A	50	50	50	50	50	50
B	50	75	65	30	30	50
C	50	30	40	85	45	50

Rata-rata nilai mahasiswa untuk setiap matakuliah (A, B, dan C) tersebut diatas adalah sama. Namun demikian kita akan memberikan interpretasi yang berbeda bila mencermati lebih jauh tentang variasi dari masing-masing matakuliah. Penghitungan ukuran penyebaran atau ukuran dispersi ini penting karena:

- 1) Akan diperoleh informasi tambahan tentang penyebaran yang terjadi pada suatu distribusi.
- 2) Dapat menilai ketepatan ukuran nilai tengah dalam mewakili distribusinya. Bila suatu distribusi data memiliki dispersi yang besar maka ukuran nilai tengah kurang mewakili distribusinya, sebaliknya bila nilai dispersi semakin kecil maka semakin tepat nilai tengah mewakili distribusi datanya.

Ukuran penyebaran yang akan dijelaskan adalah rentang (*range*), rata-rata simpangan (*mean deviation*), simpangan baku (*standard deviation*), variansi (*variance*) dan koefisien variasi (*Coefficient of variation*).

A. Rentang

Nilai rentang ini menunjukkan selisih antara data yang paling tinggi dengan data yang paling rendah. Dengan melihat ukuran ini maka dapat diketahui gambaran secara kasar tentang variasi suatu distribusi data. Nilai range ini sangat kasar, karena tidak mempertimbangkan nilai-nilai yang lain selain nilai ekstimnya. Adapun rumusnya adalah sebagai berikut:

$$range = data\ terbesar - data\ terkecil$$

Contoh 3.15.

Pada data Tabel satu diperoleh *range* pada masing masing mata kuliah adalah:

Matakuliah	Max – min	Range
A	50 - 50	0
B	75 - 30	45
C	85 - 30	55

Maka secara kasar dapat disimpulkan bahwa sebaran nilai pada matakuliah C paling bervariasi dibandingkan dengan nilai mata kuliah A dan B. Nilai range sama dengan 0 pada matakuliah A menunjukkan bahwa distribusi nilai A adalah homogen. Semakin besar nilai rentang maka distribusi data semakin bervariasi.

B. Simpangan Rata-rata

Untuk menutup kekurangan dari nilai range maka bisa dihitung nilai simpangan rata-rata (Mean Deviation). Simpangan rata-rata (SR) memperhitungkan nilai-nilai lain selain nilai ekstim distribusi data. Adapun rumusnya adalah sebagai berikut:

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

Sebagai contoh mata kuliah B mempunyai rata-rata = 50 maka simpangan rata-rata :

$$SR = \frac{|50 - 50| + |75 - 50| + |65 - 50| + |30 - 50| + |30 - 50|}{5}$$

$$SR = \frac{|0| + |25| + |15| + |-20| + |-20|}{5}$$

sehingga SR matakuliah B adalah 16.

Dengan menggunakan rumus yang sama maka SR mata kuliah A adalah 0 dan mata kuliah C adalah 14. Maka dapat disimpulkan bahwa nilai B lebih bervariasi daripada nilai A dan C. Mata kuliah A mempunyai distribusi nilai yang homogen.

C. Koefisien Variansi (Coefficient of Variance)

Koefisien variasi merupakan suatu ukuran variansi yang dapat digunakan untuk membandingkan suatu distribusi data yang mempunyai satuan yang berbeda. Kalau kita membandingkan berbagai variansi atau dua variabel yang mempunyai satuan yang berbeda maka tidak dapat dilakukan dengan menghitung ukuran penyebaran yang sifatnya absolut.

Sebagai contoh pada suatu pengukuran tinggi badan mahasiswa diperoleh rerata 165 cm dengan standar deviasi 2,5 cm dan hasil penimbangan diperoleh rerata berat badannya adalah 56 kg dengan standar deviasi 1,2 kg. Dari hasil pengamatan ini kita tidak bisa menyimpulkan bahwa tinggi badan mahasiswa lebih bervariasi bila dibandingkan dengan berat badannya.

Untuk mengatasi permasalahan ini maka harus dihitung suatu ukuran penyebaran relatif, yakni Koefisien Variansi (KV). Rumusnya dapat dilihat sebagai berikut:

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

Dalam contoh kasus diatas maka :

- KV tinggi badan adalah 1,5 %
- KV berat badan adalah 2,14 %

Maka dapat disimpulkan bahwa data berat badan lebih bervariasi bila dibandingkan dengan tinggi badan.

Pada suatu pengukuran tinggi badan mahasiswa diperoleh rerata 165 cm dengan standar deviasi 2,5 cm dan hasil penimbangan diperoleh rerata berat badannya adalah 56 kg dengan standar deviasi 1,2 kg.

D. Simpangan Baku dan Variansi

Simpangan baku ini merupakan ukuran penyebaran yang paling banyak digunakan. Ukuran ini dikenalkan oleh *Karl Pearson*. Dengan menggunakan simpangan rata-rata hasil pengamatan penyebaran sudah memperhitungkan seluruh nilai yang ada pada data. Namun demikian karena dalam penghitungan menggunakan nilai absolut maka tidak dapat diketahui arah penyebarannya. Maka dengan simpangan baku kelemahan ini dapat diatasi, yakni dengan cara membuat nilai pangkat 2, sehingga nilai negatif menjadi positif. Simpangan baku ini merupakan ukuran penyebaran yang paling teliti.

Adapun rumusnya adalah sebagai berikut (untuk sampel):

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

atau

$$S = \sqrt{\frac{\sum_{i=1}^n x^2 - (\sum x)^2 / n}{n-1}}$$

Sedangkan rumus yang digunakan untuk populasi hampir sama, namun pembagiannya tanpa dikurangkan dengan angka 1.

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

atau

$$\sigma = \sqrt{\frac{\sum_{i=1}^n x^2 - (\sum x)^2 / n}{n}}$$

Sebagai contoh pada data Tabel 1 dapat dihitung standar deviasinya sebagai berikut :

Matakuliah	Simpangan baku
A	0,00

B	20,31
C	20,92

Maka dapat disimpulkan bahwa nilai mata kuliah C mempunyai simpangan baku yang paling besar sehingga nilainya yang paling bervariasi.

Simpangan baku atau standard deviation merupakan bentuk akar pangkat 2 dari Variansi. Biasanya ukuran variansi ini diberi simbol sebagai S^2 . Sebenarnya yang merupakan ukuran simpangan adalah simpangan baku, namun demikian ukuran variansi ini merupakan ukuran pangkat dua dari simpangan baku, sehingga bisa juga dianggap sebagai ukuran penyebaran.

Sebagai contoh pada data sebelumnya dapat dihitung nilai variansinya adalah sebagai berikut:

Matakuliah	Simpangan baku	Variansi
A	0,00	0
B	20,31	412,5
C	20,92	437,5

Dari hasil penghitungan diatas dapat disimpulkan bahwa nilai mata kuliah C mempunyai nilai yang paling besar, sehingga dapat disimpulkan bahwa nilai mata kuliah inilah yang paling bervariasi dibandingkan dengan nilai mata kuliah A dan B.

Contoh. 3.17.

Misal suatu penelitian menghasilkan data seperti tersaji pada tabel distribusi frekuensi berikut ini.

Kelas interval	Frekuensi
17 – 20	4
21 – 24	15
25 – 28	7
29 – 32	6
33 – 36	21

Maka simpangan baku data tersebut adalah

Penyelesaian :

Data diatas akan disusun kembali dalam tabel berikut :

<i>Kelas interval</i>	f_i	d_i	$f_i d_i^2$	$f_i d_i$
17 – 20	4	-2	16	-8

21 – 24	15	-1	15	-15
25 – 28	7	0	0	0
29 – 32	6	1	6	6
33 – 36	21	2	84	42
Jumlah	53	0	121	25

Rumus untuk menghitung simpangan baku dengan persamaan :

$$s = p \sqrt{\frac{\sum_{i=1}^n f_i d_i^2}{n} - \frac{\sum_{i=1}^n f_i d_i}{n}}$$

$$s = 4 \sqrt{\frac{121}{53} - \frac{25}{53}} = 5,30$$

4

Peluang dan Distribusinya

Dalam statistika kita menggunakan istilah percobaan untuk sembarang proses yang membangkitkan data. Suatu percobaan dapat berupa pelambungan sekeping mata uang. Dalam percobaan demikian hasil yang mungkin adalah berupa munculnya sisi gambar atau sisi muka (angka). Ilustrasi lain dari suatu percobaan adalah ketika kita melakukan pengukuran terhadap curah hujan sepanjang bulan Oktober sampai April. Dalam sebagian besar kasus, hasil percobaan bergantung pada faktor kebetulan, dan oleh karena itu tidak dapat diramalkan dengan pasti, bahkan ketika seorang ahli kimia melakukan percobaan dalam kondisi yang sama, hasil pengukuran yang diperoleh tetap saja berbeda. Ini menunjukkan adanya faktor kebetulan pada hasil percobaan tersebut. Bahkan bila sekeping mata uang dilambungkan berulang-ulang, kita tidak dapat memastikan bahwa pada pelambungan tertentu akan diperoleh sisi gambar misalnya. Tetapi kita mengetahui semua kemungkinan hasil untuk setiap pelambungan.

4.1. PELUANG

Untuk mengetahui karakteristik suatu populasi sering dilakukan dengan menganalisis hanya sebagian data saja (atau sering disebut dengan sampel). Berdasarkan informasi yang terkandung dalam sampel, dilakukan pengambilan kesimpulan terhadap populasinya. Dasar logika dari proses pengambilan kesimpulan tentang suatu populasi dengan menganalisis data sampel adalah probabilitas. Oleh karena itu, pemahaman tentang teori peluang sangat diperlukan dan bersifat mendasar.

Kata “probabilitas” atau “peluang” adalah kata yang biasa dipakai dalam kehidupan sehari-hari. Suatu peristiwa yang mempunyai peluang untuk terjadi mengandung arti bahwa ada harapan peristiwa itu akan terjadi. Jika ada kepastian bahwa suatu peristiwa akan terjadi, maka peluang terjadinya peristiwa itu adalah 1. Jika tidak ada peluang sama sekali bahwa suatu peristiwa akan terjadi, maka peluangnya adalah 0.

Konsep peluang berhubungan dengan pengertian percobaan yang menghasilkan “hasil” yang tidak pasti. Artinya, percobaan yang diulang-

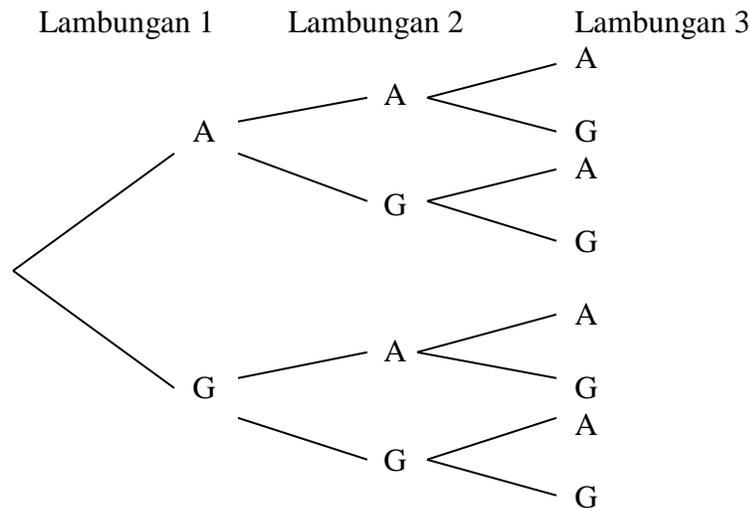
ulang dalam kondisi yang sama akan memberikan “hasil” yang dapat berbeda-beda. Beberapa definisi yang digunakan dalam mempelajari teori peluang adalah seperti:

- 1) Percobaan: Proses pengumpulan data dari sebuah fenomena yang memperlihatkan variasi pada hasilnya
- 2) Ruang sampel (*sample space*) : Himpunan semua hasil yang mungkin dari suatu percobaan.
- 3) Titik sampel : Setiap unsur / elemen / anggota dari ruang sampel.
- 4) Kejadian (peristiwa, event): Hasil dari suatu percobaan yang mempunyai sifat tertentu. Jadi, kejadian adalah himpunan bagian dari ruang sampel.

Contoh 4.1

Sebuah koin yang seimbang dilambungkan tiga kali.

- 1) Percobaan yang dilakukan adalah pelambungan sebuah koin sebanyak tiga kali.
- 2) Hasil yang diperoleh adalah sisi koin yang nampak di atas, dapat sisi “angka, A” atau sisi “gambar, G”.
- 3) Ruang sampel dapat ditentukan dengan menggunakan cara seperti berikut ini.
- 4) Ruang nol atau ruang kosong atau himpunan kosong adalah himpunan bagian ruang sampel yang tidak mengandung satu pun anggota. Kejadian ini diberi lambang khusus \emptyset , Andaikan adalah himpunan bilangan genap yang tidak habis dibagi 2 dari ruang sampel himpunan bilangan bulat, maka $A = \emptyset$.



Ada delapan titik sampel yang dapat dihimpun sebagai ruang sampel, yaitu :

$$S = \{AAA, AAG, AGA, AGG, GAA, GAG, GGA, GGG\} \text{ dan } n(S) = 8.$$

- 1) Misalkan X adalah peristiwa munculnya sisi “angka” sebanyak dua kali dalam percobaan di atas, maka $X = \{AAG, AGA, GAA\}$. Terlihat bahwa X merupakan himpunan bagian dari S.

A. Peluang suatu kejadian

Salah satu cara mendefinisikan peluang suatu kejadian adalah mengucapkannya dalam frekuensi relatif terjadinya kejadian bersangkutan. Definisi demikian ini biasanya berbunyi demikian : Misalkan bahwa suatu percobaan dengan ruang sampel S diulang-ulang di bawah kondisi yang sama. Untuk setiap kejadian A yang merupakan himpunan bagian dari S, kita definisikan $n(A)$ sebagai berapa kali kejadian A terjadi selama kali ulangan percobaan itu. Pada awalnya, peluang didefinisikan dalam pengertian yang klasik berikut ini.

Jika A adalah himpunan kejadian dan S adalah ruang sampelnya dengan $A \subseteq S$, maka :

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

Peluang klasik memiliki ciri-ciri atau sifat-sifat sebagai berikut :

1. $P(A) = \frac{n(A)}{n(S)} \geq 0$
2. $P(S) = \frac{n(S)}{n(S)} = 1$
3. Jika $A \cap B = \emptyset$, maka
 $P(A \cup B) = P(A) + P(B)$

Contoh 4.2

Dua dadu dilambungkan bersama-sama. Ruang sampel yang ada adalah pasangan dari masing-masing sisi yang muncul.

$$S = \left\{ \begin{array}{l} (1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (2,2), (2,3), (2,4), (2,5), (2,6) \\ (3,1), (3,2), (3,3), (3,4), (3,5), (3,6), (4,1), (4,2), (4,3), (4,4), (4,5), (4,6) \\ (5,1), (5,2), (5,3), (5,4), (5,5), (5,6), (6,1), (6,2), (6,3), (6,4), (6,5), (6,6) \end{array} \right\}$$

$$n(S) = 36$$

Jika A adalah himpunan dengan definisi sebagai berikut :

$A = \{x | x \text{ pasangan sisi yang sama}\}$ maka A adalah

$A = \{(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)\}$ sehingga $n(A) = 6$

Tentukan peluang dari himpunan A !

Dari rumus awal, diperoleh

$$P(A) = \frac{n(A)}{n(S)}$$
$$P(A) = \frac{6}{36} = \frac{1}{6}$$

Jika B adalah himpunan dengan definisi sebagai berikut :

$$B = \{x \mid x \text{ jumlah pasangan yang kurang dari } 5\}$$
$$B = \{(1,1), (1,2), (1,3), (2,1), (2,2), (3,1)\} \text{ dengan } n(B) = 6$$

Peluang dari himpunan B adalah

$$P(B) = \frac{n(B)}{n(S)} = \frac{6}{36} = \frac{1}{6}$$

Jika C adalah himpunan yang berisi jumlah kedua sisi adalah bilangan prima, maka :

$$C = \left\{ \begin{array}{l} (1,1), (1,2), (1,3), (1,6), (2,1), (2,3), (2,5), (3,2), (3,4), (4,1), (4,3), \\ (5,2), (5,6), (6,1), (6,5) \end{array} \right\}$$

$$n(C) = 15$$

$$P(C) = \frac{n(C)}{n(S)} = \frac{15}{36} = \frac{5}{12}$$

B. Peluang Bersyarat

Peluang bersyarat B, bila A diketahui dilambangkan dengan $P(B|A)$, didefinisikan sebagai:

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

Contoh 4.3

Peluang suatu penerbangan yang telah terjadwal teratur berangkat tepat waktu $P(B) = 0,83$, peluang sampai tepat waktu $P(S) = 0,82$ dan peluang berangkat dan sampai tepat waktu $P(B \cap S) = 0,78$. Cari peluang bahwa pesawat :

- Sampai tepat waktu bila diketahui berangkat tepat waktu ?
- Berangkat tepat waktu jika diketahui sampai tepat waktu ?

Jawab:

$$(a) P(S|B) = P(B \cap S)/P(B) = 0,94$$

$$(b) P(B|S) = P(B \cap S)/P(S) = 0,95$$

Kejadian saling bebas

Dua kejadian A dan B dikatakan bebas bila $P(B|A) = P(B)$ atau $P(A|B) = P(A)$. Bila hal ini tidak dipenuhi A dan B dikatakan dua kejadian yang tidak bebas

Kaidah Penggandaan. Bila dalam suatu percobaan kejadian A dan B keduanya dapat terjadi sekaligus, maka $P(A \cap B) = P(A)P(B|A)$.

Kaidah penggandaan khusus. Bila dua kejadian A dan B bebas saling bebas, maka

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

Kaidah penggandaan umum. Jika dalam suatu percobaan kejadian-kejadian A_1, A_2, \dots, A_k dapat terjadi, maka :

$$P(A_1 \cap A_2 \cap A_3 \dots \cap A_k) = P(A_1).P(A_2|A_1)P(A_3|A_1 \cap A_2) \dots P(A_k|A_1 \cap A_2 \cap \dots \cap A_{k-1})$$

Jika kejadian-kejadian $A_1, A_2, A_3, \dots, A_k$ semuanya saling bebas, maka

$$P(A_1 \cap A_2 \cap A_3 \dots \cap A_k) = P(A_1).P(A_2)P(A_3) \dots P(A_k)$$

Contoh 4.4

Bila sebuah dadu dan sekeping mata uang yang tidak seimbang dilempar sekaligus, berapa peluang kejadian munculnya sisi mata dadu yang kurang dari 3 dan sisi muka, jika munculnya sisi muka 2 kali lebih sering daripada sisi belakang ?

Jawab:

Misalkan A muncul sisi mata dadu yang kurang dari 3, $A = \{1,2\}$, $P(A) = 1/3$. B muncul sisi muka, $B = \{M\}$, tetapi karena munculnya sisi muka dua kali lebih sering dari sisi belakang, maka $P(B) = 2/3$.

Karena A dan B merupakan dua kejadian yang saling bebas, maka :

$$P(A \cap B) = P(A).P(B) = 2/9$$

C. Kaidah Bayes

Bila kejadian-kejadian $B_1, B_2, \dots, B_k, \neq 0$ untuk $i=1,2,\dots,k$, maka untuk sembarang kejadian yang merupakan himpunan bagian S berlaku

$$P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + \dots + P(B_k)P(A|B_k)$$

Jika kejadian-kejadian B_1, B_2, \dots, B_k merupakan sekatan dari ruang S dengan $P(B_i) \neq 0$ untuk $i=1,2,\dots,k$, maka untuk sembarang kejadian A yang bersifat $P(A) \neq 0$,
 Untuk $r=1,2,\dots,k$

$$P(B_r | A) = \frac{P(B_r)P(A | B_r)}{P(B_1)P(A | B_1) + P(B_2)P(A | B_2) + \dots + P(B_k)P(A | B_k)}$$

Contoh 4.5

Tiga anggota formatur Dewan Mahasiswa dicalonkan menjadi ketua. Peluang Ali terpilih adalah 0,3; peluang Bambang terpilih adalah 0,5; sedangkan peluang Candra adalah 0,2. Apabila Ali terpilih, maka peluang kenaikan iuran organisasi adalah 0,8. Apabila Bambang atau Candra yang terpilih, maka peluang kenaikan iuran adalah masing-masing 0,1 dan 0,4. Berapakah peluang iuran akan naik ?

Jawab :

Perhatikan kejadian sebagai berikut.

- A = Orang yang terpilih menaikkan iuran
- B_1 = Ali yang terpilih
- B_2 = Bambang yang terpilih
- B_3 = Candra yang terpilih.

Berdasarkan teorema jumlah probabilitas, maka diperoleh :

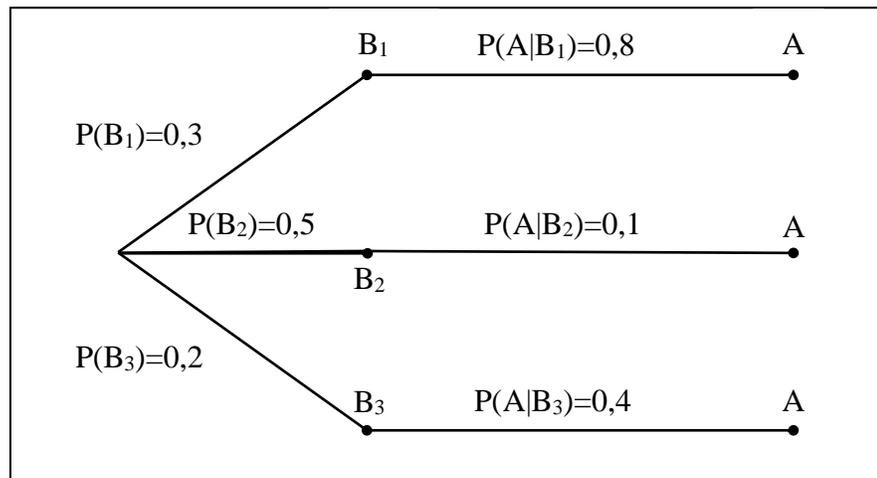
$$P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3)$$

Dengan melihat diagram pohon pada Gambar di bawah ini, terlihat bahwa ketiga cabang mempunyai probabilitas

$$P(B_1)P(A|B_1) = (0,3)(0,8) = 0,24$$

$$P(B_2)P(A|B_2) = (0,5)(0,1) = 0,05$$

$$P(B_3)P(A|B_3) = (0,2)(0,4) = 0,08.$$



Jadi $P(A) = 0,24 + 0,05 + 0,08 = 0,37$.

4.2. DISTRIBUSI PELUANG

Hasil percobaan diberikan secara lengkap oleh suatu ruang sampel S dengan fungsi peluang $P(\cdot)$ pada peristiwa atau kejadiannya. Dengan mengamati suatu $\omega \in S$ yang dipilih berdasarkan $P(\cdot)$. Kita dapat menjawab pertanyaan mengenai percobaan tersebut

A. Variabel Random Diskret

Dari suatu percobaan seringkali kita tertarik terutama pada suatu fungsi hasil percobaannya, bukan pada hasil-percobaannya itu sendiri. Sebagai contoh, dalam pelemparan dua dadu kita sering tertarik pada jumlah mata dadu yang muncul dari kedua dadu tersebut, dan bukan pada hasil masing-masing dadu. Di dalam pelemparan uang logam, kita mungkin tertarik pada total banyaknya sisi gambar yang muncul dan tidak peduli pada urutan gambar-angka yang muncul sesungguhnya. Besaran-besaran yang menjadi perhatian kita ini, dalam hal ini jumlah mata dadu yang muncul untuk pelemparan dadu dan total banyaknya sisi gambar yang muncul dalam pelemparan mata uang, atau lebih formalnya disebut fungsi bernilai real yang didefinisikan pada ruang percobaannya ini, dikenal sebagai variabel random.

Misalkan, sebuah koin dilambungkan tiga kali. X adalah banyaknya sisi “angka” yang muncul dalam percobaan tadi. Maka X akan menjalani nilai 0 jika hasilnya GGG, bernilai 1 jika AGG, GGA, dan GAG, bernilai 2 jika AAG, AGA, dan GAA, serta bernilai 3 jika muncul AAA. Variabel X dikatakan variabel random, dituliskan $X = i$, dengan $i = 0, 1, 2, 3$.

Jika X adalah himpunan dari nilai-nilai variabel random yang mungkin terjadi, yang terdiri atas x_1, x_2, \dots, x_n atau x_1, x_2, \dots maka X disebut variabel random diskret. Bagi suatu variabel random diskret X , didefinisikan fungsi massa peluang $p(a)$ sebagai

$$p(a) = P(X = a)$$

Fungsi massa peluang $p(a)$ bernilai positif untuk paling banyak sejumlah tercacah nilai a . Dengan kata lain, jika X mengambil salah satu dari nilai-nilai x_1, x_2, \dots maka

$$p(x_i) \geq 0$$

$$p(x) = 0 \text{ untuk semua nilai } x \text{ yang lainnya.}$$

Karena X pasti mengambil salah satu dari nilai-nilai x_i , maka

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

Fungsi distribusi kumulatif dari variabel random X didefinisikan sebagai:

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

untuk suatu bilangan real $x \in \mathbb{R}$

Jika X adalah suatu variabel random diskret dengan nilai-nilai yang mungkin diambil ialah x_1, x_2, x_3, \dots dengan $x_1 < x_2 < x_3 < \dots$, maka fungsi distribusinya F berupa sebuah fungsi tangga. Dengan kata lain, nilai F konstan pada interval $[x_{i-1}, x_i)$ dan kemudian meloncat sebesar $p(x_i)$ di x_i . Sebagai contoh, jika X mempunyai fungsi massa probabilitas

$$p(1) = 1/4, \quad p(2) = 1/2 \quad p(3) = 1/8 \quad p(4) = 1/8$$

maka fungsi distribusi kumulatifnya adalah

$$F(a) = \begin{cases} 0 & a < 1 \\ \frac{1}{4} & 1 \leq a < 2 \\ \frac{3}{4} & 2 \leq a < 3 \\ \frac{7}{8} & 3 \leq a < 4 \\ 1 & 4 \leq a \end{cases}$$

1. Distribusi Bernoulli dan Binomial

Pada suatu percobaan yang hanya ada dua hasil yang mungkin yaitu “sukses” atau “gagal”. Percobaan yang demikian juga disebut percobaan Bernoulli. Fungsi peluang distribusi Bernoulli dituliskan dalam bentuk :

$$P(X = i) = p^i (1 - p)^{1-i}$$

dengan $i = 0$ untuk percobaan “gagal” dan $i = 1$ untuk “sukses” dan p adalah peluang akan “sukses”.

Jika suatu percobaan Bernoulli diulang-ulang sampai n kali percobaan, maka percobaan tersebut dikatakan memiliki distribusi Binomial dengan fungsi peluangnya adalah :

$$P(X = i) = \binom{n}{i} p^i (1-p)^{n-i} \quad \text{dengan } i = 0, 1, 2, \dots, n.$$

Contoh 4.6

Peluang seorang mahasiswa akan sukses memperoleh nilai A dalam ujian skripsi adalah 35%. Jika ada 10 mahasiswa yang maju ujian, hitung peluang akan ada:

- Tepat dua orang memperoleh A.
- Paling banyak delapan orang memperoleh A
- Sedikitnya tiga orang memperoleh A

Peluang seorang akan memperoleh nilai A adalah $p = 0,35$ dan banyaknya mahasiswa yang mengikuti ujian $n = 10$.

- Untuk $x = 2$ diperoleh

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$P(X = 2) = \frac{10!}{2!(10-2)!} 0,35^2 (0,65)^8 = 0,1757$$

Jika diselesaikan menggunakan MS Excel, perintah yang dituliskan adalah sebagai berikut:

$$=binom.dist(x,n,pejuang,false)$$

$$=binom.dist(2,10,0.35,false)$$

$$= 0,175653$$

- Untuk $X \leq 8$ diperoleh

$$P(X \leq 8) = 1 - P(> 8) = 1 - \{P(X = 9) + P(X = 10)\}$$

$$P(X \leq 8) = 1 - P(> 8) = 1 - \left\{ \left(\frac{10!}{1!9!} 0,35^9 (1-0,35)^1 \right) + \left(\frac{10!}{10!} 0,35^{10} (1-0,35)^0 \right) \right\}$$

$$P(X \leq 8) = 0,99946$$

Perintah menggunakan MS Excel dapat dilakukan sebagai berikut:

$$=binom.dist(8,10,0.35,true)$$

$$= 0,99946$$

- Untuk $X \geq 3$ diperoleh

$$P(X \geq 3) = 1 - P(X < 3) = 1 - P(X \leq 2)$$

$$P(X \geq 3) = 1 - \{P(X = 0) + P(X = 1) + P(X = 2)\}$$

$$P(X \geq 3) = 1 - \left\{ \binom{10!}{10!} 0,35^0 (1-0,35)^{10} + \binom{10!}{1!9!} 0,35^1 (1-0,35)^9 + \binom{10!}{2!8!} 0,35^2 (1-0,35)^8 \right\}$$

$$P(X \geq 3) = 1 - \{(0,0135) + (0,0725) + (0,1757)\} = 1 - 0,2616 = 0,7384$$

Dengan menggunakan Excel, yang kita lakukan adalah sebagai berikut:

$$\begin{aligned} &= 1 - \text{binom.dist}(3, 10, 0.35, \text{true}) \\ &= 0,7384 \end{aligned}$$

2. Distribusi Geometrik

Percobaan geometrik didefinisikan sebagai suatu percobaan pada percobaan Bernoulli yang dilakukan hingga mencapai sukses yang pertama, jika p adalah peluang kejadian akan sukses. Fungsi peluang untuk distribusi Geometrik didefinisikan sebagai :

$$P(X = x) = (1-p)^{x-1} p$$

dengan $i = 1, 2, 3, \dots$ dan $0 \leq p \leq 1$

Contoh 4.7

Peluang seorang pelamar kerja akan berhasil dalam wawancara adalah 0,3 (selama ini dia gagal dalam tahap tersebut). Berapa peluang dia akan diterima setelah mengikuti tahap wawancara sampai lima kali?

Jawab :

$$P(X = x) = (1-p)^{x-1} p$$

$$P(X = 5) = (1-0,3)^{5-1} 0,3$$

$$P(X = 5) = (0,7)^4 0,3 = 0,072$$

3. Distribusi Poisson

Jika variabel random X menyatakan banyaknya hasil percobaan yang terjadi dalam selang waktu tertentu, dan μ adalah rata-rata banyaknya hasil percobaan dalam selang waktu tersebut, maka X berdistribusi Poisson dengan distribusi peluang

$$P(X = x) = \frac{e^{-\mu} \mu^x}{x!}, \quad x = 0, 1, 2, \dots$$

Nilai harapan dan variansi dari variabel random yang berdistribusi Poisson keduanya sama dengan μ .

Ciri-ciri percobaan Poisson :

- 1) Banyaknya hasil percobaan yang terjadi dalam suatu selang waktu tertentu, tidak tergantung pada banyaknya hasil percobaan yang terjadi pada selang waktu lain yang terpisah.
- 2) Peluang terjadinya suatu hasil percobaan selama selang waktu yang singkat, sebanding dengan panjang selang waktu tersebut, dan tidak tergantung pada banyaknya hasil percobaan yang terjadi di luar selang waktu tersebut.
- 3) Peluang lebih dari satu hasil percobaan akan terjadi dalam selang waktu yang singkat, dapat diabaikan.

Contoh 4.8

Dalam satu minggu, rata-rata terdapat 3 mahasiswa yang siap maju ujian skripsi di Jurusan Tarbiyah. Tentukan berapa peluang dalam satu minggu akan terdapat:

- a. Tepat 4 mahasiswa maju skripsi
- b. Paling banyak 3 mahasiswa
- c. Sedikitnya 3 mahasiswa

Jawab :

Diketahui $\mu = 3$

a. Untuk $X = 4$, diperoleh :

$$P(X = 4) = \frac{e^{-3} 3^4}{4!} = 0,168$$

Jadi peluang akan terdapat 4 mahasiswa yang siap ujian skripsi dalam satu minggu adalah 0,168.

b. Paling banyak tiga mahasiswa, dapat dituliskan dalam bentuk :

$$P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$$

$$P(X \leq 3) = \frac{e^{-3} 3^0}{0!} + \frac{e^{-3} 3^1}{1!} + \frac{e^{-3} 3^2}{2!} + \frac{e^{-3} 3^3}{3!}$$

$$P(X \leq 3) = e^{-3} \left(1 + 3 + \frac{9}{2} + \frac{27}{6} \right) = e^{-3} (13) = 0,647$$

Jadi, peluang paling banyak tiga mahasiswa yang siap maju adalah 0,647.

c. Sedikitnya tiga mahasiswa, dapat dirumuskan sebagai:

$$P(X \geq 3) = 1 - P(X < 3)$$

$$P(X \geq 3) = 1 - P(X \leq 2) = 1 - (P(X = 0) + P(X = 1) + P(X = 2))$$

$$P(X \geq 3) = 1 - \left(e^{-3} \left(1 + 3 + \frac{9}{2} \right) \right) = 0,577$$

Distribusi Poisson dapat digunakan untuk menyelesaikan kasus Binomial yang memiliki peluang sukses sangat kecil dan banyaknya percobaan sangat besar. Pendekatan yang dilakukan dengan cara menentukan parameter dari Poisson $\mu = n.p$.

Contoh 4.9

Peluang seorang mahasiswa akan mengalami Drop Out (DO) karena suatu masalah administrasi adalah 0,002. Carilah peluang jika 2000 mahasiswa aktif, terdapat 5 orang akan mengalami DO !

Jawab :

Diketahui $p = 0,002$ dan $n = 2000$

Jika digunakan Binomial,

$$P(X = 5) = \binom{2000}{5} 0,002^5 (1 - 0,002)^{2000-5}$$

$$P(X = 5) = \frac{2000!}{5!1995!} 0,002^5 \cdot 0,998^{1995}$$

Hitungan di atas sulit diperoleh dengan hitungan manual, sehingga kita perlu menggunakan pendekatan Poisson.

$$\mu = n.p = 2000 \cdot 0,002 = 4$$

$$P(X = 5) = \frac{e^{-4} 4^5}{5!} = e^{-4} \cdot 8,533 = 0,156$$

B. Variabel Random Kontinu

Variabel random X dikatakan variabel random kontinu jika terdapat fungsi $f(x)$ yang disebut sebagai fungsi densitas peluang dari X, sedemikian sehingga fungsi distribusi kumulatifnya dapat ditulis sebagai:

$$F(x) = \int_{-\infty}^x f(t)dt$$

Di bawah ini beberapa distrib usi variabel kontinu yang sering digunakan dalam penelitian pendidikan.

1. Distribusi Eksponensial

Variabel random X disebut mempunyai distribusi eksponensial dengan parameter λ bila memiliki fungsi kepadatan peluang :

$$f(x) = \frac{e^{-\frac{x}{\lambda}}}{\lambda}$$

dengan $x \geq 0$

Distribusi eksponensial dalam praktek sering muncul sebagai distribusi lamanya waktu sampai suatu kejadian tertentu terjadi. Sebagai contoh, lamanya waktu (mulai sekarang) sampai terjadi gempa bumi, atau sampai pecah perang baru, atau sampai deringan telepon yang Diterima ternyata salah sambung semuanya adalah contoh kasus variabel random eksponensial.

Contoh 4.10

Misalkan bahwa panjang pembicaraan telepon, dalam menit, adalah suatu variabel random eksponensial dengan parameter $\lambda = 10$. Jika seseorang tiba dahulu dari kita di sebuah wartel, hitunglah peluang bahwa kita terpaksa harus menunggu

- a. lebih dari 10 menit
- b. antara 10 dan 20 menit

Jawab :

Misalkan X adalah lamanya pembicaraan oleh orang pada wartel tersebut. Maka peluang yang dicari adalah

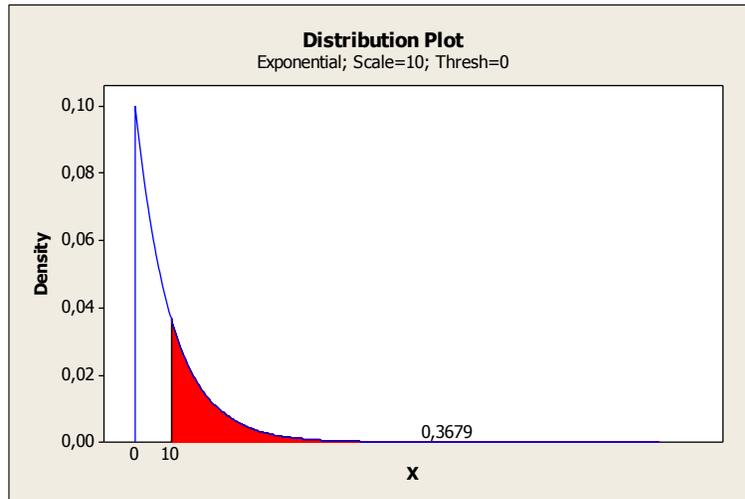
$$a. P\{X > 10\} = \int_{10}^{\infty} \frac{1}{10} e^{-x/10} dx = -e^{-x/10} \Big|_{10}^{\infty} = e^{-1} \approx 0,3679$$

Jika diselesaikan dengan Minitab, langkah-langkahnya sebagai berikut:

- 1) Pilih Graph
- 2) Pilih probability Distribution Plot
- 3) Pilih view probability, kemudian OK
- 4) Pilih pada kotak distribution : Exponential

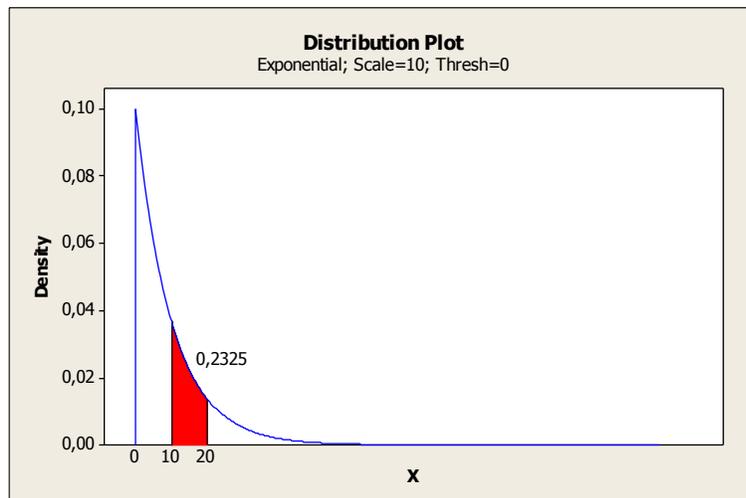
- 5) Tuliskan dalam kotak scale : 10
- 6) Pindahkan ke shaded area. Pilih right tile
- 7) Klik Xvalue, masukkan dalam kotak X value : 10.
- 8) Klik OK

Hasilnya seperti gambar di bawah ini.



$$b. P\{10 < X < 20\} = \int_{10}^{20} \frac{1}{10} e^{-x/10} dx = -e^{-x/10} \Big|_{10}^{20} = 0,2325$$

Dengan cara yang sama seperti langkah (a), pilihlah middle area pada kotak shaded area. Pilihlah X value dengan X value 1 : 10 dan X value 2: 20. Hasilnya seperti gambar di bawah ini.



2. Distribusi Normal dan Normal Baku

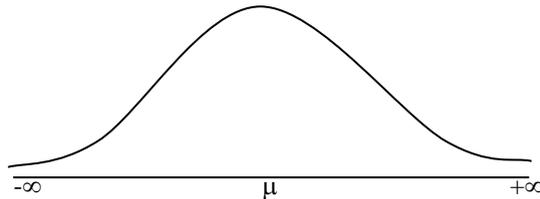
Apabila X merupakan sebuah variabel yang mengikuti distribusi normal dengan rata-rata μ dan simpangan baku σ , maka fungsi densitas untuk X adalah:

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

Distribusi Normal memiliki ciri-ciri sebagai berikut:

- Variabelnya merupakan variabel kuantitatif kontinu, yang harganya bisa mengambil setiap harga dari $-\infty$ sampai $+\infty$.
- Bentuk kurva distribusi peluangnya adalah simetrik sekitar harga $X = \mu$ (μ adalah rata-rata untuk X), tersebar sepanjang harga-harga X dengan simpangan baku sebesar σ .

Kurva distribusi normal dengan rata-rata μ dan simpangan baku σ



Menghitung peluang langsung dari fungsi di atas merupakan tugas yang sukar, karena melibatkan penghitungan integral yang tidak sederhana. Oleh karena itu peluang untuk harga-harga X dicari melalui suatu bentuk transformasi ke dalam distribusi Normal Standar (baku). Distribusi normal baku (distribusi normal dengan harga rata-rata 0 dan simpangan baku 1) diperoleh melalui transformasi variabel, dalam bentuk:

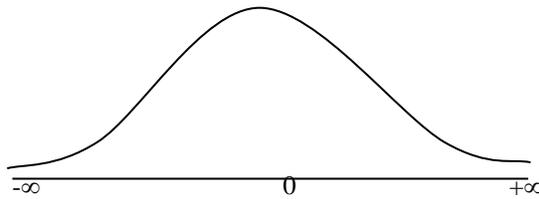
$$Z = \frac{X - \mu}{\sigma}$$

Variabel Z akan mengikuti distribusi normal dengan rata-rata 0 dan simpangan baku 1, dengan fungsi densitas berbentuk

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}z^2); -\infty < z < \infty$$

Bentuk kurva distribusi peluangnya adalah simetrik sekitar 0 (harga rata-ratanya), tersebar sepanjang harga variabel yang bersangkutan, dengan simpangan baku sekitar 1.

Bentuk kurva distribusi normal dengan rata-rata 0 dan simpangan baku 1, normal standar seperti gambar di bawah ini.

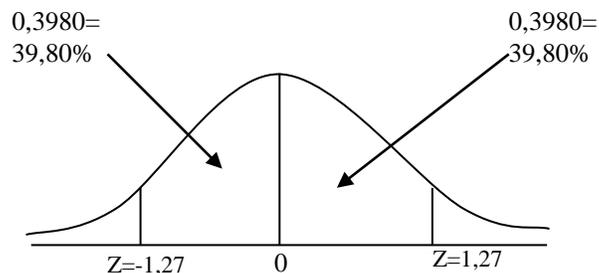


Besarnya peluang untuk harga-harga Z tertentu telah dibuatkan tabelnya, yang untuk selanjutnya akan kita sebut tabel luas daerah di bawah kurva normal baku, disingkat menjadi tabel Normal Standar.

Cara Mempergunakan Tabel Normal

Bilangan-bilangan di samping dan di bawah huruf Z menyatakan harga-harga Z yang mungkin. Jika $Z = 1,27$, ke bawah Z kita cari 1,2 dan ke kanan Z dicari 0,07.

Bilangan-bilangan dalam badan tabel memperlihatkan luas daerah-di bawah-kurva-normal dinyatakan dalam bentuk proporsi (persentase), dari titik $Z = 0,00$ (yang merupakan harga rata-rata distribusi normal baku) ke titik Z tertentu. Jadi, misalnya diperoleh bilangan 0,3980. Ini artinya bahwa luas-daerah-di bawah-kurva normal dari titik $Z = 0,00$ ke $Z = 1,27$ meliputi 0,3980 bagian, atau 39,80%, dari seluruh luas daerah di bawah kurva.



Oleh karena kurva normal baku adalah simetris sekitar titik 0 (rata-ratanya, dalam hal ini $Z = 0,00$), maka apabila diperoleh harga $Z 1,27$, luas daerah di bawah kurva dari titik $Z = 0,00$ ke kiri sampai $Z = -1,07$ adalah juga 0,3980 atau 39,80%.

Jadi bilangan-bilangan dalam badan tabel menyatakan luas daerah kurva (dalam bentuk proporsi/persentase), dari titik $Z = 0,00$ ke kiri (ke arah Z -negatif) sampai suatu titik Z negatif tertentu, atau, dari titik $Z = 0,00$ ke kanan (arah Z -positif) sampai titik Z -positif tertentu.

Luas daerah di bawah kurva ini oleh statistika diinterpretasikan sebagai peluang.

Jadi

$$\begin{aligned} F(0 < Z < 1,27) &= \text{luas di bawah kurva normal (baku) dari titik } Z \\ &= 0,00 \text{ ke } Z = 1,27 \\ &= 0,3980 \end{aligned}$$

$$\begin{aligned} F(-1,27 < Z < 0) &= \text{Luas daerah kurva normal (baku) dari titik } Z \\ &= -1,27 \text{ sampai titik } Z = 0,00 \\ &= 0,3980. \end{aligned}$$

Untuk menghitung nilai peluang x terletak diantara dua nilai a dan b , adalah

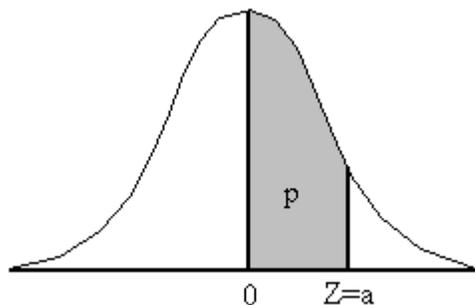
$$P(a \leq x \leq b) = \int_a^b \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

Gunakan transformasi $Z = \frac{(x - \mu)}{\sigma}$, sehingga

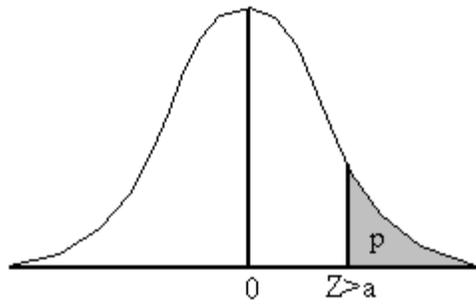
$$\begin{aligned} P(a \leq x \leq b) &= P\left(\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right) \\ &= \int_{z_1}^{z_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z^2} dz \end{aligned}$$

Nilai-nilai tersebut dapat dihitung dengan menggunakan daftar distribusi normal baku. Besarnya nilai peluang distribusi normal Z dapat digambarkan seperti berikut ini

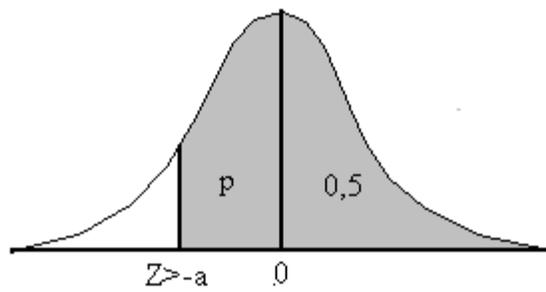
a. $P(Z = a) = p$



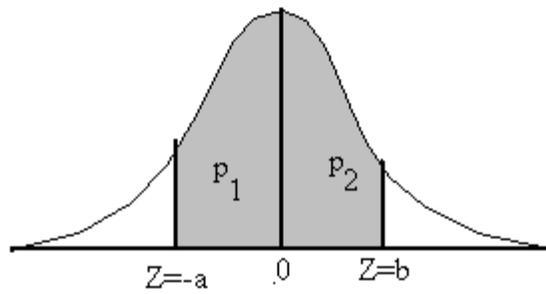
b. $P(Z \geq a) = 0,5 - p$



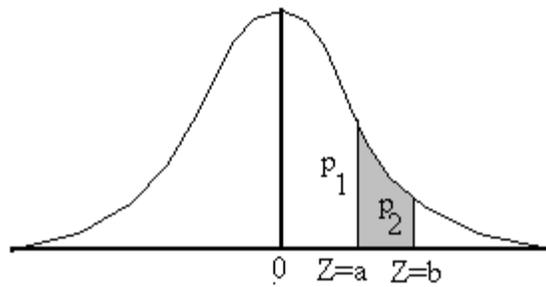
c. $P(Z \geq -a) = 0,5 + p$



d. $P(-a \leq Z \leq b) = p_1 + p_2$



e. $P(a \leq Z \leq b) = p_2 - p_1$



Contoh 4.11.

Berat kiriman paket dalam suatu biro jasa pengiriman barang dapat dianggap berdistribusi normal dengan rata-rata 12,5 kg dengan simpangan baku 0,8 kg. Berapa persenkah berat barang kiriman yang memiliki berat antara 11 sampai 13 kg

Jawab:

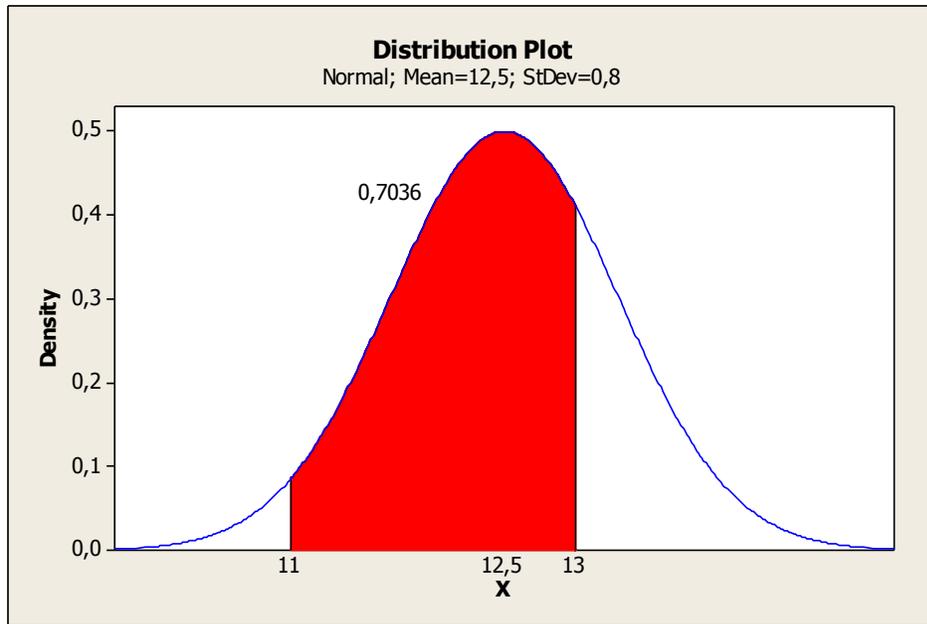
$$\begin{aligned}P(11 < X < 13) &= P\left(Z < \frac{13-12,5}{0,8}\right) - P\left(Z < \frac{11-12,5}{0,8}\right) \\&= P(Z < 0,63) - P(Z < -1,88) \\&= 0,2357 + 0,4679 \\&= 0,7036\end{aligned}$$

Jadi berat paket antara 11 sampai 13 kg adalah 70,36%.

Jika kita menggunakan paket Minitab, langkah-langkahnya adalah sebagai berikut:

- 1) Pada menu bar, pilih graph
- 2) Pilih probability Distribution Plot
- 3) Pilih view probability, kemudian OK
- 4) Pilih pada kotak distributin : Normal
- 5) Masukkan mean : 12,5
- 6) Masukkan standart deviation : 0,8
- 7) Klik shaded area
- 8) Klik x value
- 9) Pilih middle
- 10) Masukkan X value 1 : 11 dan X value 2 : 13, klik OK

Hasil yang diperoleh seperti gambar di bawah ini.



3. Distribusi Chi Kuadrat

Variabel random X berdistribusi chi-kuadrat jika fungsi padatnya

$$\text{diberikan oleh } f(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-x/2} & x > 0 \\ 0 & x \text{ lainnya} \end{cases}$$

dengan n bilangan bulat positif. Distribusi Chi-Kudrat ini juga dilambangkan dengan χ^2 . Yang perlu diperhatikan dalam menentukan nilai dari distribusi χ^2 adalah derajat kebebasan (ν) dengan $\nu = n - 1$ dan nilai peluang (α).

Peluang suatu kejadian dalam distribusi Chi Kuadrat dapat dihitung dengan $P(a \leq X \leq b) = \int_a^b f(x) dx = \int_a^b \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-x/2} dx$ atau dapat dicari

lewat tabel distribusi Chi-Kuadrat.

Contoh 4.12

Diketahui variabel random X berdistribusi chi-kuadrat dengan derajat kebebasan 8. Dengan menggunakan tabel distribusi chi-kuadrat, hitunglah:

- a. $P(X < 17,5)$
- b. $P(X > 15,5)$
- c. $P(15,5 < X < 17,5)$

Penyelesaian :

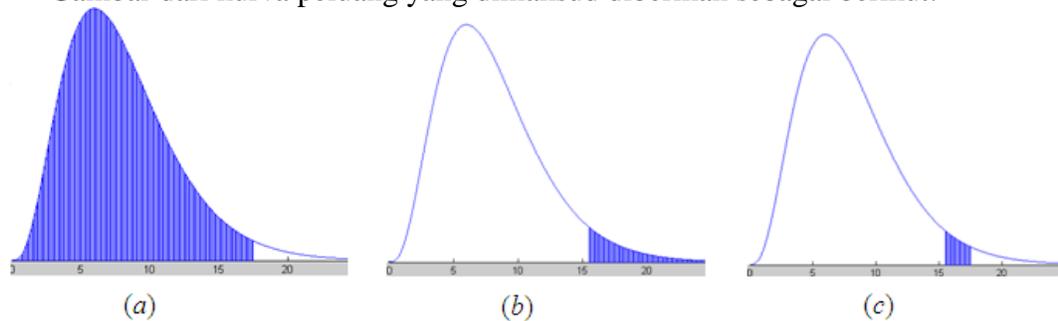
- a. Pada kolom pertama (kolom derajat kebebasan) pada tabel, pilih angka 8, kemudian tarik ke kanan hingga diperoleh angka 17,5. Selanjutnya tarik ke atas dan diperoleh nilai 0.975.

Jadi $P(X < 17.5) = 0.975$

- b. Dengan cara yang sama, dari angka 8, tarik kekanan hingga diperoleh nilai 15,5. Selanjutnya tarik ke atas dan diperoleh nilai 0.95. Nilai tersebut adalah $P(X < 15,5)$. Hubungan antara $P(X < 15,5)$ dengan $P(X > 15,5)$ adalah $P(X < 15,5) + P(X > 15,5) = 1$. Jadi nilai dari $P(X > 15,5) = 1 - P(X < 15,5) = 1 - 0.95 = 0.05$

- c. $P(15,5 < X < 17,5) = P(X < 17,5) - P(X < 15,5) = 0.975 - 0.95 = 0.025$

Gambar dari kurva peluang yang dimaksud diberikan sebagai berikut:



Gambar 5.15 Luas daerah di bawah kurva chi-kuadrat dengan (a) $P(X < 17.5) = 0.975$; (b) $P(X < 15,5) = 0.95$ dan (c) $P(15,5 < X < 17,5) = 0.7$

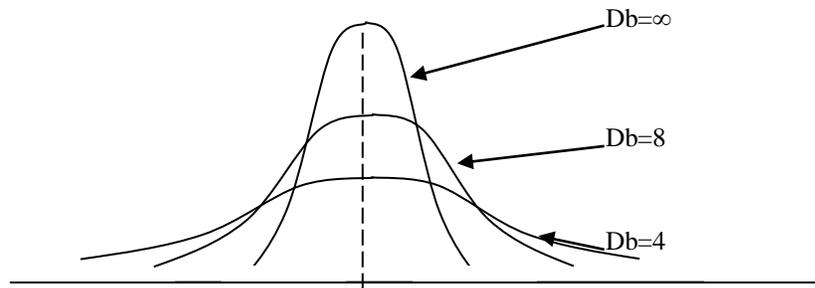
4. Distribusi student' t

Distribusi yang tak kalah pentingnya dengan distribusi normal adalah distribusi t-student. Sekilas tentang distribusi t-student ini berasal dari W.S Gosset. W.S Gosset pada tahun 1908 telah mempublikasikan suatu makalah tentang distribusi bagi sampel kecil dan dikenal dengan distribusi t-student. Kata student digunakan karena dalam publikasi itu Gosset menggunakan samaran “student”. Statistik sampel yang penting, seperti rata-rata sampel, bila ukuran sampelnya kecil (< 30) tidak lagi mengikuti distribusi normal, tetapi mendekati distribusi t-student atau biasa disingkat dengan distribusi student. Distribusi t juga simetrik terhadap mean = 0 tetapi variansinya lebih besar dari 1. Semakin ukuran sampel (n) semakin besar variansinya dan sebaliknya semakin besar ukuran sampel semakin kecil variansinya. Bila $n \rightarrow \infty$ maka kurva t menyerupai kurva Z.

Fungsi densitas peluang X yang berdistribusi t student :

$$f(x) = \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})} \frac{1}{\sqrt{v\pi}} \frac{1}{\left(1 + \frac{x^2}{v}\right)^{(v+1)/2}}, \quad -\infty < x < \infty, \quad v = 1, 2, \dots,$$

dengan v adalah derajat bebas, dan $v = n - 1$.



Gambar 4.4.2. Kurva Z dan t.

Sama halnya dengan distribusi Normal, pada distribusi student ini fungsinya cukup kompleks, tetapi tabel distribusi student sudah tersedia sehingga dengan sangat mudah kita dapat mencari nilai peluang. Berikut merupakan contoh sederhana cara membaca Tabel t.

Contoh 4.13

Distribusi t diketahui derajat bebas = 19 dan $t = 0,127$, maka luas daerah di sebelah kiri t adalah

Penyelesaian :

Dari tabel distribusi t dapat diperoleh sebagai berikut :

V	$t_{0,995}$	$t_{0,55}$
1	63,66	0,158
.....
19	2,86	0,127

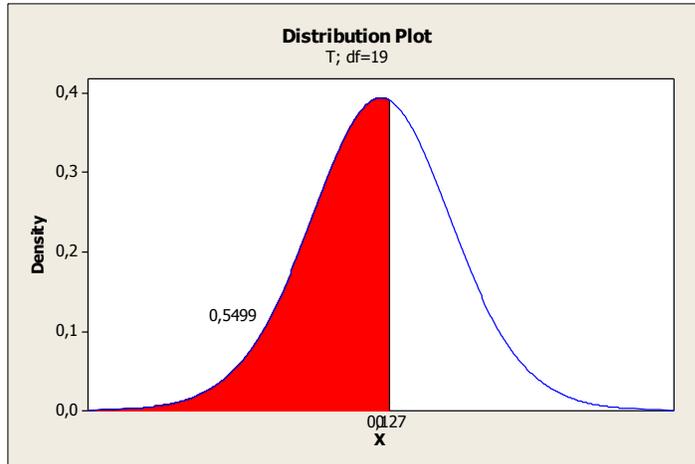
Luas daerah di sebelah kirinya adalah adalah 0,55.

Jika kita menggunakan Minitab, langkah-langkahnya adalah sebagai berikut.

- 1) Pilih Graph >> Probability Distribution Plot >> View Probability >> OK
- 2) Pilih distribution : t, degrees of freedom : 19

- 3) Klik shaded area : pilih left Tail dan klik X value. Masukkan nilai $t : 0,127$. Klik OK

Hasilnya adalah sebagai berikut.



Luas daerah yang diarsir adalah $0,5499 \approx 0,55$.

Contoh 4.14

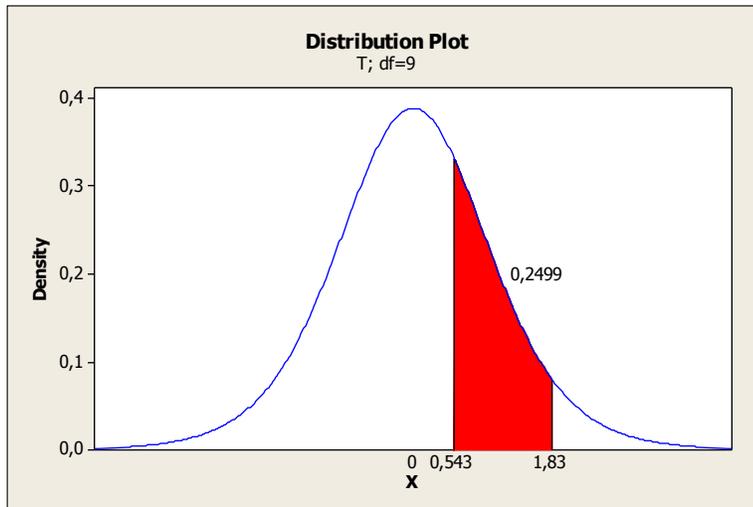
Misal banyaknya data 10, maka luas daerah yang dibatasi $1,830 > t > 0,543$ adalah

Penyelesaian :

Jika $v = n - 1 = 10 - 1 = 9$, maka untuk $t = 0,543$ dicapai saat $t_{0,70}$ dan $t = 1,830$ dicapai saat $t_{0,95}$. Sehingga luas antara $1,830 > t > 0,543$ adalah $0,95 - 0,70 = 0,250$.

- 1) Pilih Graph >> Probability Distribution Plot >> View Probability >> OK
- 2) Pilih distribution : t, degrees of freedom : 9
- 3) Klik shaded area : pilih Middle area dan klik X value. Masukkan nilai $t_1 : 0,543$ dan $t_2 : 1,830$. Klik OK

Hasilnya adalah sebagai berikut.



5. Distribusi F

Variabel random X dikatakan berdistribusi F jika fungsi peluangnya dapat didefinisikan sebagai:

$$f(x) = \begin{cases} \frac{\tau\left(\frac{v_1 + v_2}{2}\right) \left(\frac{v_1}{v_2}\right)^{\frac{v_1}{2}} x^{\frac{v_1}{2}-1}}{\tau\left(\frac{v_1}{2}\right) \tau\left(\frac{v_2}{2}\right) \left(1 + \frac{v_1 x}{v_2}\right)^{\frac{v_1 + v_2}{2}}} & 0 < x < \infty \\ 0 & x \text{ lainnya} \end{cases}$$

dengan $v_1 = n_1 - 1$, $v_2 = n_2 - 1$ bilangan bulat positif dimana v_1 disebut sebagai derajat kebebasan pembilang dan v_2 disebut sebagai derajat kebebasan penyebut.

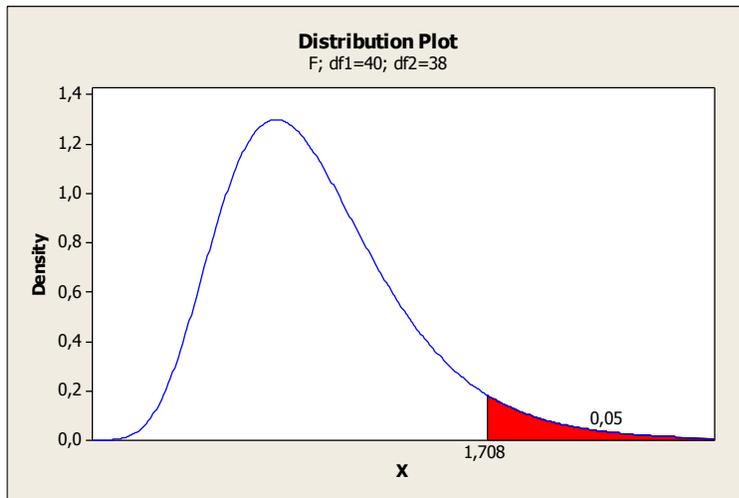
Contoh 4.15.

Nilai $F_{0,05(40;38)}$ adalah

Dari tabel distribusi F untuk dk pembilang 40, dk penyebut 38 dan $\alpha = 0,05$ diperoleh nilai $F = 1,71$.

Jika digunakan Minitab, caranya adalah:

- 1) Pilih Graph >> Probability Distribution Plot
- 2) View Probability >> OK
- 3) Pilih distribution : F, Numerator df : 40, denominator df : 38
- 4) Shaded area, pilih probability, pilih right tail, isikan probability : 0,05. Klik OK



Contoh 4.16.

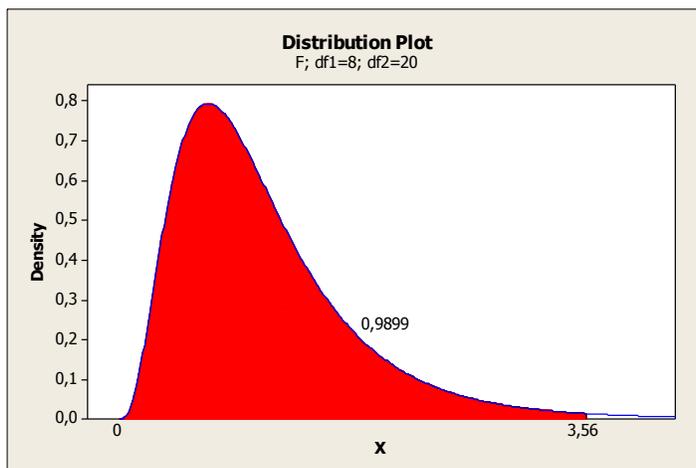
Jika derajat pembilang sama dengan 8 dan derajat penyebut = 20, nilai $F = 3,56$, maka luas daerah di sebelah kiri F adalah

Penyelesaian :

Diketahui dk pembilang = 8, dk penyebut = 20 dan $F = 3,56$. Hal ini akan dicapai saat $\alpha = 0,01$ yaitu 3,56 (angka yang bawah). Sehingga luas dari F ke kiri adalah 0,99.

Dengan menggunakan Minitab,

- 1) Pilih Graph >> Probability Distribution Plot
- 2) View Probability >> OK
- 3) Pilih distribution : F, Nemurator df : 8, denemurator df : 20
- 4) Shadded area, pilih X Value, pilih left tail, isikan X-value : 3,56. Klik OK



5

Sampling dan Distribusinya

5.1. SAMPLING

Ketika mempelajari pengertian statistika pada Bab 1, kita menemukan istilah populasi dan sampel. Populasi dalam definisi statistika adalah seluruh kumpulan objek atau orang yang akan diteliti. Sedangkan sampel berarti suatu bagian yang diambil dari suatu populasi. Populasi sangat banyak macamnya dan memiliki ukuran dari terbatas sampai tak terhingga. Apabila kita mengumpulkan data dari seluruh elemen dalam suatu populasi, kita akan memperoleh data yang sesungguhnya, yang biasa dikenal dengan istilah parameter, sedangkan jika kita melakukan penarikan sampel (mengumpulkan data sebagian elemen dari suatu populasi), kita akan memperoleh hasil yang berupa pendugaan yang biasanya disebut statistik. Pengambilan sampel (sampling) dari populasi merupakan proses utama dalam statistika induktif. Sampling dilakukan karena seorang peneliti tidak mungkin meneliti seluruh populasi, apalagi jika populasi tersebut relatif besar. Dengan pengambilan sampel, seorang peneliti dapat menghemat waktu dan biaya penelitian. Apakah sampel yang diambil cukup memberikan gambaran yang sama dengan populasinya? Dengan teknik pengambilan sampel yang baik, peneliti dapat memperoleh sampel yang representatif.

Penarikan sampel merupakan proses pemilihan sejumlah elemen dari populasi, sehingga dengan mempelajari sampel, suatu pemahaman karakteristik subjek sampel akan dimungkinkan untuk menggeneralisasi karakteristik elemen populasi. Teknik sampling merupakan teknik pengambilan sampel. Untuk menentukan sampel yang akan digunakan dalam penelitian, terdapat berbagai teknik sampling yang dipakai. Karakteristik populasi seperti μ (rata-rata populasi atau *mean*), σ (*population standard deviation*) dan σ^2 (*population variance*) disebut parameter populasi (*population parameter*). Kecendrungan-kecendrungan pusat (*central tendencies*), dispersi (*dispersions*), dan statistik-statistik lainnya dalam suatu sampel yang diteliti diperlukan sebagai pendekatan-pendekatan terhadap kecendrungan-kecendrungan pusat, dispersi dan parameter-parameter lainnya dalam populasi.

Istilah pengambilan sampel merujuk pada strategi-strategi atau metode-metode yang memungkinkan kita mengambil sebagian atau subbagian dari suatu kelompok yang lebih besar dan menggunakannya sebagai dasar untuk membuat kesimpulan. Hal ini sejalan dengan tujuan penelitian, yaitu ingin menggeneralisasikan populasi yang didasarkan pada pengamatan atau observasi terhadap sampel. Strategi pengambilan sampel bukan hanya membuat kemungkinan pengambilan data dari sejumlah kelompok yang lebih kecil, tetapi juga strategi-strategi yang memungkinkan kita memahami secara mendalam kelompok yang lebih kecil tersebut melalui pertanyaan yang mendalam. Apabila subkelompok demikian ini dipakai untuk membuat kesimpulan terhadap kelompok yang lebih besar atau lebih banyak, subkelompok ini harus benar-benar representatif terhadap kelompok yang lebih besar.

A. *PROBABILITY SAMPLING*

Probability Sampling adalah teknik sampling atau pengambilan sampel yang memberikan peluang yang sama bagi setiap unsur atau anggota populasi untuk dipilih menjadi sampel. Terdapat beberapa metoda penarikan sampel probabilitas, yaitu :

1. **Sampling acak sederhana (*simple random sampling*)**

Teknik sampling acak sederhana termasuk teknik sampling yang paling baik dan representatif. Kebaikan teknik ini bukan saja pada teori yang mendasarinya, tetapi juga pada bukti empiris yang dihasilkan. Dalam teknik ini setiap individu memiliki peluang atau kesempatan yang sama untuk dijadikan subjek penelitian. Alat untuk menentukan sampel secara acak ini bisa berupa kalkulator yang memiliki program untuk bilangan acaknya. Apabila populasi terbatas, peluang acak bisa diberikan secara individual. Sebaliknya, jika populasi sangat besar atau banyak dan berkelompok, peluang acak pertama-tama dilakukan dengan cara mengambil sejumlah kelompok yang ada, kemudian pengambilan sampel acak dilakukan pada kelompok tersebut. Cara tersebut akan membantu pekerjaan peneliti memahami karakteristik setiap subjek penelitian dari masing-masing kelompok yang ditetapkan sebagai sampel penelitian.

Contoh yang paling sederhana adalah cara undian. Misalnya, kita ingin menarik sampel sebanyak 35 secara acak dari suatu populasi yang berjumlah 100, yang terdiri dari sejumlah mahasiswa Jurusan Syari'ah, Tarbiyah dan Dakwah. Masing-masing jurusan diberi nomor sampai dengan 100, kemudian setiap nomor ditulis pada secarik kertas dan selanjutnya kertas-kertas bernomor tersebut dimasukkan ke dalam sebuah kotak. Setelah seluruh kertas dalam kotak tersebut dikocok dengan baik,

selanjutnya dipilih sebanyak 35 sampel yang prosedur penarikannya dilakukan 35 kali. Cara lain adalah dengan menggunakan tabel bilangan acak.

2. Sampling acak berstrata proporsional (*proportioned stratified random sampling*)

Metode sampling acak berstrata merupakan suatu prosedur dimana subsampel-subsampel acak sederhana ditarik dari setiap strata yang kurang lebih sama dalam beberapa karakteristik. Ada dua macam sampling acak berstrata, yaitu sampling acak berstrata proporsional dan disproporsional.

a. Sampling acak berstrata proporsional

Teknik ini digunakan bila populasi mempunyai anggota atau unsur yang tidak homogen dan berstrata secara proporsional. Untuk suatu organisasi yang mempunyai pegawai dengan latar belakang pendidikan berstrata, populasi pegawai itu berstrata.

Misalnya, jumlah pegawai yang lulus S1 = 45, S2 = 30, STM = 800, SMP = 900, SMK = 400 dan SD = 300. Jumlah sampel yang harus diambil meliputi strata pendidikan tersebut. Contoh lain, populasi = 1000 yang terdiri dari 700 orang wanita dan 300 orang pria. Sampel yang diperlukan = 100. Secara proporsional, sampel yang dapat ditarik adalah wanita = $\frac{700}{1000} \times 100 = 70$ dan pria = $\frac{300}{1000} \times 100 = 30$.

b. Sampling acak berstrata disproporsional

Teknik ini digunakan untuk menentukan jumlah sampel bila populasi berstrata, tetapi kurang proporsional. Berdasarkan kasus di atas, secara disproporsional dapat ditarik sampel, misalnya untuk wanita 60% = 60 dan pria 40% = 40. Prinsip sampling disproporsional adalah :

1. Semakin besar suatu strata, semakin besar sampel
2. Semakin tinggi variabilitas di dalam suatu sampel, semakin besar sampel

Misalnya, pegawai dari unit kerja tertentu mempunyai 3 orang lulusan S3, 4 orang lulusan S2, 90 orang lulusan S1, 800 orang lulusan SMA dan 700 orang lulusan SMP. Dalam hal ini, 3 orang lulusan S3 dan 4 orang lulusan S2 diambil semuanya sebagai sampel, karena dua kelompok ini terlalu kecil bila dibandingkan dengan kelompok S1, SMA dan SMP.

3. Metode sampling berkelompok (*cluster sampling*)

Metode sampling berkelompok merupakan prosedur penarikan sampel probabilitas yang memilih subpopulasi yang disebut klaster, kemudian setiap elemen di dalam kelompok dipilih sebagai anggota sampel. Teknik sampling ini digunakan untuk menentukan sampel bila objek yang akan

diteliti sumber datanya sangat luas, misalnya penduduk dari suatu Negara, provinsi atau kabupaten. Untuk menentukan penduduk mana yang akan dijadikan sumber data, samplingnya dilakukan berdasarkan daerah populasi yang telah ditetapkan.

Pada sampling klaster, kriteria yang digunakan dalam pembentukan kelompok (*cluster*) bertolak belakang dengan apa yang digunakan dalam sampling berstrata. Dalam sampling berstrata, kelompok-kelompok yang dibentuk harus bersifat homogen. Contoh, strata toko berdasarkan ukuran (besar, menengah, kecil), strata usia 0-19, 20-39, dsb. Dalam teknik sampling ini, sebelum memilih klaster, populasi harus dibagi ke dalam kelompok-kelompok yang bersifat *mutually exclusive*. Selanjutnya, baru kita memilih kelompok-kelompok tersebut secara acak sebagai sampel. Sebagai contoh, kita memiliki suatu populasi yang terdiri atas 20 elemen yang dibagi dalam empat kelompok berukuran sama

Kelompok	Jumlah elemen populasi
Kel 1	1, 2, 3, 4, 5
Kel 2	6, 7, 8, 9, 10
Kel 3	11, 12, 13, 14, 15
Kel 4	16, 17, 18, 19, 20

Selanjutnya dari pengelompokan populasi tersebut, dipilih secara acak kelompok-kelompok yang akan dijadikan sampel. Kemudian, dari kelompok yang terpilih, anggota-anggota kelompok tersebut dipilih secara acak untuk dijadikan sampel. Misalnya, Indonesia terdiri dari 30 provinsi dan sampelnya akan menggunakan 15 provinsi. Pengambilan 15 provinsi tersebut dilakukan secara acak. Tetapi karena provinsi di Indonesia adalah berstrata (tidak sama), sehingga perlu menggunakan sampling acak berstrata. Ada provinsi di Indonesia yang penduduknya padat, ada yang tidak, ada yang mempunyai hutan banyak, ada yang tidak, ada yang kaya bahan tambang, dan ada yang tidak. Karakteristik semacam ini perlu diperhatikan sehingga pengambilan sampel menurut strata populasi dapat ditetapkan.

B. NONPROBABILITY SAMPLING

Sampling nonprobabilitas merupakan suatu prosedur penarikan sampel yang bersifat subjektif. Dalam hal ini, probabilitas pemilihan elemen-elemen populasi tidak dapat ditentukan. Hal ini disebabkan setiap elemen populasi tidak memiliki peluang yang sama untuk dipilih sebagai sampel. Sampling nonprobabilitas bisa menghemat waktu dan biaya karena tidak memerlukan kerangka sampling, namun hasilnya bisa mengandung bias dan ketidakpastian. Dalam kasus-kasus tertentu, kemungkinan sampling nonprobabilitas merupakan satu-satunya cara yang dapat dilakukan. Misalnya, dalam suatu penelitian terhadap para pengunjung mal atau pusat-pusat perbelanjaan.

1. Sampling Sistematis

Merupakan teknik sampling berdasarkan urutan anggota populasi yang telah diberi nomor urut. Dalam penarikan sampel sistematis, populasi dibagi dengan ukuran sampel yang diperlukan (n) dan sampel diperoleh dengan cara mengambil setiap subjek ke- n .

Contoh, populasi 100, ukuran sampel 10. Bagi populasi dengan ukuran sampel, $100/10 = 10$. Selanjutnya, pilih nomor antara 1 dan 10, misalnya 5. Kemudian pilih yang ke 10, setelah itu hingga 10 dipilih 5, 15, 25, 35, 45, 55, 65, 75, 85, 95.

2. Sampling Wilayah

Merupakan bentuk sampling kluster dalam suatu wilayah. Suatu kota yang menunjukkan wilayah-wilayah dapat dijadikan dasar pembentukan sampel wilayah dan selanjutnya bisa diperoleh data dari penduduk yang berada dalam wilayah-wilayah tertentu.

Contoh, sebuah stasiun radio melakukan survei profil dan perilaku pendengar radio. Penelitian dapat menggunakan peta kota, kemudian mengidentifikasi wilayah-wilayah kota tersebut hingga ke kecamatan, kelurahan, RW dan RT yang terpilih secara acak. Hal ini dilakukan jika daftar yang lengkap dan akurat sulit diperoleh atau membutuhkan biaya yang besar untuk memperolehnya. Selanjutnya sampel dipilih secara acak dari setiap kluster tersebut.

3. Sampling Kemudahan

Untuk mendapatkan informasi dengan cepat, mudah dan murah, sering digunakan sampling kemudahan. Prosedurnya adalah dengan langsung menghubungi unit-unit sampling yang mudah dijumpai, seperti mahasiswa di suatu kelas, jemaah tempat-tempat ibadah, rekan-rekan, para tetangga, dll. Sering kali teknik sampling ini dilakukan untuk menguji kuesioner atau digunakan dalam penelitian eksplorasi.

Disamping kemudahan dan kelebihan dari teknik sampling ini juga memiliki keterbatasan. Hal ini mengingat pemilihan unit sampling dengan metode ini dapat dilakukan dengan mengambil siapa saja yang dapat ditemui oleh peneliti sehingga bila dalam prosesnya tidak dilakukan lebih detail, hasil yang diperoleh dapat menimbulkan bias dalam pengambilan keputusannya. Disamping itu, teknik ini tidak tepat digunakan bila populasinya dapat didefinisikan, karena kondisi ini dimungkinkan untuk menyediakan kerangka sampel, sehingga dalam tindak lanjutnya lebih disarankan untuk menggunakan sampling probabilitas. Teknik ini juga tidak cocok digunakan untuk penelitian

yang bersifat deskriptif dan kasual. Sebagai tambahan, metode ini juga menuntut kehati-hatian dalam penerjemahan hasil penelitian.

4. Sampling Pertimbangan

Merupakan bentuk penarikan sampel nonprobabilitas yang didasarkan pada kriteria-kriteria tertentu. Misalnya dalam suatu penelitian tentang masalah sumber daya manusia, peneliti mungkin hanya ingin memperoleh informasi dari pegawai-pegawai yang memiliki karakteristik tertentu. Dalam kaitannya dengan sampling pertimbangan dikenal juga sampling ahli (*expert sampling*) dan sampling bertujuan (*purposive sampling*). Kendala yang dihadapi dalam penggunaan sampling pertimbangan ini adalah tuntutan adanya kejelian dari peneliti dalam mendefinisikan populasi dan membuat pertimbangannya. Pertimbangan atau *judgement* harus masuk akal dan relevan dengan maksud penelitian.

5. Sampling Kuota

Merupakan bentuk lain dari sampling pertimbangan. Prinsipnya adalah karakteristik-karakteristik tertentu yang relevan yang menjelaskan dimensi-dimensi populasi. Dalam hal ini, distribusi populasi harus diketahui. Misalnya kita ingin menarik sampel sebanyak 1000 orang penduduk kota Bandung. Jika diketahui penyebaran penduduk secara geografis, sampelnya dapat ditarik persentase distribusi yang sama.

Kelebihan sampling kuota adalah rendahnya biaya penelitian, keleluasaan peneliti untuk menentukan elemen-elemen setiap kuotanya, bahkan pada kondisi tertentu, hasil penelitian dapat menyamai hasil penelitian yang dilakukan dengan teknik sampling probabilitas.

6. Sampling Bola Salju

Merupakan suatu metode penarikan sampel yang dalam hal ini responden yang berhasil diperoleh diminta untuk menunjukkan responden lainnya secara berantai. Teknik ini sangat tepat digunakan bila populasinya sangat spesifik, yaitu cara pengambilan sampel dilakukan dengan cara berantai, mulai dari ukuran sampel yang kecil, semakin lama semakin besar seperti halnya bola salju yang menggelinding menuruni lereng gunung atau bukit. Dalam pelaksanaannya, pertama-tama dilakukan wawancara terhadap suatu kelompok atau seorang responden yang relevan, dan untuk selanjutnya yang bersangkutan diminta untuk menyebutkan atau menunjuk calon responden yang berikutnya yang memiliki spesifikasi atau spesialisasi yang sama.

Hasil penelitian dari teknik ini pada umumnya menghasilkan bias yang relatif kecil. Kendala utama pada metode sampling bola salju ini terletak pada besarnya waktu dan biaya yang dibutuhkan untuk memperoleh informasi, karena populasinya spesifik serta tersebar.

C. **PENENTUAN UKURAN SAMPEL**

Dalam suatu penelitian, penentuan ukuran sampel (*sample size*) merupakan hal yang sangat penting untuk diperhatikan, terutama jika penelitian dilakukan dengan cara survei dan bertujuan untuk memperkirakan nilai karakteristik dari populasi yang diteliti. Ukuran sampel yang terlalu besar akan kurang efisien dari aspek sumber daya penelitian (waktu, tenaga, biaya), sebaliknya ukuran sampel yang terlalu kecil akan kurang efisien dari aspek ketelitian pengukuran dan keakuratan hasil penelitian. Oleh karena itu, penentuan besar sampel tidak akan pernah terlepas dari pertimbangan dua aspek tersebut. Penentuan ukuran sampel penelitian merupakan bagian dari penyusunan metode penelitian, terutama berkaitan dengan metode pengumpulan data. Namun, sebelum hal ini dilakukan, kita harus merumuskan masalah dan tujuan penelitian secara jelas serta mendefinisikan populasi yang akan menjadi cakupan dari penelitian. Dalam menentukan besar sampel, ada banyak alternatif rumus penentuan besar sampel yang bisa dipakai.

1. **Tabel Krejcie dan Morgan**

Krejcie dan Morgan membuat daftar yang bisa dipakai untuk menentukan jumlah sampel sebagai berikut (Lihat Tabel)

Populasi (N)	Sampel (n)	Populasi (N)	Sampel (n)	Populasi (N)	Sampel (n)
10	10	220	140	1200	291
15	14	230	144	1300	297
20	19	240	148	1400	302
25	24	250	152	1500	306
30	28	260	155	1600	310
35	32	270	159	1700	313
40	36	280	162	1800	317
45	40	290	165	1900	320
50	44	300	169	2000	322
55	48	320	175	2200	327
60	52	340	181	2400	331
65	56	360	186	2600	335
70	59	380	191	2800	338
75	63	400	196	3000	341
80	66	420	201	3500	346

85	70	440	205	4000	351
90	73	460	210	4500	354
95	76	480	214	5000	357
100	80	500	217	6000	361
110	86	550	226	7000	364
120	92	600	234	8000	367
130	97	650	242	9000	368
140	103	700	248	10000	370
150	108	750	254	15000	375
160	113	800	260	20000	377
170	118	850	265	30000	379
180	123	900	269	40000	380
190	127	950	274	50000	381
200	132	1000	278	75000	382
210	136	1100	285	1000000	384

Sumber : Uma Sekaran. 2006. Metodologi Penelitian Untuk Bisnis. Jakarta : Salemba Empat.

2. Tabel Isaac dan Michael

Tabel penentuan jumlah sampel dari Isaac dan Michael memberikan kemudahan penentuan jumlah sampel berdasarkan tingkat kesalahan 1%, 5% dan 10%. Dengan tabel ini, peneliti dapat secara langsung menentukan besaran sampel berdasarkan jumlah populasi dan tingkat kesalahan yang dikehendaki.

N	Siginfikasi			N	Siginfikasi		
	1%	5%	10%		1%	5%	10%
10	10	10	10	280	197	155	138
15	15	14	14	290	202	158	140
20	19	19	19	300	207	161	143
25	24	23	23	320	216	167	147
30	29	28	28	340	225	172	151
35	33	32	32	360	234	177	155
40	38	36	36	380	242	182	158
45	42	40	39	400	250	186	162
50	47	44	42	420	257	191	165
55	51	48	46	440	265	195	168
60	55	51	49	460	272	198	171
65	59	55	53	480	279	202	173
70	63	58	56	500	285	205	176
75	67	62	59	550	301	213	182
80	71	65	62	600	315	221	187

85	75	68	65	650	329	227	191
90	79	72	68	700	341	233	195
95	83	75	71	750	352	238	199
100	87	78	73	800	363	243	202
110	94	84	78	850	373	247	205
120	102	89	83	900	382	251	208
130	109	95	88	950	391	255	211
140	116	100	92	1000	399	258	213
150	122	105	97	1100	414	265	217
160	129	110	101	1200	427	270	221
170	135	114	105	1300	440	275	224
180	142	119	108	1400	450	279	227
190	148	123	112	1500	460	283	229
200	154	127	115	1600	469	286	232
210	160	131	118	1700	477	289	234
220	165	135	122	1800	485	292	235
230	171	139	125	1900	492	294	237
240	176	142	127	2000	498	297	238
250	182	146	130	2200	510	301	241
260	187	149	133	2400	520	304	243
270	192	152	135	2600	529	307	245

Sumber : Uma Sekaran. 2006. Metodologi Penelitian Untuk Bisnis. Jakarta : Salemba Empat.

3. Aturan Slovin

Berbeda dengan dua tabel di atas, kita juga bisa menentukan ukuran atau jumlah sampel dengan formula yang dikembangkan oleh **Slovin** dengan margin error yang diperkenankan berkisar antara 5 - 10%, yakni dengan rumus :

$$n = \frac{N}{1 + Nd^2}$$

keterangan :

n : ukuran sampel yang dibutuhkan

N : jumlah populasi

d : batas kesalahan yang diperkenankan (5% atau 10%)

5.2. DISTRIBUSI SAMPLING

Bidang statistic inferensi sesungguhnya berhubungan dengan generalisasi dari peramalan. Sebagai misal, kita mungkin menyatakan, berdasarkan pendapat beberapa orang yang kita wawancarai, bahwa dalam pemilihan mendatang 60% pemilih menyukai calon tertentu. Dalam hal ini kita

menghitung sebuah statistik dari suatu sampel yang terambil dari suatu populasi, dan berdasarkan statistik kita membuat pernyataan mengenai nilai parameter populasi. Karena banyak sekali sampel yang terambil dari populasi yang sama, maka kita dapat membayangkan bahwa setiap statistik apakah itu \bar{X} atau s^2 akan sedikit bervariasi dari sampel satu ke sampel lainnya. Jadi suatu statistik merupakan variabel acak yang hanya bergantung pada sampel yang teramati.

Distribusi peluang suatu statistik disebut distribusi sampling. Distribusi peluang bagi \bar{X} disebut distribusi sampling bagi mean μ . Sudah menjadi kelaziman untuk menyebut simpangan baku bagi distribusi penarikan sampel suatu statistik sebagai sisaan/galat baku statistik tersebut. Distribusi penarikan sampel suatu statistik akan bergantung pada ukuran populasi, ukuran sampel, dan metode pengambilan sampelnya.

Ada empat jenis distribusi sampling yang akan dibahas, diantaranya: distribusi sampling rata-rata, beda 2 rata-rata, 1 proporsi dan beda 2 proporsi. Distribusi sampling rata-rata adalah distribusi probabilitas yang berisi daftar semua rata-rata sampel yang mungkin jika kita mengambil sejumlah sampel dari populasi, beserta dengan probabilitas setiap rata-rata sampel. Jika kita memiliki populasi yang terdiri dari tiga orang, yaitu A, B, C, lalu mengambil sampel sebanyak 2 orang, ada 3 kombinasi sampel yang mungkin terpilih, yaitu AB, AC dan BC. Jika kita mengambil 10 sampel dari populasi sebanyak 60, maka kita mempunyai kombinasi 60 dari 10. Kita tahu bahwa ukuran populasi (N) biasanya sangat besar dan ukuran sampel (n) relatif lebih kecil. Oleh karenanya, jika kita melakukan pengambilan sampel n dan N populasi, kita akan memiliki kombinasi N terhadap n . Setiap kombinasi sampel memiliki ukuran (statistik sampel), misalnya rata-rata sampel. Maka jika kita mengambil n sampel dari N populasi, kita akan memiliki rata-rata sampel yang cukup banyak.

Telah disebutkan bahwa rata-rata sampel merupakan variabel acak sehingga mempunyai distribusi sendiri. Menurut dalil batas memusat, jika populasi terdistribusi secara normal, rata-rata sampel yang banyak tersebut juga akan terdistribusi secara normal. Distribusi rata-rata sampel adalah distribusi sampling atau tepatnya distribusi sampling rata-rata. Distribusi sampling rata-rata merupakan distribusi normal, yang berbentuk lonceng, simetris dan memiliki rata-rata dan deviasi standar.

A. TEOREMA LIMIT PUSAT

Teorema limit pusat adalah suatu dalil yang sangat penting peranannya dalam distribusi sampling, yang menyatakan bahwa untuk suatu populasi dengan rata-rata μ dan variansi σ^2 , distribusi sampling rata-rata dari semua kemungkinan sampel berukuran n yang diambil dari populasi akan terdistribusi secara normal dengan rata-rata μ_x dan deviasi standar σ_x , dimana nilai μ_x sama dengan rata-rata populasi (μ) dan σ_x sama dengan

deviasi standar populasi dibagi akar n atau $\frac{\sigma}{\sqrt{n}}$, dengan asumsi bahwa ukuran sampel cukup besar.

Bila \bar{x} rata-rata (mean) sampel acak berukuran n yang diambil dari populasi dengan mean μ dan variansi σ^2 , maka bentuk limit dari distribusi

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

bila $n \rightarrow \infty$, ialah distribusi Normal Baku $N(0, 1)$.

Dengan kata lain, dalam pemilihan sampel acak sederhana dengan ukuran n dari suatu populasi yang berasal dari distribusi apapun, distribusi dari rata-rata sampel dapat didekati dengan distribusi probabilitas normal untuk ukuran sampel yang besar.

Jika \bar{x} dan s^2 masing-masing adalah mean dan varians dari sampel acak berukuran n yang diambil dari populasi normal dengan mean μ dan varians σ^2 , maka

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

merupakan nilai variabel random T yang berdistribusi t dengan derajat bebas $v = n - 1$.

Jika S^2 variansi sampel random berukuran n yang diambil dari populasi normal dengan varians σ^2 , maka statistik

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

berdistribusi Chi Kuadrat χ^2 dengan derajat bebas $db = n - 1$.

Jika S_1^2 dan S_2^2 adalah varians dua sampel acak bebas yang berukuran n_1 dan n_2 yang ditarik dari populasi normal dengan varians σ_1^2 dan σ_2^2 , maka :

$$f = \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}$$

berdistribusi $F_{(v_1, v_2)}$ dengan derajat bebas $v_1 = n_1 - 1$ dan $v_2 = n_2 - 1$.

B. DISTRIBUSI SAMPLING RATA-RATA

Pada hakikatnya, distribusi sampling rata-rata adalah distribusi probabilitas rata-rata sejumlah C sampel, N adalah ukuran populasi dan n adalah ukuran sampel yang diambil dari populasi. Distribusi ini memiliki rata-rata μ_x dan deviasi standar σ_x . Sementara itu menurut Teorema Limit

Pusat, $\mu_x = \mu$ dan $\sigma_x = \frac{\sigma}{\sqrt{n}}$. Jika disusun ke dalam suatu distribusi, rata-

rata tersebut sama seperti nilai-nilai dalam distribusi skor mentah. Distribusi semacam ini disebut distribusi sampel rata-rata (*sample distribution of means*). Selanjutnya dapat dihitung rata-rata dari distribusi sampel rata-rata (*means of sample distribution*). Rata-rata distribusi ini akan sama dengan rata-rata populasi.

Contoh 5.1

Untuk keperluan monitoring dan evaluasi jurusan, pihak Rektorat menyarankan Jurusan Tarbiyah untuk menghitung rata-rata dan simpangan baku IPK semua mahasiswa yang aktif. Jurusan tarbiyah sendiri memiliki 600 orang mahasiswa aktif, dengan rata-rata IPK 3,12 dengan simpangan baku 0,62. Seorang dosen melakukan cross cek data yang dimiliki jurusan, dengan mengambil sampel sebanyak 100 mahasiswa, berapa probabilitas jika :

- Rata-rata sampel akan terletak antara 3,00 dan 3,25
- Rata-rata sampel akan lebih kecil dari 2,95
- Rata-rata sampel akan lebih besar dari 3,30

Jawab :

Dosen mengambil sampel 100 dari 600 populasi, akan terdapat $\binom{600}{100}$

kombinasi sampel yang mungkin. Dengan kata lain, akan terdapat

sebanyak $\binom{600}{100}$ rata-rata sampel. Jumlah rata-rata sampel tersebut cukup

banyak sehingga distribusinya normal (hal ini konsisten dengan Teorema Limit Pusat). Distribusi sampling rata-rata ini memiliki rata-rata dan simpangan baku sebagai berikut :

$$\mu_x = \mu = 3,12$$

dan

$$\sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{0,62}{\sqrt{100}} = \frac{0,62}{10} = 0,062$$

- a. Selanjutnya, untuk menyelesaikan soal tersebut, kita akan menerapkan konsep menghitung luas daerah kurva normal

$$Z = \frac{\bar{X} - \mu_x}{\sigma_x} = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}$$

Untuk $\bar{X} = 3,00$

$$Z_1 = \frac{\sqrt{n}(\bar{X}_1 - \mu)}{\sigma} = \frac{\sqrt{100}(3,00 - 3,12)}{0,62} = \frac{10(-0,12)}{0,62} = -1,94$$

Untuk $\bar{X} = 3,25$

$$Z_2 = \frac{\sqrt{n}(\bar{X}_2 - \mu)}{\sigma} = \frac{\sqrt{100}(3,25 - 3,12)}{0,62} = \frac{10(0,13)}{0,62} = 2,10$$

Selanjutnya,

$$P(3,00 \leq \bar{X} \leq 3,25) = P(-1,94 \leq Z \leq 2,10) = 0,4744 + 0,4821 = 0,96$$

- b. $P(\bar{X} \leq 2,95)$

$$Z = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} = \frac{\sqrt{100}(2,95 - 3,12)}{0,62} = \frac{10(-0,17)}{0,62} = -2,74$$

Selanjutnya,

$$P(\bar{X} \leq 2,95) = P(Z \leq -2,74) = 0,5 - 0,4969 = 0,0031$$

- c. $P(\bar{X} \geq 3,30)$

$$Z = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} = \frac{\sqrt{100}(3,30 - 3,12)}{0,62} = \frac{10(0,18)}{0,62} = 2,90$$

Selanjutnya,

$$P(\bar{X} \geq 3,30) = P(Z \geq 2,90) = 0,5 - 0,4981 = 0,0019$$

C. FAKTOR KOREKSI UNTUK POPULASI TERBATAS

Jika populasi sangat besar, kita asumsikan populasi tersebut tak terbatas (*infinite*). Bagaimana jika populasi tidak tak terbatas atau tidak sangat besar? Dalam kasus ini, kita harus melakukan beberapa penyesuaian/koreksi terhadap deviasi standar dari distribusi sampling dengan cara mengalikan $\frac{\sigma}{\sqrt{n}}$ dengan suatu faktor koreksi sebesar

$\sqrt{\left(\frac{N-n}{N-1}\right)}$, sehingga dapat dirumuskan sebagai :

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

Mengapa faktor koreksi ini perlu dan apa efeknya? Jika sampel adalah suatu persentase yang cukup besar dari populasinya, kita mengharapkan ukurannya akan lebih tepat daripada ukuran suatu sampel yang lebih kecil. Perhatikan efek dari faktor koreksi, misalnya kita mengambil sampel dengan ukuran 100 dari populasi berukuran 1000, besar faktor koreksinya adalah 0,9492. Jika dikalikan dengan faktor koreksi tersebut, deviasi standar distribusi sampling rata-rata akan berkurang sebesar $1 - 94,92\% = 5\%$. Semakin besar ukuran sampel, semakin besar pengurangan kesalahan baku tersebut, demikian pula sebaliknya.

Ukuran sampel (n)	Bagian dari Populasinya (n/N)	Faktor Koreksi
10	1%	99.55%
25	2.5%	98.79%
50	5%	97.52%
100	10%	94.92%
200	20%	89.49%
500	50%	70.75%

Jika $\frac{n}{N}$ lebih kecil dari 5%, faktor koreksi mendekati 1 sehingga muncul

aturan jika $\frac{n}{N}$ lebih kecil dari 55, faktor koreksi tidak perlu digunakan, walaupun digunakan, pengaruhnya tidak akan banyak, karena nilainya mendekati 1.

Contoh 5.2

Bila sampel acak dengan $n=10$ dipilih dari populasi sebesar 40 dengan rata-rata 5.5 dan deviasi standar 2.9155, berapa rata-rata dan deviasi distribusi sampling rata-rata?

Jawab

Menurut Teorema Limit Pusat dan penyesuaian terhadap koreksi,

$$\mu_x = \mu = 5,5$$

dan

$$\sigma_x = \frac{\sigma}{\sqrt{n}} \sqrt{\left(\frac{N-n}{N-1}\right)} = \frac{2,9155}{\sqrt{10}} \sqrt{\left(\frac{40-10}{40-1}\right)} = 0,2773$$

D. DISTRIBUSI SAMPLING BEDA RATA-RATA

Misalnya, kita sedang meneliti 2 populasi, yaitu populasi 1 dan populasi 2 yang masing-masing memiliki ukuran N_1 dan N_2 , kemudian dari masing-masing populasi tersebut kita mengambil sampel dengan ukuran n_1 dan n_2 . Pengambilan sampel n_1 dari populasi N_1 menghasilkan kombinasi sampel sebanyak $\binom{N_1}{n_1}$ dan pengambilan sampel n_2 dari

populasi N_2 menghasilkan kombinasi sampel sebanyak $\binom{N_2}{n_2}$. Dengan

kata lain, kita memiliki rata-rata sampel dari populasi 1 dan rata-rata sampel dari populasi 2 yang cukup banyak. Jika \bar{X} adalah selisih \bar{X}_1 dan \bar{X}_2 , yaitu $\bar{X} = \bar{X}_1 - \bar{X}_2$, kita akan memiliki \bar{X} yang banyak sekali yang membentuk suatu distribusi normal yang disebut distribusi sampling beda rata-rata. Distribusi sampling ini memiliki rata-rata $\mu_{X_1-X_2}$ dan deviasi standar atau kesalahan baku $\sigma_{X_1-X_2}$. Keduanya dirumuskan sebagai berikut:

$$\mu_{X_1-X_2} = \mu_1 - \mu_2$$

dan

$$\sigma_{X_1-X_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Contoh 5.3

Dua buah populasi masing-masing diketahui nilai rata-ratanya 143 dan 132 dengan simpangan baku masing-masing 3,2 dan 2,9. Dari kedua populasi tersebut diambil secara acak sampel masing-masing berukuran 120 dan 140. Tentukan kekeliruan baku selisih rata-rata ($\sigma_{\bar{X}-\bar{Y}}$)

Jawab

Diketahui :

$$n_1 = 120$$

$$n_2 = 140$$

$$\mu_1 = 143$$

$$\mu_2 = 132$$

$$\sigma_1 = 3,2$$

$$\sigma_2 = 2,9$$

$$(\sigma_{\bar{X}-\bar{Y}}) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$(\sigma_{\bar{X}-\bar{Y}}) = \sqrt{\frac{3,2^2}{120} + \frac{2,9^2}{140}} = 0,3813$$

E. DISTRIBUSI SAMPLING PROPORSI

Proporsi populasi, dinotasikan dengan π atau P, dapat dicari dengan rumus $\pi = \frac{n(X)}{n(S)}$ dengan $n(X)$ adalah banyaknya individu yang berkarakteristik X dan $n(S)$ adalah total seluruh individu yang diamati. Sebagai contoh, total mahasiswa adalah 100 orang, jika 30 mahasiswa diantaranya merokok, proporsi mahasiswa yang merokok adalah $\frac{30}{100}$ atau 30%.

Proses pembentukan distribusi sampling proporsi sama dengan distribusi sampling rata-rata yang telah dibahas sebelumnya.

Pada dasarnya terdapat kemiripan antara distribusi sampling dari \bar{X} dan proporsi. Pada distribusi \bar{X} , nilai X_1, X_2, \dots, X_n berasal dari distribusi Normal. Distribusi proporsi berasal dari nilai-nilai X yang berdistribusi Bernouli, yaitu bernilai 0 dan 1.

Andaikan X adalah banyaknya yang berhasil, dan $\hat{p} = \frac{X}{n}$, maka distribusi bagi $\hat{p} = \frac{X}{n}$ dapat ditentukan dengan cara sebagai berikut:

Mean :

$$\mu_{\hat{p}} = E(\hat{P}) = E\left(\frac{X}{n}\right) = \frac{np}{n} = p$$

Variansi :

$$\sigma_{\hat{p}}^2 = \sigma_{X/n}^2 = \frac{\sigma_X^2}{n^2} = \frac{pq}{n}, \text{ dimana } q = 1 - p$$

Contoh 5.4

Di suatu kota, diketahui 6% dari siswa SMA terkena penyakit flu di setiap akhir tahun ajaran. Sampel acak berukuran 250 diteliti. Nilai kemungkinan bahwa dari 250 siswa akan terdapat paling sedikit 25 siswa terkena flu adalah

Penyelesaian:

Diketahui :

$$\pi = 0,06$$

$$n = 250$$

$$x = 25$$

$$\begin{aligned} P(X \leq 25) &= P\left(\frac{X}{n} \leq \frac{25}{250}\right) = P\left(\frac{X/n}{\sqrt{\frac{\pi(1-\pi)}{n}}} \leq \frac{0,1-0,06}{\sqrt{\frac{0,06 \times 0,94}{250}}}\right) \\ &= P\left(Z \leq \frac{\sqrt{250} \times 0,04}{\sqrt{0,237}}\right) \\ &= P(Z \leq 2,76) = 0,4962 = 49,62\% \end{aligned}$$

F. DISTRIBUSI SAMPLING BEDA PROPORSI

Pada dasarnya, proses terbentuknya distribusi sampling beda proporsi ini sama dengan pembentukan distribusi sampling beda rata-rata. Misalnya terdapat 2 populasi binomial (populasi yang dibedakan menjadi 2 kelompok, seperti merokok dan tidak merokok, setuju dan tidak setuju, dsb) dengan ukuran N_1 dan N_2 . Dari kedua populasi tersebut, masing-masing kita ambil sampel n_1 dan n_2 . Kita akan memiliki kemungkinan kombinasi sampel sebesar $\binom{N_1}{n_1}$ dan $\binom{N_2}{n_2}$. Dengan kata lain kita

memiliki proporsi sampel 1 (p_1) sebanyak $\binom{N_1}{n_1}$ dan proporsi sampel 2 (p_2) sebanyak $\binom{N_2}{n_2}$. Jika p adalah selisih p_1 dan p_2 , kita akan mendapatkan $p_1 - p_2$ banyak yang membentuk distribusi normal dengan rata-rata $\mu_{p_1-p_2}$ dan deviasi standar $\sigma_{p_1-p_2}$.

Misalkan dari populasi 1 diperoleh estimasi bagi proporsi, $\hat{p}_1 = \frac{X_1}{n_1}$, dimana X_1 adalah banyaknya sukses pada sampel 1. Estimasi bagi proporsi poluasi 2, $\hat{p}_2 = \frac{X_2}{n_2}$, dimana . Distribusi sampling bagi selisih dua proporsi, dimana X_2 adalah banyaknya sukses pada sampel 2.

Distribusi sampling bagi selisih dua proporsi, $\hat{p}_1 - \hat{p}_2$ adalah :

$$\mu_{\hat{p}_1 - \hat{p}_2} = E(\hat{p}_1 - \hat{p}_2)$$

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

Jadi $\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$ (menurut sifat-sifat dari ekspektasi)

Variansi bagi $\hat{p}_1 - \hat{p}_2$:

$$\sigma_{\hat{p}_1 - \hat{p}_2}^2 = \text{Var}(\hat{p}_1 - \hat{p}_2) = \text{Var}(\hat{p}_1) + \text{Var}(\hat{p}_2)$$

$$\text{Jadi } \sigma_{\hat{p}_1 - \hat{p}_2}^2 = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$$

Sehingga,

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

Contoh 5.5

Sebuah populasi berisikan 70% kategori A akan diambil dua buah sampel berukuran sama besar. Dikehendaki bahwa kekeliruan baku selisih kedua sampel tidak lebih dari 0,04, tentukan ukuran sampel terkecilnya!

Penyelesaian :

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

Diketahui dari soal bahwa $\sigma_{p_1-p_2} = 0,04$ $p_1 = p_2 = 0,7$ dan $n_1 = n_2 = n$, sehingga

$$0,04 = \sqrt{\frac{0,7(1-0,7)}{n} + \frac{0,7(1-0,7)}{n}} = \sqrt{\frac{0,21}{n} + \frac{0,21}{n}} = \sqrt{\frac{0,42}{n}}$$

$$0,0016 = \frac{0,42}{n}$$

$$n = \frac{0,42}{0,0016} = 262,5$$

Jadi ukuran sampel terkecilnya adalah 263.

6

Pengujian Hipotesis

Sering permasalahan yang dihadapi oleh peneliti tidak banyak menyangkut estimasi parameter suatu populasi, tetapi menyangkut cara pengambilan keputusan berdasarkan data mengenai suatu sistem ilmu. Dalam hal ini peneliti membuat suatu dugaan mengenai suatu sistem. Dalam setiap kasus harus melibatkan penggunaan data percobaan dan pengambilan keputusan berdasarkan data tersebut. Resminya dalam setiap kasus dugaan dapat dirumuskan dalam bentuk hipotesis statistik.

6.1. KOMPONEN-KOMPONEN DALAM HIPOTESIS STATISTIK

A. HIPOTESIS STATISTIK

Hipotesis Statistik didefinisikan sebagai suatu anggapan atau pernyataan yang mungkin benar/salah mengenai satu atau lebih populasi. Benar salahnya hipotesis tidak akan pernah diketahui dengan pasti, kecuali bila kita memeriksa seluruh populasi. Untuk mengetahui apakah anggapan yang telah kita buat benar atau salah sehingga kita menerima atau menolak hipotesis, diperlukan pengujian dengan data sampel. Oleh karena memakai sampel, maka kebenaran suatu hipotesis statistik tidak pernah diketahui dengan pasti.

Hal ini tentu saja dalam kebanyakan situasi tidak dapat dilakukan. Oleh karena itu kita dapat mengambil sampel dari populasi tersebut untuk memutuskan apakah hipotesis kita benar atau salah yang ditentukan (dibuktikan) oleh data yang dikumpulkan.

Ada dua hipotesis, yaitu hipotesis nol (H_0) dan hipotesis alternatif atau hipotesis tandingan H_1 . Hipotesis nol adalah anggapan sementara yang dianggap benar dan akan diuji melalui data. Dalam proses pengujian, sebenarnya yang diharapkan adalah kita ingin menolak Hipotesis nol (H_0).

Apabila suatu penelitian dilakukan untuk menunjukkan pernyataan yang didukung kuat oleh data sampel, maka lawan dari pernyataan tersebut diambil sebagai H_0 dan pernyataan penelitian itu diambil sebagai H_1 . Kita mulai dari ilustrasi contoh berikut ini.

Metode mengajar cara konvensional di suatu kelas, memberikan nilai rata-rata ulangan harian siswa sebesar 75,25. Dengan menggunakan metode

mengajar yang baru, guru memperagakan cara mengajar yang lebih atraktif. Hal ini dilakukan dengan tujuan agar nilai ulangan harian siswa bisa lebih baik lagi. Setelah dilakukan ujian, diperoleh rata-rata nilai ulangan harian sebesar 80,25. Apakah metode baru tadi dapat dikatakan mampu menaikkan nilai ulangan harian siswa?

Contoh diatas mengilustrasikan bahwa guru tadi akan membuktikan pernyataan bahwa metode mengajar cara atraktif dapat menaikkan nilai ulangan harian. Pernyataan tadi merupakan hipotesis penelitian dari guru tadi, sehingga dapat dituliskan sebagai Hipotesis alternatif, H_1 . Sebagai komplemennya, kita dapat merumuskan pernyataan bahwa metode mengajar cara atraktif tidak dapat menaikkan nilai ulangan harian. Pernyataan ini disebut sebagai Hipotesis nol, H_0 .

Suatu uji hipotesis statistik yang alternatifnya bersifat satu-arah dinyatakan sebagai :

$$H_0 : \theta = \theta_0$$

$$H_1 : \theta > \theta_0.$$

Bentuk hipotesis di atas, biasa disebut sebagai uji sisi kanan.

atau

$$H_0 : \theta = \theta_0$$

$$H_1 : \theta < \theta_0$$

disebut uji sisi kiri.

Sedangkan uji hipotesis statistik yang alternatifnya bersifat dua-arah seperti

$$H_0 : \theta = \theta_0$$

$$H_1 : \theta \neq \theta_0$$

disebut uji dua arah.

B. JENIS-JENIS KESALAHAN DALAM UJI HIPOTESIS

Keputusan yang dicapai dari suatu uji hipotesis selalu dibayangi kemungkinan yang salah, yaitu menolak H_0 padahal H_0 benar atau menerima H_0 padahal H_0 salah. Kedua jenis kesalahan tadi, secara matematis dirumuskan sebagai:

a. Kesalahan jenis kesatu, yang disimbolkan dengan α , yaitu:

$$\alpha = P(\text{menolak } H_0 | H_0 \text{ benar})$$

b. Kesalahan jenis kedua, disimbolkan dengan β , yaitu

$$\beta = P(\text{menerima } H_0 | H_0 \text{ salah})$$

Dengan mempertimbangkan hakekat hipotesis nol, H_0 , dan kemungkinan keputusan yang diambil, salah satu situasi pasti akan timbul seperti pada diagram di bawah ini.

Persidangan			Uji hipotesis		
	Kenyataan			Kenyataan	
Putusan	Innocent	Guilty	Putusan	H_0 benar	H_0 salah
Innocent	Benar	Salah	Menerima H_0	$1 - \alpha$	β
Guilty	Salah	Benar	Menolak H_0	α	$1 - \beta$

Kesalahan tipe 1, α dan tipe 2, β , memiliki relasi yang berkebalikan. Idealnya kedua kesalahan minimal tetapi jika kesalahan yang satu diperkecil yang lain membesar. Dalam kenyataannya, nilai yang berpengaruh dalam uji hipotesis adalah α , yang lebih sering disebut sebagai tingkat kesalahan dan $1 - \alpha$ dikenal sebagai tingkat kepercayaan. Pada umumnya α ditentukan terlebih dahulu seperti $\alpha = 0,05$ dan $\alpha = 0,01$.

C. STATISTIK UJI DAN DAERAH KRITIK

Dalam uji hipotesis, sikap kita adalah menganggap H_0 benar, kecuali secara kuat data sampel tidak mendukungnya. Suatu uji hipotesis H_0 adalah aturan yang apabila harga-harga sampel telah diperoleh akan mengarah pada suatu keputusan menerima atau menolak H_0 itu. Statistik yang harganya digunakan untuk menentukan keputusan ini disebut sebagai statistik uji, dan himpunan harga-harga yang membuat H_0 ditolak dinamakan daerah penolakan (daerah kritik) uji tersebut.

Ada beberapa statistik uji yang dapat digunakan dengan syarat-syarat tertentu, diantaranya adalah uji Z, uji t, dan uji χ^2 .

6.2. UJI HIPOTESIS UNTUK MEAN SATU POPULASI

Jika peubah acak X berdistribusi Normal dengan mean μ dan simpangan baku σ , maka uji hipotesis yang perlu dilakukan adalah seperti langkah-langkah di bawah ini.

1. Menyusun hipotesis

Dengan mencermati pernyataan yang akan diuji, kita dapat merumuskan hipotesis dalam tiga kondisi, uji dua sisi, sisi kiri dan sisi kanan.

Uji dua sisi :

$H_0 : \mu = \mu_0$ yang dapat diartikan sebagai pernyataan rata-rata populasi masih sama, sebesar μ_0 , dengan tandingan

$H_1 : \mu \neq \mu_0$ yang dapat dituliskan sebagai rata-rata populasi sudah berubah, tidak sebesar μ_0 lagi.

Uji sisi kiri :

$H_0 : \mu = \mu_0$ atau kadang-kadang dituliskan dalam bentuk $\mu \geq \mu_0$ yang dapat diartikan sebagai rata-rata populasi tidak kurang dari μ_0

$H_1 : \mu < \mu_0$ yang diartikan sebagai rata-rata populasi kurang dari μ_0 .

Uji sisi kanan :

$H_0 : \mu = \mu_0$ atau kadang-kadang dituliskan sebagai $\mu \leq \mu_0$ yang diartikan sebagai rata-rata populasi tidak lebih dari μ_0

$H_1 : \mu > \mu_0$ yang diartikan sebagai rata-rata populasi lebih dari μ_0 .

2. Menentukan taraf signifikansi

Tingkat signifikansi atau tingkat kesalahan biasanya sudah ditentukan oleh peneliti, misalnya ada yang mengambil $\alpha = 5\%$ yang diartikan sebagai taraf nyata dan pengambilan $\alpha = 1\%$ yang diartikan sebagai taraf sangat nyata.

3. Menentukan statistik uji

Untuk uji rata-rata, kita mengenal dua bentuk statistik uji, uji Z dan uji t, yang masing-masing digunakan dalam kondisi yang berbeda-beda.

Dalam kondisi simpangan baku populasi σ diketahui, kita dapat melakukan pengujian hipotesis dengan menggunakan statistik uji Z, dengan persamaan

$$Z_{hit} = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}$$

dengan Z_{hit} berdistribusi normal baku. Sebaliknya, jika σ tidak diketahui kita dapat menggunakan informasi dari ukuran sampelnya, n . Jika n sangat besar, lebih dari 30, digunakan statistik uji

$$Z_{hit} = \frac{\sqrt{n}(\bar{X} - \mu)}{s}$$

dengan s adalah simpangan baku sampel.

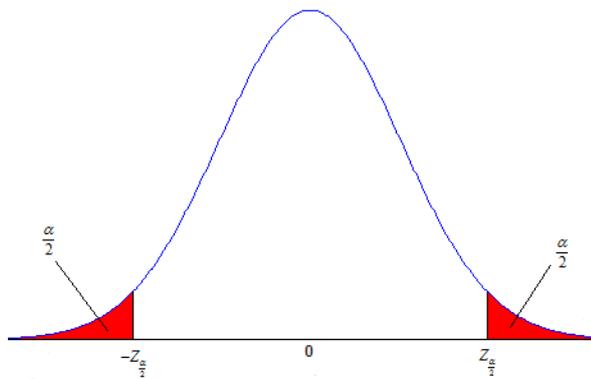
Khusus untuk n yang kurang dari 30, digunakan statistik uji

$$t_{hit} = \frac{\sqrt{n}(\bar{X} - \mu)}{s}$$

dengan t_{hit} berdistribusi student t dengan derajat bebas, $db = n - 1$.

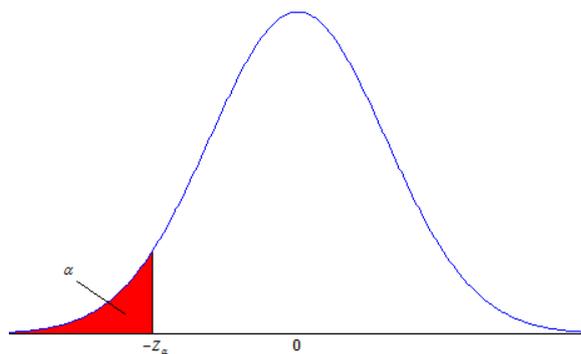
4. Merumuskan daerah kritis

Dengan memperhatikan hipotesis yang akan diuji, kita dapat merumuskan bentuk-bentuk daerah kritisnya. Jika uji yang dilakukan adalah uji dua sisi, daerah kritisnya adalah sebagai berikut:

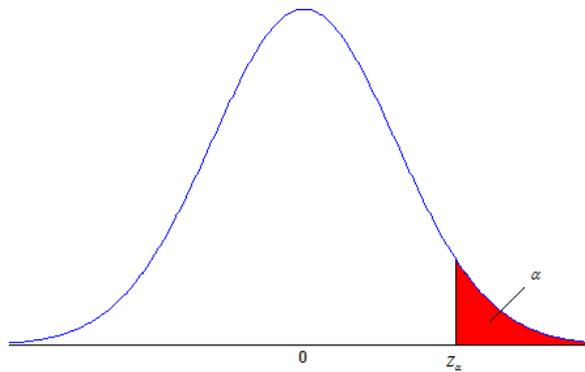


Keputusan yang diambil, H_0 ditolak jika $Z_{hit} < -Z_{\frac{\alpha}{2}}$ atau $Z_{hit} > Z_{\frac{\alpha}{2}}$

Untuk uji sisi kiri, H_0 ditolak jika $Z_{hit} < -Z_{\alpha}$ dengan gambar daerah kritisnya sebagai berikut.



Terakhir, jika dilakukan uji sisi kanan, H_0 ditolak jika $Z_{hit} > Z_{\alpha}$ dengan gambar daerah kritisnya adalah



5. Membuat kesimpulan

Dengan membandingkan hasil hitungan statistik uji dengan titik kritisnya dan kriteria penolakan H_0 , kesimpulan dapat diambil.

Contoh 6.1

Seorang peneliti pendidikan mengamati waktu belajar siswa dalam setahun rata-rata selama 800 jam dengan simpangan baku 60 jam. Akhir-akhir ini timbul dugaan bahwa perubahan lingkungan dan sistem kompetisi di sekolah telah mempengaruhi lama belajar siswa. Untuk menentukan hal ini, dilakukan penelitian dengan mengamati 50 siswa. Ternyata rata-ratanya 792 jam. Selidikilah dengan taraf nyata 0,05 apakah lama belajar sudah berubah atau belum.

Jawab :

Dengan memisalkan lama belajar siswa berdistribusi normal, maka kita akan menguji hipotesis :

$H_0 : \mu = 800$ jam, berarti lama belajar siswa masih sekitar 800 jam per tahunnya, terhadap hipotesis tandingan :

$H_1 : \mu \neq 800$ jam, berarti lama belajar siswa telah berubah dan bukan 800 jam lagi per tahunnya.

Diketahui dari soal informasi sebagai berikut : $\mu_0 = 800$, $\sigma = 60$, $n = 50$, $\bar{X} = 792$ dan $\alpha = 0,05$.

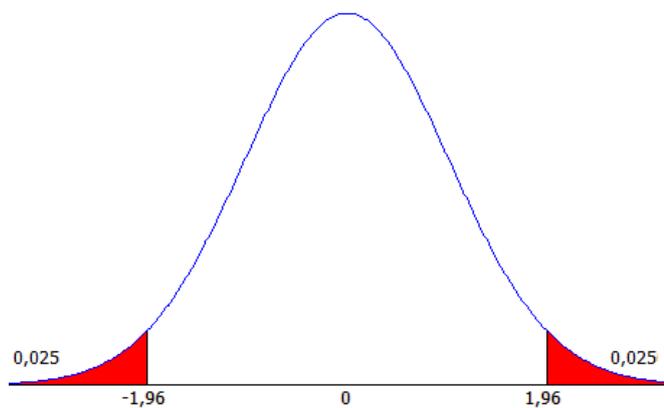
Ternyata, informasi tentang simpangan baku populasi diketahui, $\sigma = 60$, sehingga digunakan uji Z.

Selanjutnya dihitung besarnya nilai statistik uji Z,

$$Z_{hit} = \frac{\sqrt{n}(X - \mu_0)}{\sigma}$$

$$Z_{hit} = \frac{\sqrt{50}(792-800)}{60} = -0,94$$

Digunakan $\alpha = 0,05$. Karena merupakan uji dua sisi, terdapat dua daerah penolakan H_0 dan dua titik kritis. Luas masing-masing daerah penolakan adalah $\frac{\alpha}{2} = 0,025$. Titik kritis ditentukan dengan cara membaca Tabel distribusi Normal baku. Diperoleh titik kritis sebelah kanan $Z_{\frac{\alpha}{2}} = 1,96$ dan sebelah kiri $-Z_{\frac{\alpha}{2}} = -1,96$.



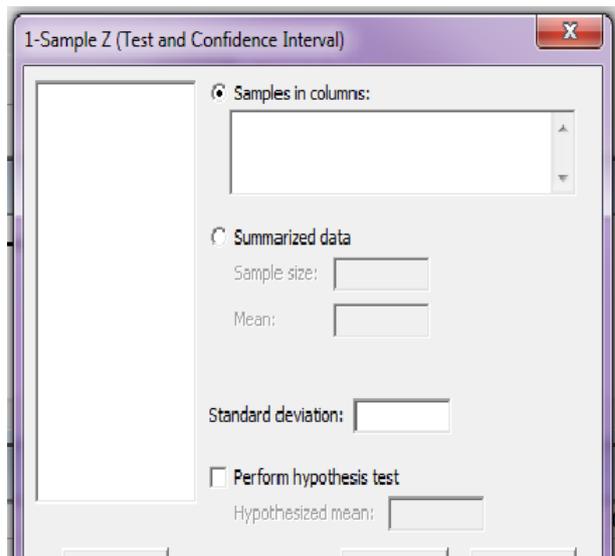
H_0 diterima jika Z_{hit} terletak antara -1,96 dan 1,96. Sebaliknya H_0 akan ditolak jika $Z_{hit} < -1,96$ atau $Z_{hit} > 1,96$.

Dari perhitungan statistik uji, diperoleh $Z_{hit} = -0,94$ dan hal ini jelas terletak dalam daerah penerimaan H_0 . Jadi H_0 diterima.

Ini berarti dalam taraf nyata 0,05 penelitian memperlihatkan bahwa memang lama belajar siswa masih sekitar 800 jam. Jadi tidak berubah.

Jika permasalahan tadi diselesaikan dengan menggunakan Minitab, langkah-langkahnya adalah sebagai berikut:

- 1) Pilih Stat >> Basic Statistics >> 1 Z- 1 Sample Z...



- 2) Ada dua cara untuk memasukkan informasi data, menggunakan list data asli yang dimasukkan dalam kolom data dan ringkasan data. Dari soal, yang diketahui adalah ringkasan data, dengan ukuran sampel $n = 50$, dan mean $\bar{X} = 792$. Masukkan simpangan baku populasi $\sigma = 60$. Masukkan mean populasi dengan cara klik pada Perform hypothesis test dan ketikkan di hypothesized mean 800.
- 3) Pilih option, masukkan confidence level : 95,00 dan H_1 diisikan not equal. Klik OK.
- 4) Hasilnya sebagai berikut ini.

One-Sample Z

Test of $\mu = 800$ vs not = 800
 The assumed standard deviation = 60

N	Mean	SE Mean	95% CI	Z	P
50	792,00	8,49	(775,37; 808,63)	-0,94	0,346

Interpretasi outputnya sebagai berikut:

Baris $\mu = 800$ vs not = 800 merupakan bentuk hipotesis yang akan diuji. $Z = -0,94$ adalah nilai statistik uji, yang akan dibandingkan dengan titik kritisnya. Jika diambil $\alpha = 0,05$, karena uji dua sisi, titik kritisnya adalah $Z_{\frac{\alpha}{2}} = 1,96$.

Nilai p adalah suatu nilai peluang yang dihitung dengan rumus :

$$p_{value} = P(Z \leq Z_{hit})$$

Dalam pengambilan keputusan, kita lebih suka menggunakan p_{value} dengan cara akan menolak H_0 jika $p_{value} < \alpha$.

Dari contoh kasus di atas, kesimpulan kita adalah menerima H_0 karena

$$p_{value} = 0,346 > \frac{\alpha}{2} = 0,025.$$

Contoh 6.2

Sebanyak 140 siswa di suatu SD diberi paket susu kemasan yang bertuliskan isi bersih 500 ml. Para siswa tersebut merasa bahwa volumenya kurang dari 500 ml. Untuk keperluan penyelidikan, sebelum susu diminum diukur dulu volumenya, ternyata dari 140 sampel tadi memiliki rata-rata volumenya adalah 480 ml dengan simpangan baku 150 ml. Dengan menggunakan $\alpha = 0,05$, apakah keluhan para siswa tadi terbukti?

Penyelesaian :

Hipotesis penelitian yang dapat disusun berasal dari informasi “siswa tersebut merasa bahwa volumenya kurang dari 500 ml”, sehingga bentuk dari H_1 adalah $\mu < 500$. Perumusan hipotesis secara lengkap adalah sebagai berikut:

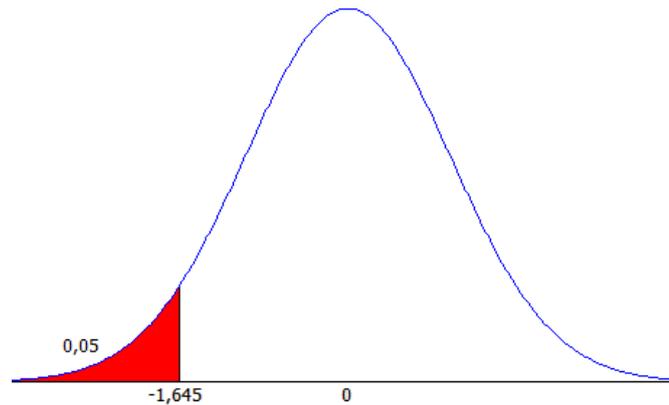
$H_0 : \mu = 500$ atau $\mu \geq 500$ ml, berarti bahwa volume susu kemasan tidak kurang dari 500 ml.

$H_1 : \mu < 500$ jam, berarti volume susu kemasan kurang dari 500 ml.

Informasi yang dapat dihimpun dari soal adalah $\mu_0 = 500$, $n = 140$, $\bar{X} = 480$, $s = 150$ dan $\alpha = 0,05$. Ternyata informasi tentang simpangan baku populasi σ tidak diketahui, tetapi ukuran sampelnya, $n = 140$ lebih dari 30, sehingga digunakan statistik uji Z_{hit} .

$$Z_{hit} = \frac{\sqrt{n}(X - \mu_0)}{\sigma} = \frac{\sqrt{140}(480 - 500)}{150} = -1,5776$$

Diketahui $\alpha = 0,05$ dan hipotesisnya merupakan uji sisi kiri, sehingga titik kritisnya adalah $-Z_{\alpha} = -1,645$ dengan gambar daerah kritis seperti berikut ini.



Karena merupakan uji sisi kiri, H_0 akan ditolak jika $Z_{hit} < -Z_{\alpha}$. Karena $Z_{hit} = -1,5776 > -Z_{\alpha} = -1,645$, maka H_0 diterima. Kesimpulan yang dapat diambil adalah rata-rata volume susu kemasan tidak kurang dari 500 ml, atau volumenya masih tetap.

Contoh 6.3

Tahun lalu rata-rata nilai UN tingkat SD di Kabupaten Banyumas adalah 23,75. Untuk menyelidiki kondisi ujian tahun ini, diambil sampel sebanyak 20 SD dan ternyata diperoleh rata-ratanya adalah 25,80 dengan simpangan baku 1,24. Dengan mengambil $\alpha = 5\%$ apakah tahun ini di Banyumas terjadi peningkatan nilai rata-rata UN?

Penyelesaian :

Hipotesis penelitian yang dapat disusun berasal dari informasi “apakah tahun ini di Banyumas terjadi peningkatan nilai rata-rata UN”, sehingga bentuk dari H_1 adalah $\mu > 23,75$. Perumusan hipotesis secara lengkap adalah sebagai berikut:

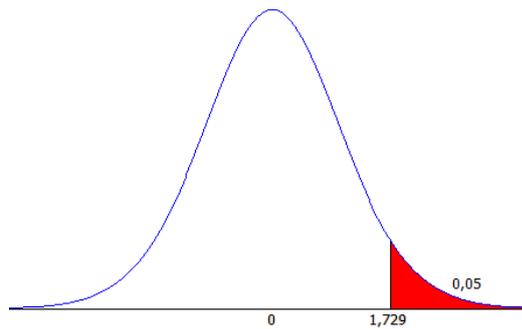
$H_0 : \mu = 23,75$ atau $\mu \leq 23,75$ ml, berarti bahwa rata-rata nilai UN tidak lebih dari 23,75.

$H_1 : \mu > 23,75$ berarti rata-rata nilai UN meningkat, lebih dari 23,75.

Informasi tentang simpangan baku populasi σ tidak diketahui dan ukuran sampelnya kecil, $n = 20$ kurang dari 30, sehingga digunakan statistik uji t.

$$t_{hit} = \frac{\sqrt{n}(\bar{X} - \mu_0)}{s} = \frac{\sqrt{20}(25,80 - 23,75)}{1,24} = 7,3935$$

Titik kritis diperoleh dari tabel distribusi student t dengan derajat bebas sebesar $db = n - 1$, sehingga untuk $n = 20$, $db = 19$ dan $\alpha = 0,05$ diperoleh: $t_{(1-\alpha;n-1)} = t_{(0,95;19)} = 1,729$. Secara grafis, daerah kritisnya dapat dilukiskan seperti gambar di bawah ini.



Karena merupakan uji sisi kanan, H_0 akan ditolak jika $t_{hit} > t_{(1-\alpha; n-1)}$. Karena $t_{hit} = 7,3935 > t_{(0,95;19)} = 1,729$, maka H_0 ditolak. Kesimpulan yang dapat diambil adalah memang benar bahwa rata-rata nilai ujian UN SD telah meningkat.

6.3. UJI HIPOTESIS UNTUK MEAN DUA POPULASI

Misalkan dua kelompok sampel random yang saling independen, yang pertama berukuran n_1 yang diambil dari populasi 1 dengan mean μ_1 dan yang kedua berukuran n_2 yang diambil dari populasi 2 dengan mean μ_2 . Kita akan membandingkan μ_1 dan μ_2 melalui uji hipotesis.

1. Untuk uji dua sisi

Bentuk hipotesisnya adalah:

- $H_0 : \mu_1 = \mu_2$
- $H_1 : \mu_1 \neq \mu_2$

Statistik ujinya jika σ_1 dan σ_2 diketahui adalah :

$$Z_{hit} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Daerah kritis, atau kriteria penolakannya adalah H_0 ditolak jika $Z_{hit} < -Z_{\frac{\alpha}{2}}$ atau $Z_{hit} > Z_{\frac{\alpha}{2}}$

Jika σ_1 dan σ_2 tidak diketahui dan dianggap bahwa $\sigma_1 = \sigma_2$ (artinya kedua populasi adalah homogen) dan $n_1 \geq 30$ dan $n_2 \geq 30$ statistik ujinya adalah :

$$Z_{hit} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \left[\frac{1}{n_1} + \frac{1}{n_2} \right]}}$$

Jika kedua populasi heterogen (artinya $\sigma_1 \neq \sigma_2$) statistik ujinya adalah

$$Z_{hit} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Jika σ_1 dan σ_2 tidak diketahui dan $n_1 < 30$ dan $n_2 < 30$ statistik uji dan daerah kritisnya adalah :

$$t_{hit} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \left[\frac{1}{n_1} + \frac{1}{n_2} \right]}}$$

Daerah kritisnya adalah H_0 ditolak jika $t_{hit} < -t_{\left(1-\frac{\alpha}{2}; n_1+n_2-2\right)}$ atau

$$t_{hit} > t_{\left(1-\frac{\alpha}{2}; n_1+n_2-2\right)}$$

2. Untuk uji sisi kanan

Bentuk hipotesisnya adalah:

- $H_0 : \mu_1 \leq \mu_2$
- $H_1 : \mu_1 > \mu_2$

Statistik ujinya sama dengan kasus uji dua sisi, hanya bedanya pada daerah kritisnya.

Daerah kritis untuk uji Z, atau kriteria penolakannya adalah H_0 ditolak jika $Z_{hit} > Z_\alpha$ dan untuk uji t, $t_{hit} > t_{(1-\alpha; n_1+n_2-2)}$

3. Untuk uji sisi kiri

Bentuk hipotesisnya adalah:

- $H_0 : \mu_1 \geq \mu_2$
- $H_1 : \mu_1 < \mu_2$

Daerah kritis untuk uji Z, atau kriteria penolakannya adalah H_0 ditolak jika $Z_{hit} < -Z_\alpha$ dan untuk uji t, $t_{hit} < -t_{(1-\alpha; n_1+n_2-2)}$

Contoh 6.4

Untuk membandingkan tingkat pemahaman materi ajar matematika terhadap jenis kelamin siswa, diteliti sebanyak 35 siswa putra dan 30 siswa putri di suatu SMP. Diperoleh informasi sebagai berikut: rata-rata nilai siswa putra sebesar 75,27 dengan simpangan baku 5,90 dan rata-rata siswa putri 77,26 dengan simpangan baku 7,12. Jika digunakan $\alpha = 5\%$ apakah kita dapat mengatakan bahwa siswa putri memiliki rata-rata nilai yang lebih tinggi?

Penyelesaian :

Jika μ_1 adalah rata-rata populasi nilai siswa putri dan μ_2 adalah rata-rata nilai siswa putra, maka dapat disusun hipotesis sebagai berikut:

- $H_0 : \mu_1 \leq \mu_2$ artinya rata-rata nilai siswa putri tidak lebih tinggi dari siswa putra
- $H_1 : \mu_1 > \mu_2$ artinya rata-rata nilai siswa putri lebih tinggi dari siswa putra

yang merupakan bentuk hipotesis uji sisi kanan.

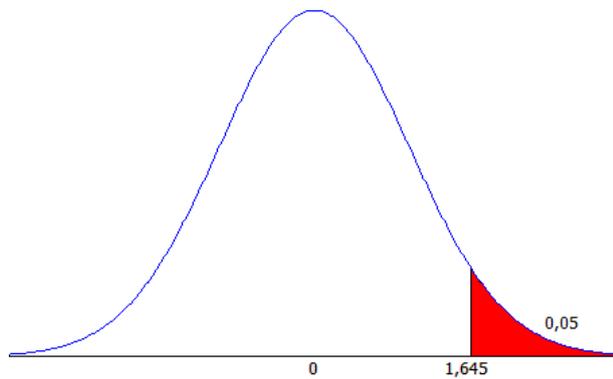
Diketahui $n_1 = 30$, $n_2 = 35$, $\bar{X}_1 = 77,26$, $\bar{X}_2 = 75,27$, $s_1 = 7,12$ dan $s_2 = 5,90$. Ternyata informasi untuk σ_1 dan σ_2 tidak diketahui dan $n_1 \geq 30$ dan $n_2 \geq 30$, sehingga digunakan uji Z. Karena siswa tadi diambil dari sekolah yang sama, maka kedua populasi dianggap homogen, sehingga statistik uji yang digunakan adalah:

$$Z_{hit} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \left[\frac{1}{n_1} + \frac{1}{n_2} \right]}}$$

$$Z_{hit} = \frac{77,26 - 75,27}{\sqrt{\frac{(30 - 1)7,12^2 + (35 - 1)5,90^2}{30 + 35 - 2} \left[\frac{1}{30} + \frac{1}{35} \right]}} = \frac{1,99}{\sqrt{42,12187 \left[\frac{1}{30} + \frac{1}{35} \right]}}$$

$$= 1,2324$$

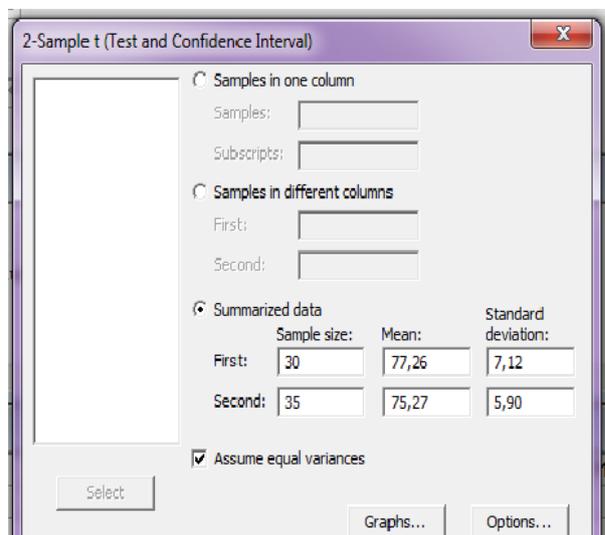
Karena uji Z sisi kanan dengan $\alpha = 0,05$, maka titik kritisnya adalah $Z_{0,05} = 1,645$. Secara grafis terlukis daerah kritis seperti gambar berikut.



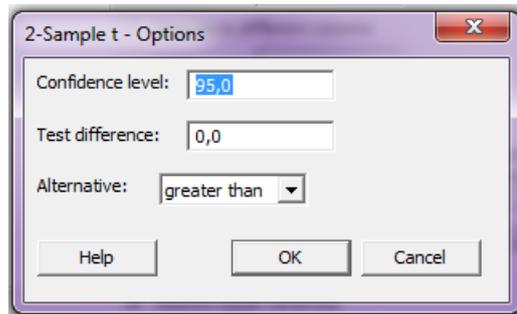
Dalam uji sisi kanan, H_0 ditolak jika $Z_{hit} > Z_{\alpha}$. Karena $Z_{hit} = 1,2324 < Z_{0,05} = 1,645$ maka H_0 diterima. Artinya tidak cukup bukti kalau rata-rata siswa perempuan lebih tinggi dari rata-rata siswa laki-laki. Atau, dengan kata lain, rata-rata nilai kedua jenis kelamin masih sama.

Jika digunakan Minitab, langkahnya adalah sebagai berikut:

- 1) Stat >> Basic Statistics >> 2t - 2 sample t..... Langkah ini kita pilih, karena dalam Minitab 16 tidak tersedia uji 2 Z. Hasil yang diberikan akan sama, karena untuk n besar distribusi t juga dekat ke distribusi Normal standar.
- 2) Masukkan semua informasi yang ada seperti pada gambar.



- 1) Pilih Option, masukkan semua informasi yang sesuai.



Hasilnya adalah sebagai berikut :

Two-Sample T-Test and CI

Sample	N	Mean	StDev	SE Mean
1	30	77,26	7,12	1,3
2	35	75,27	5,90	1,0

Difference = mu (1) - mu (2)
 Estimate for difference: 1,99
 95% lower bound for difference: -0,71
 T-Test of difference = 0 (vs >): T-Value = 1,23 P-Value = 0,111 DF = 63
 Both use Pooled StDev = 6,4901

Dengan $p_{value} = 0,111$, maka H_0 diterima.

6.4. UJI HIPOTESIS UNTUK PROPORSI SATU POPULASI

Uji mengenai proporsi banyak dipakai dalam berbagai bidang. Disini akan dibahas persoalan pengujian hipotesis bahwa proporsi sukses dalam suatu percobaan binomial, $p = \frac{X}{n}$, dengan p adalah proporsi “sukses” suatu kategori/peristiwa tertentu, X adalah banyaknya kategori yang sukses, dan n adalah banyaknya percobaan atau pengamatan yang dilakukan.

Distribusi sampling harga proporsi $p = \frac{X}{n}$ yang mempunyai mean π_0 dan

simpangan baku $\sqrt{\frac{\pi_0(1-\pi_0)}{n}}$ dan untuk ukuran sampel yang besar, akan

berdistribusi normal baku dengan sampling statistiknya $Z = \frac{\frac{X}{n} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$.

Langkah-langkah yang ditempuh dalam uji hipotesis ini serupa dengan uji hipotesis rata-rata satu populasi, yaitu dengan menyusun:

1. Hipotesis

Ada tiga bentuk uji, yaitu:

- Uji sisi kanan $H_0 : \pi = \pi_0$ vs $H_1 : \pi \neq \pi_0$
- Uji sisi kanan $H_0 : \pi \leq \pi_0$ atau $\pi = \pi_0$ vs $H_1 : \pi > \pi_0$
- Uji sisi kiri $H_0 : \pi \geq \pi_0$ atau $\pi = \pi_0$ vs $H_1 : \pi < \pi_0$

2. Statistik uji

Statistik uji yang dipakai hanya satu, yaitu uji Z dengan rumus:

$$Z_{hit} = \frac{\frac{X}{n} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$$

3. Daerah kritis

- Untuk uji dua sisi, H_0 akan ditolak jika $Z_{hit} < -Z_{\frac{\alpha}{2}}$ atau $Z_{hit} > Z_{\frac{\alpha}{2}}$
- Untuk uji sisi kanan, H_0 akan ditolak jika $Z_{hit} > Z_{\alpha}$
- Untuk uji sisi kiri, H_0 akan ditolak jika $Z_{hit} < -Z_{\alpha}$.

Contoh 6.5

Dinyatakan bahwa 30% mahasiswa baru di suatu universitas memilih untuk ikut SPMB lagi. Pernyataan ini dibantah oleh bagian kemahasiswaan, yang menyatakan bahwa mahasiswa yang ikut SPMB kurang dari 30%. Diamati 500 mahasiswa baru, dan ternyata diperoleh fakta bahwa 124 mahasiswa mengikuti SPMB lagi. Dengan tingkat signifikansi 2,5% apa kesimpulan kita?

Penyelesaian:

Diketahui $\pi_0 = 0,3$, $n = 500$, $X = 124$, dan $\alpha = 0,025$.

Hipotesis yang akan diuji adalah sebagai berikut:

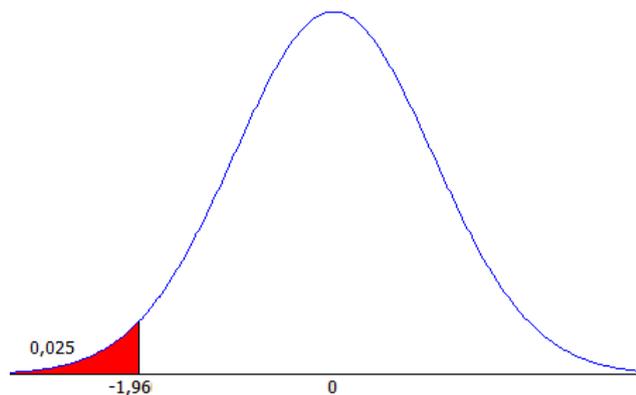
- $H_0 : \pi \geq 0,3$ atau $\pi = 0,3$
artinya 30% mahasiswa baru di universitas tersebut masih ikut SPMB lagi, terhadap hipotesis penelitian :
- $H_1 : \pi < 0,3$ artinya kurang dari 30% mahasiswa baru yang ikut SPMB lagi.

Nilai statistik ujinya adalah :

$$Z_{hit} = \frac{\frac{X}{n} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{\sqrt{n} \left(\frac{X}{n} - \pi_0 \right)}{\sqrt{\pi_0(1-\pi_0)}}$$

$$Z_{hit} = \frac{\sqrt{500} \left(\frac{124}{500} - 0,3 \right)}{\sqrt{0,3(1-0,3)}} = -2,5373$$

Untuk $\alpha = 0,025$ dan uji sisi kiri, nilai titik kritisnya adalah $-Z_{0,025} = -1,96$.



H_0 akan ditolak jika $Z_{hit} < -Z_{\alpha}$. Karena $Z_{hit} = -2,5373 < -Z_{\alpha} = -1,96$, maka H_0 ditolak, artinya memang benar bahwa mahasiswa yang ikut SPMB lagi sudah kurang dari 30%.

Jika kita gunakan Minitab, langkahnya adalah sebagai berikut:

- 1) Stat >> Basic Statistics >> 1P – 1 Proportion...
- 2) Pilihlah summarized data, karena kita hanya memiliki ringkasan data. Masukkan number of events banyaknya kejadian yang “sukses” yaitu 124 dan number of trials banyaknya percobaan, yaitu 500.
- 3) Aktifkan pilihan Perform hypothesis test, dan masukkan nilai $\pi_0 = 0,3$
- 4) Klik option, masukkan nilai signifikansi 97,5%, karena kita mengambil $\alpha = 0,025$
- 5) Pilihlah less than untuk alternative, karena kita menggunakan uji sisi kiri. Klik Use test and interval based on normal distribution, kemudian klik OK.

Hasilnya sebagai berikut:

Test of $p = 0,3$ vs $p < 0,3$

Sample	X	N	Sample p	97,5% Upper Bound	Z-Value	P-Value
1	124	500	0,248000	0,285853	-2,54	0,006

Using the normal approximation.

Jika kita menggunakan informasi dari p_{value} , kesimpulan kita lebih mudah, karena hanya dibandingkan dengan nilai α . Karena $p_{value} = 0,006 < \alpha = 0,026$, maka H_0 ditolak.

6.5. UJI HIPOTESIS UNTUK PROPORSI DUA POPULASI

Kadang-kadang dalam penelitian kita seringkali membandingkan suatu kategori dari dua populasi yang berbeda. Misalkan, kita akan membandingkan kategori lulusan yang berpredikat cumlaude dari dua program studi yang berbeda. Program Studi PAI pada acara wisuda bulan Juli tahun lalu mendapatkan 12 wisudawan yang cumlaude dari 98 wisudawan, sedangkan Prodi PGMI ada 9 yang cumlaude dari 81 wisudawan. Apakah dapat disimpulkan bahwa proporsi wisudawan yang cumlaude dari dua prodi tadi berbeda?

Untuk menjawab permasalahan tadi, kita dapat melakukan uji sebagai berikut. Misalkan π_1 adalah proporsi dari populasi 1 dan π_2 adalah proporsi dari populasi 2. Kita dapat membentuk suatu statistik Z yang berdistribusi Normal standar, dengan persamaan :

$$Z = \frac{\left(\frac{X_1}{n_1} - \frac{X_2}{n_2} \right) - (\pi_1 - \pi_2)}{\sqrt{\left(\frac{\frac{X_1}{n_1} \left(1 - \frac{X_1}{n_1} \right)}{n_1} + \frac{\frac{X_2}{n_2} \left(1 - \frac{X_2}{n_2} \right)}{n_2} \right)}}$$

Langkah-langkah yang dilakukan dalam uji hipotesis analog dengan uji rata-rata ataupun uji yang lainnya.

Statistik uji yang digunakan, ada beberapa orang yang menuliskan dalam bentuk :

$$Z_{hit} = \frac{\left(\frac{X_1}{n_1} - \frac{X_2}{n_2} \right)}{\sqrt{\left(\hat{p}(1 - \hat{p}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \right)}}$$

dengan

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2}$$

dan

$$\pi_1 - \pi_2 = 0 \text{ dibawah asumsi } H_0.$$

Contoh 6.6

Kita pandang ilustrasi pemahaman di atas. Misalkan, kita akan membandingkan kategori lulusan yang berpredikat cumlaude dari dua program studi yang berbeda. Program Studi PAI pada acara wisuda bulan Juli tahun lalu mendapatkan 12 wisudawan yang cumlaude dari 98 wisudawan, sedangkan Prodi PGMI ada 9 yang cumlaude dari 81 wisudawan. Apakah dapat disimpulkan bahwa proporsi wisudawan yang cumlaude dari dua prodi tadi berbeda, dengan menggunakan tingkat kesalahan 5%?

Hipotesis yang dapat disusun adalah :

- $H_0 : \pi_1 = \pi_2$ yang berarti proporsi lulusan yang cumlaude dari Prodi PAI dengan Prodi PGMI adalah sama.
- $H_1 : \pi_1 \neq \pi_2$ yang berarti ada perbedaan proporsi lulusan yang cumlaude dari Prodi PAI dengan Prodi PGMI.

Statistik uji dapat dihitung dengan cara :

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{12 + 9}{98 + 81} = \frac{21}{179} = 0,1173$$

$$Z_{hit} = \frac{\left(\frac{X_1}{n_1} - \frac{X_2}{n_2} \right)}{\sqrt{\left(\hat{p}(1-\hat{p}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \right)}} = \frac{\left(\frac{12}{98} - \frac{9}{81} \right)}{\sqrt{(0,1173)(0,8827) \left(\frac{1}{98} + \frac{1}{81} \right)}} = 0,2346$$

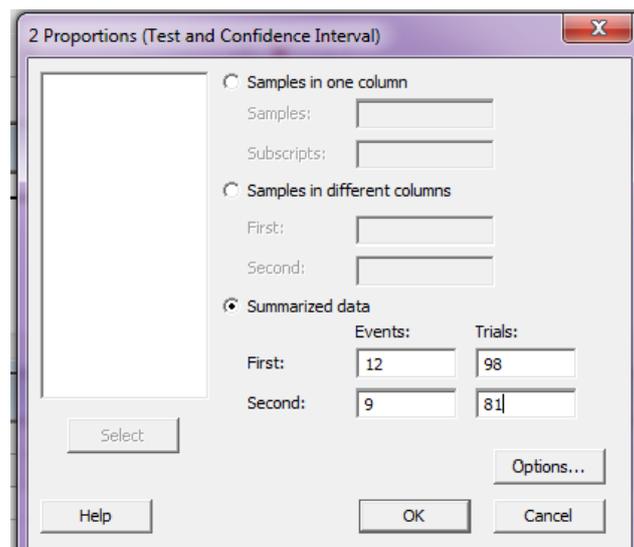
Dengan mengambil $\alpha = 0,05$ diperoleh titik-titik kritisnya adalah $-Z_{\frac{\alpha}{2}} = -1,96$ dan $Z_{\frac{\alpha}{2}} = 1,96$.

H_0 akan ditolak jika $Z_{hit} > Z_{\frac{\alpha}{2}}$ atau jika $Z_{hit} < -Z_{\frac{\alpha}{2}}$.

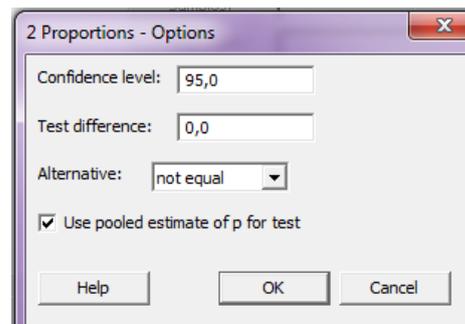
Karena $Z_{hit} = 0,2346$ berada dalam daerah penerimaan, maka H_0 diterima. Artinya proporsi lulusan yang cumlaude dari dua prodi tadi masih dianggap sama.

Jika digunakan Minitab, kita lakukan langkah-langkah sebagai berikut:

- 1) Stat >> Basic Statistics >> 2P - 2 Proportions.....
- 2) Masukkan angka-angka yang diketahui ke dalam kotak yang sesuai. Lihat gambar di bawah ini.



- 3) Pilih Option, masukkan seperti pada gambar



Hasilnya adalah sebagai berikut:

Test and CI for Two Proportions

Sample	X	N	Sample p
1	12	98	0,122449
2	9	81	0,111111

Difference = p (1) - p (2)
 Estimate for difference: 0,0113379
 95% CI for difference: (-0,0829812; 0,105657)
 Test for difference = 0 (vs not = 0): Z = 0,23 P-Value = 0,814

Dari output di atas, kesimpulan yang dapat diambil adalah menerima H_0 , karena $p\text{-value} = 0,814 > \alpha = 0,05$, berarti proporsi lulusan yang cumlaude dari Prodi PAI dengan Prodi PGMI adalah sama

6.6. UJI HIPOTESIS UNTUK VARIANSI SATU POPULASI

Sama halnya pada uji rata-rata dan proporsi, uji variansi juga memiliki langkah-langkah sebagai berikut:

1) Menyusun hipotesis

1) Uji dua sisi memiliki bentuk hipotesis sebagai berikut:

$$H_0 : \sigma^2 = \sigma_0^2 \text{ terhadap } H_1 : \sigma^2 \neq \sigma_0^2 .$$

2) Bentuk hipotesis uji sisi kanan adalah :

$$H_0 : \sigma^2 \leq \sigma_0^2 \text{ atau } \sigma^2 = \sigma_0^2 \text{ terhadap } H_1 : \sigma^2 > \sigma_0^2$$

3) Uji sisi kiri bentuk hipotesisnya adalah :

$$H_0 : \sigma^2 \geq \sigma_0^2 \text{ atau } \sigma^2 = \sigma_0^2 \text{ terhadap } H_1 : \sigma^2 < \sigma_0^2$$

2) Statistik uji

Dalam uji variansi, kita hanya mengenal satu bentuk statistik uji, yaitu :

$$\chi^2_{hit} = \frac{(n-1)s^2}{\sigma_0^2} \text{ dimana } \chi^2_{hit} \text{ berdistribusi chi kuadrat dengan derajat}$$

bebas db = $n - 1$.

3) Daerah kritis

a. Untuk uji dua sisi, H_0 akan ditolak jika

$$\chi^2_{hit} < \chi^2_{\left(\frac{\alpha}{2}; n-1\right)} \text{ atau } \chi^2_{hit} > \chi^2_{\left(1-\frac{\alpha}{2}; n-1\right)}$$

b. Untuk uji sisi kanan, H_0 akan ditolak jika $\chi^2_{hit} > \chi^2_{(1-\alpha; n-1)}$

c. Untuk uji sisi kiri, H_0 akan ditolak jika $\chi^2_{hit} < \chi^2_{(\alpha; n-1)}$

Contoh 6.7

Deviasi nilai UN SMP tahun lalu sudah dapat dikatakan ideal karena memiliki variansi 0,25. Setelah mengamati hasil ujian tahun ini, ternyata dari 25 sampel diperoleh variansi 0,32. Apakah hasil tahun ini sudah dapat dikatakan tidak ideal lagi? Gunakan tingkat kesalahan 5%.

Penyelesaian :

Suatu pengamatan dikatakan memiliki efisiensi yang baik apabila memiliki nilai variansi yang kecil, sehingga kita dapat merumuskan hipotesis penelitiannya adalah ada peningkatan nilai variansi (karena diduga ada penurunan efisiensi).

1) Hipotesis yang dapat dirumuskan adalah sebagai berikut:

$$H_0 : \sigma^2 \leq 0,25 \text{ atau } \sigma^2 = 0,25 \text{ terhadap } H_1 : \sigma^2 > 0,25$$

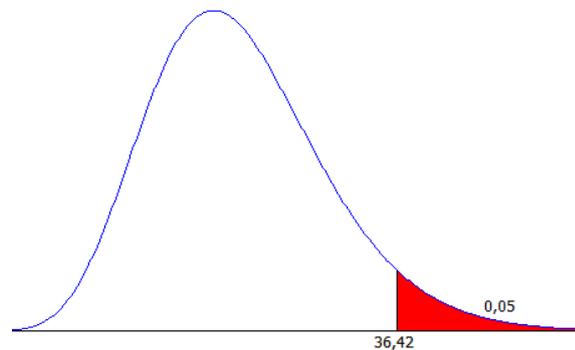
2) Statistik uji dapat dihitung dengan cara sebagai berikut:

$$\chi^2_{hit} = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(25-1)32}{25} = 30,72$$

Daerah kritis kita tentukan dengan menetapkan titik kritisnya terlebih dahulu. Karena uji ini termasuk uji sisi kanan, titik kritisnya adalah :

$$\chi^2_{(1-\alpha;n-1)} = \chi^2_{(1-0,05;25-1)} = \chi^2_{(0,95;24)}$$

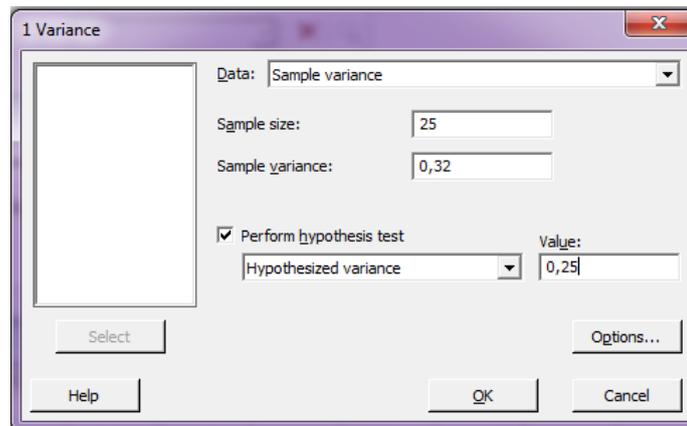
Dengan melihat tabel distribusi χ^2 , diperoleh $\chi^2_{(0,95;24)} = 36,4150$



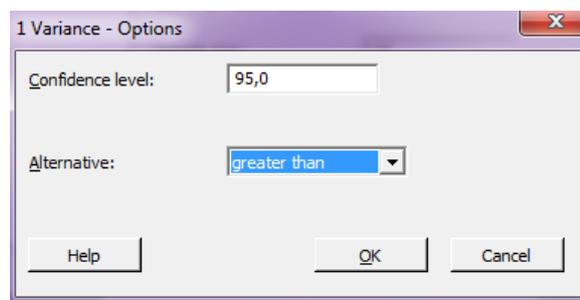
Kriteria penolakan H_0 pada uji sisi kanan, H_0 ditolak jika $\chi^2_{hit} > \chi^2_{tabel}$. Karena $\chi^2_{hit} = 30,72 < \chi^2_{tabel} = 36,42$ maka H_0 diterima. Artinya tidak cukup bukti kalau dikatakan terjadi penurunan efisiensi. Atau, dalam kalimat lain dapat diartikan bahwa sebaran nilai UN masih ideal.

Langkah-langkah jika menggunakan Minitab adalah sebagai berikut:

- 1) Stat >> Basic Statistics >> σ^2 1 variance
- 2) Masukkan semua informasi seperti pada gambar di bawah ini.



3) Pilih Option, masukkan informasi yang sesuai.



4) Klik OK, hasilnya adalah seperti di bawah ini.

Test and CI for One Variance

Method

Null hypothesis Sigma-squared = 0,25
 Alternative hypothesis Sigma-squared > 0,25

The chi-square method is only for the normal distribution.
 The Bonett method cannot be calculated with summarized data.

Statistics

N	StDev	Variance
25	0,566	0,320

95% One-Sided Confidence Intervals			
Method	Lower Bound	for StDev	Lower Bound
			for Variance
Chi-Square	0,459		0,211

Tests

Method	Test		
	Statistic	DF	P-Value
Chi-Square	30,72	24	0,162

Karena $p_{value} = 0,162 > \alpha = 0,05$, maka H_0 diterima.

6.7. Uji Hipotesis Untuk Variansi Dua Populasi

Salah satu tujuan dari uji hipotesis dua variansi bertujuan untuk menyelidiki apakah dua populasi tersebut memiliki variansi yang sama (homogen) atau berbeda (heterogen).

Sama halnya dengan uji rata-rata dan proporsi, uji hipotesis dua variansi juga mempunyai tiga bentuk hipotesis, yaitu uji dua sisi, sisi kiri dan sisi kanan.

1) Penyusunan Hipotesis

a. Uji dua sisi

$H_0 : \sigma_1^2 = \sigma_2^2$ yang berarti dua populasi adalah homogen, terhadap

$H_1 : \sigma_1^2 \neq \sigma_2^2$ yang berarti kedua populasi heterogen.

b. Uji sisi kanan

$H_0 : \sigma_1^2 = \sigma_2^2$, atau $\sigma_1^2 \leq \sigma_2^2$ yang berarti populasi 1 memiliki variansi yang tidak lebih besar dari variansi populasi 2, terhadap

$H_1 : \sigma_1^2 > \sigma_2^2$ yang berarti populasi 1 memiliki variansi yang lebih besar daripada variansi populasi 2

c. Uji sisi kiri

$H_0 : \sigma_1^2 = \sigma_2^2$, atau $\sigma_1^2 \geq \sigma_2^2$ yang berarti populasi 1 memiliki variansi yang tidak lebih kecil dari variansi populasi 2, terhadap

$H_1 : \sigma_1^2 < \sigma_2^2$ yang berarti populasi 1 memiliki variansi yang lebih besar daripada variansi populasi 2

2) Statistik Uji

Statistik uji yang digunakan adalah :

$$F_{hit} = \frac{s_1^2}{s_2^2}$$

yang berdistribusi F dengan derajat bebas pembilangnya, $db_1 = n_1 - 1$ dan derajat bebas penyebutnya, $db_2 = n_2 - 1$.

3) Daerah Kritis

a. Uji dua sisi

$$H_0 \text{ ditolak jika } F_{hit} > F_{\left(n_1-1; n_2-1; \frac{\alpha}{2}\right)} \text{ atau } F_{hit} < \frac{1}{F_{\left(n_2-1; n_1-1; \frac{\alpha}{2}\right)}}$$

b. Uji sisi kanan

$$H_0 \text{ ditolak jika } F_{hit} > F_{(n_1-1; n_2-1; \alpha)}$$

c. Uji sisi kiri

$$H_0 \text{ ditolak jika } F_{hit} < \frac{1}{F_{(n_2-1; n_1-1; \alpha)}}$$

Contoh 6.8

Seorang guru memperagakan metode mengajar yang baru, berbeda dengan kebanyakan guru yang ada di suatu sekolah tersebut. Metode tadi dicobakan pada dua kelas paralel yang berbeda, yaitu di kelas 2A dan kelas 2B. Diambil sampel random dari masing-masing kelas, 10 siswa dari kelas 2A dan 16 siswa dari kelas 2B, hasilnya adalah kelas 2A memiliki variansi 0,36 dan kelas 2B variansinya 0,87. Apakah kedua kelas percobaan tadi memiliki perbedaan variansi? Gunakan tingkat kesalahan 2%.

Penyelesaian :

Hipotesis yang akan diuji dapat dirumuskan sebagai: $H_0 : \sigma_1^2 = \sigma_2^2$ yang berarti sebaran nilai kelas 2A dan 2B sama, terhadap $H_0 : \sigma_1^2 \neq \sigma_2^2$ yang berarti terdapat perbedaan sebaran nilai pada kedua kelas pengamatan. Di sini kita akan melakukan uji dua sisi.

Statistik uji dihitung dari informasi sampel sebagai berikut :

$$F_{hit} = \frac{s_1^2}{s_2^2} = \frac{0,36^2}{0,87^2} = 0,17$$

Titik-titik kritis kita tentukan dengan cara membaca tabel distribusi F seperti berikut.

Titik kritis yang kanan,

$$F_{\left(n_1-1; n_2-1; \frac{\alpha}{2}\right)} = F_{\left(10-1; 16-1; \frac{0,02}{2}\right)} = F_{(9; 15; 0,01)}$$

Dengan derajat bebas pembilang 9 dan penyebut 15 dan $\frac{\alpha}{2} = 0,01$, diperoleh $F_{tabel} = 3,89$.

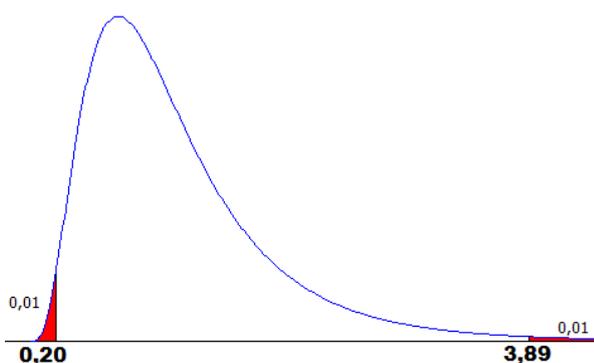
Titik kritis yang kiri,

$$\frac{1}{F_{\left(n_2-1; n_1-1; \frac{\alpha}{2}\right)}} = \frac{1}{F_{\left(16-1; 10-1; \frac{\alpha}{2}\right)}} = \frac{1}{F_{(15;9;0,01)}}$$

Dengan derajat bebas pembilang 15 dan penyebut 9 dan $\frac{\alpha}{2} = 0,01$, diperoleh $F_{(15;9;0,01)} = 4,96$, sehingga

$$\frac{1}{F_{(15;9;0,01)}} = \frac{1}{4,96} = 0,20$$

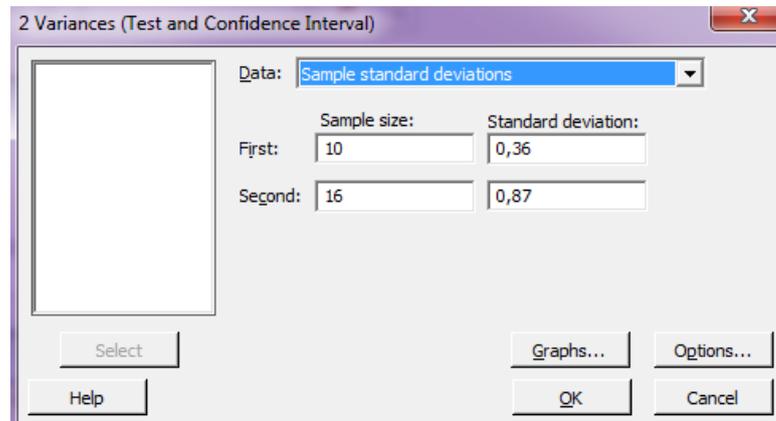
Jadi, H_0 akan ditolak jika $F_{hit} > 3,89$ atau $F_{hit} < 0,20$.



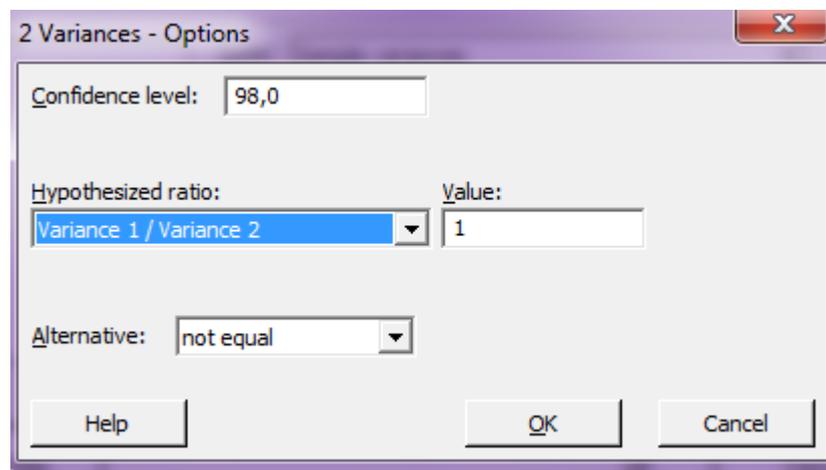
Karena $F_{hit} = 0,17 < 0,20$, maka H_0 ditolak. Artinya bahwa memang ada perbedaan sebaran nilai di dua kelas tadi karena metode mengajar guru. Dengan melihat nilai variansi dari masing-masing sampel, dapat dikatakan bahwa kelas 2A memiliki sebaran nilai yang lebih baik jika dibandingkan dengan kelas 2B.

Jika digunakan Minitab, prosedurnya adalah sebagai berikut:

- 1) Stat >> Basic Statistics >> 2 variances
- 2) Masukkan semua informasi seperti pada gambar di bawah ini.



3) Pilih Option, masukkan informasi yang sesuai.



4) Klik OK, hasilnya adalah seperti di bawah ini.

Test and CI for Two Variances

Method

Null hypothesis Variance(1) / Variance(2) = 1
 Alternative hypothesis Variance(1) / Variance(2) not = 1
 Significance level Alpha = 0,02

Statistics

Sample	N	StDev	Variance
1	10	0,360	0,130
2	16	0,870	0,757

Ratio of standard deviations = 0,414
 Ratio of variances = 0,171

98% Confidence Intervals

Distribution of Data	CI for StDev Ratio	CI for Variance Ratio
Normal	(0,210; 0,922)	(0,044; 0,850)

Tests

Method	DF1	DF2	Test Statistic	P-Value
F Test (normal)	9	15	0,17	0,011

Terlihat di output, bahwa $p_{value} = 0,011$. Ternyata $p_{value} = 0,011 < \alpha = 0,02$ sehingga H_0 ditolak.

7

Regresi Linier

Analisa pasangan variabel membutuhkan data yang terdiri dari dua kelompok hasil observasi atau pengukuran. Data yang dapat diperoleh dari berbagai kegiatan yang menghasilkan pasangan observasi sebanyak n , dinyatakan sebagai pasangan data $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$.

Regresi linear adalah alat statistik yang dipergunakan untuk mengetahui pengaruh antara satu atau beberapa variabel terhadap satu buah variabel. Variabel yang mempengaruhi sering disebut variabel bebas, variabel independen atau variabel penjelas. Variabel yang dipengaruhi sering disebut dengan variabel terikat atau variabel dependen.

Secara umum regresi linear terdiri dari dua, yaitu regresi linear sederhana yaitu dengan satu buah variabel bebas dan satu buah variabel terikat; dan regresi linear berganda dengan beberapa variabel bebas dan satu buah variabel terikat. Analisis regresi linear merupakan metode statistik yang paling sering digunakan dalam penelitian-penelitian eksak, sosial, terutama penelitian ekonomi dan pendidikan.

7.1. REGRESI LINIER SEDERHANA

Analisis regresi linear sederhana dipergunakan untuk mengetahui pengaruh antara satu buah variabel bebas terhadap satu buah variabel terikat. Persamaan umumnya adalah:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Suatu variabel dikategorikan sebagai variabel bebas apabila variabel itu fungsinya menerangkan variabel lainnya. Atau meramalkan variabel lainnya, oleh karena itu dalam penelitian variabel bebas sering disebut prediktor. Suatu variabel disebut variabel tak bebas apabila dalam hubungannya dengan variabel lain, dia merupakan variabel yang diterangkan (diramalkan) oleh variabel lainnya, oleh karena itu variabel tak bebas sering disebut respons.

1. Membentuk Persamaan Regresi

Hubungan antara 2 variabel atau lebih secara matematis mengikuti sesuatu fungsi matematis tertentu. Dalam matematika kita mempunyai fungsi $Y = f(X_1, X_2, \dots, X_n)$. Ini artinya semua perubahan yang terjadi pada Y sepenuhnya bisa diterangkan oleh variabel X_1, X_2, \dots, X_n . Fungsi yang seperti itu disebut fungsi deterministik.

Apabila kita mempunyai 2 buah variabel X dan Y, X merupakan variabel bebas dan Y variabel tidak bebas. Sedangkan fungsinya berbentuk persamaan linier. Disebut linier karena baik variabel bebas maupun tidak bebas kedua-duanya pangkatnya 1. (Definisi ini bukan merupakan definisi ketat, tetapi semata-mata hanya definisi “bebas”, sebab secara ketat linieritas dari model regresi bukan ditentukan oleh variabelnya tetapi oleh parameternya). Disebut sederhana karena dalam model itu hanya ada sebuah variabel bebas dan sebuah variabel tidak bebas. Dalam model populasi β_0 dan β_1 merupakan parameter, dalam model sampel b_0 dan b_1 merupakan statistik, sehingga tujuan kita sebenarnya akan membentuk suatu persamaan regresi yang dibentuk berdasarkan informasi sampel :

$$\hat{Y} = b_0 + b_1 X .$$

Untuk menentukan nilai b_0 dan b_1 digunakan metode kuadrat terkecil. Nilai b_0 dan b_1 dalam persamaan regresi linier biasanya disebut sebagai koefisien regresi.

Koefisien-koefisien regresi b_0 dan b_1 untuk regresi linier, dapat dihitung dengan:

$$b_0 = \frac{(\sum X)(\sum X^2) - (\sum X)(\sum XY)}{n\sum X^2 - (\sum X)^2}$$

dan

$$b_1 = \frac{n\sum XY - (\sum X)(\sum Y)}{n\sum X^2 - (\sum X)^2}$$

Bentuk rumus-rumus yang lain untuk menentukan koefisien regresi adalah sebagai berikut:

$$b_1 = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}$$

atau

$$b_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

atau

$$b_1 = \frac{\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}}{\sum X_i^2 - \frac{(\sum X_i)^2}{n}} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

dan

$$b_o = \bar{Y} - b_1 \bar{X}$$

Bentuk rumus yang lain jika menggunakan jumlah kuadrat (SS) adalah sebagai berikut:

$$b_1 = \frac{SS_{XY}}{SS_X}$$

dengan

$$SS_X = \sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - \frac{\left(\sum_{i=1}^n X_i\right)^2}{n}$$

dan

$$SS_{XY} = \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^n X_i Y_i - \frac{\left(\sum_{i=1}^n X_i\right)\left(\sum_{i=1}^n Y_i\right)}{n}$$

Jika ukuran sampel, n cukup besar, maka simpangan titik-titik koordinat dari garis regresi (s^2), yang merupakan nilai taksiran dari standart error

populasi (σ^2) dapat ditentukan dengan persamaan dari jumlah kuadrat error (SSE) :

$$s^2 = \frac{SSE}{n-2}$$

dengan $n - 2$ adalah derajat bebasnya.

Nilai SSE ditentukan dengan rumus:

$$SSE = SS_Y - b_1 SS_{XY}$$

dengan

$$SS_Y = \sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n Y_i^2 - \frac{\left(\sum_{i=1}^n Y_i\right)^2}{n}$$

Contoh 7.1.

Seorang guru SMA ingin melihat hubungan tinggi badan dan berat badan siswa yang diasuhnya. Secara umum, orang mengatakan bahwa Ada hubungan linier antara berat badan dengan tinggi badang, yang terkenal dengan rumus berat badan ideal. Untuk keperluan itu dipilih secara random sebanyak 12 orang. Orang-orang tersebut diukur masing-masing tinggi badannya dan berat badannya yang memberikan pengukuran sebagai berikut :

No	Tinggi badan (cm)	Berat badan (kg)
1.	160	60
2.	172	65
3.	155	50
4.	162	60
5.	167	63
6.	159	54
7.	172	70
8.	181	76
9.	163	60
10.	167	64
11.	165	60
12.	155	52

Dalam penelitian ini timbul beberapa masalah bagaimana persamaan matematis hubungan antara berat badan atas tinggi badan itu.

Penyelesaian:

Langkah awal yang perlu dicermati adalah penentuan variabel bebas dan variabel terikat. Berat badan merupakan variabel yang dapat berubah-ubah (naik turun), sedangkan tinggi badan tidak mungkin naik turun, sehingga yang dijadikan sebagai variabel bebas (X) adalah tinggi badan. Variabel berat badan dijadikan variabel terikat (Y). Untuk keperluan perhitungan, kita susun tabel seperti berikut ini.

No	X	Y	XY	X^2	Y^2
1	160	60	9600	25600	3600
2	172	65	11180	29584	4225
3	155	50	7750	24025	2500
4	162	60	9720	26244	3600
5	167	63	10521	27889	3969
6	159	54	8586	25281	2916
7	172	70	12040	29584	4900
8	181	76	13756	32761	5776
9	163	60	9780	26569	3600
10	167	64	10688	27889	4096
11	165	60	9900	27225	3600
12	155	52	8060	24025	2704
Jumlah	1978	734	121581	326676	45486

Dari tabel di atas, diperoleh :

$$\begin{aligned}
 n &= 12 \\
 \sum X_i &= 1978 \\
 \sum Y_i &= 734 \\
 \sum X_i Y_i &= 121581 \\
 \sum X_i^2 &= 326676 \\
 \sum Y_i^2 &= 45486
 \end{aligned}$$

Untuk menentukan nilai b_0 dan b_1 , masukkan ke dalam rumus :

$$\begin{aligned}
 \bar{X} &= \frac{\sum X_i}{n} = \frac{1978}{12} = 164,833 \\
 \bar{Y} &= \frac{\sum Y_i}{n} = \frac{734}{12} = 61,167
 \end{aligned}$$

$$b_1 = \frac{\sum X_i Y_i - \frac{(\sum X_i)(\sum Y_i)}{n}}{\sum X_i^2 - \frac{(\sum X_i)^2}{n}} = \frac{121581 - \frac{(1978)(734)}{12}}{326676 - \frac{(1978)^2}{12}} = 0,933$$

$$b_0 = \bar{Y} - b_1 \bar{X} = 61,167 - (0,933)(164,833) = -92,689$$

Jadi,

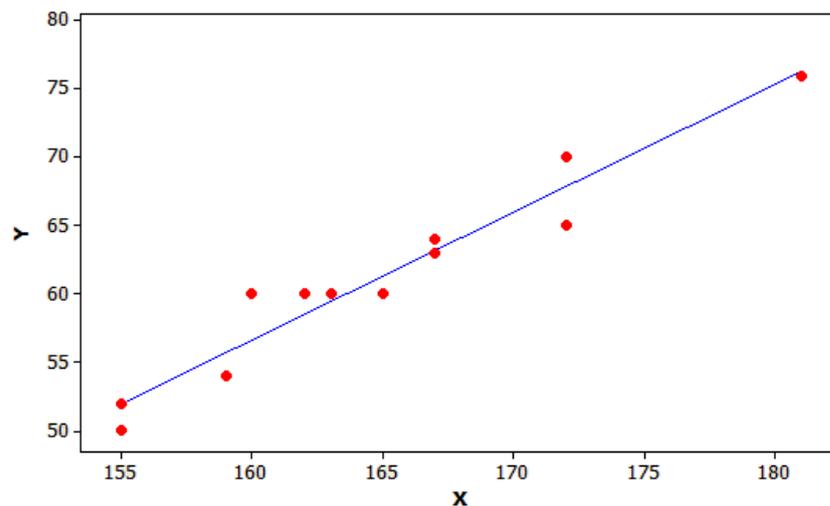
Persamaan regresi yang dapat menggambarkan hubungan antara berat badan dan tinggi badan di kelas tersebut adalah :

$$\hat{Y} = -92,689 + 0,933X$$

Intersep bernilai -92,789. Nilai intersep ini tidak memiliki arti. Jangan diartikan sebagai berikut: jika ada seorang dengan tinggi badan 0, maka berat badannya adalah -92,689. Pengertian ini tidak benar, mengapa?

Kemiringan garis regresi (slope) bernilai 0,933. Artinya jika ada dua siswa yang memiliki selisih tinggi badan 1 cm, tinggi badan yang lebih tinggi akan memiliki perkiraan berat badan 0,933 kg lebih tinggi dari temannya.

Garis regresi yang terbentuk di atas, dapat dilukiskan dalam gambar berikut beserta plotting datanya.



Nilai taksiran variansinya, dihitung dengan cara:

$$SS_Y = \sum_{i=1}^n Y_i^2 - \frac{\left(\sum_{i=1}^n Y_i\right)^2}{n} = 45486 - \frac{(734)^2}{12} = 589,667$$

$$SS_{XY} = \sum_{i=1}^n X_i Y_i - \frac{\left(\sum_{i=1}^n X_i\right)\left(\sum_{i=1}^n Y_i\right)}{n} = 121581 - \frac{(1978)(734)}{12} = 593,333$$

$$SSE = SS_Y - b_1 SS_{XY} = 589,667 - (0,933)(593,333) = 35,889$$

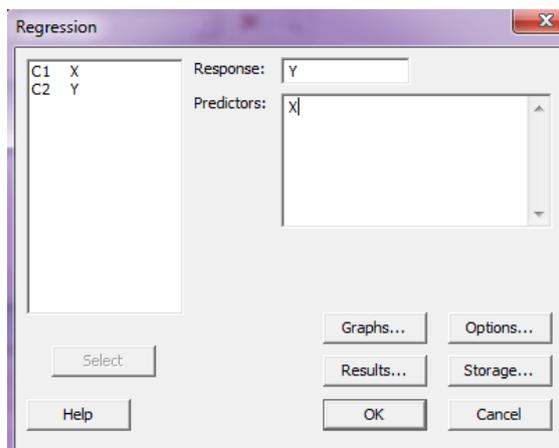
$$s^2 = \frac{SSE}{n-2} = \frac{35,889}{12-2} = 3,5889$$

atau,

$$s = \sqrt{3,5889} = 1,89334$$

Jika kita selesaikan dengan bantuan Minitab, langkah-langkah yang dilakukan adalah sebagai berikut:

- 1) Masukkan data ke dalam kolom 1 untuk X dan kolom 2 untuk Y
- 2) Pilih Stat >> Regression >> Regression
- 3) Pindahkan variable X ke kotak predictors dan Y ke Response



- 4) Klik OK, hasilnya sebagai berikut.

Regression Analysis: Y versus X

The regression equation is
 $Y = -92,7 + 0,933 X$

Predictor	Coef	SE Coef	T	P
Constant	-92,69	12,39	-7,48	0,000
X	0,93340	0,07510	12,43	0,000

S = 1,89334 R-Sq = 93,9% R-Sq(adj) = 93,3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	553,82	553,82	154,49	0,000
Residual Error	10	35,85	3,58		
Total	11	589,67			

2. Koefisien Regresi

Setelah diperoleh b_0 dan b_1 , kesimpulan mengenai persamaan regresi belum dapat kita katakan sesuai dengan kondisi data sampel. Kita harus melakukan pengujian mengenai koefisien regresi terlebih dahulu.

Menguji koefisien kemiringan garis regresi (β_1).

Langkah-langkah pengujian hipotesis koefisien regresi adalah sebagai berikut:

a. Merumuskan hipotesis

Ada tiga bentuk hipotesisnya, yaitu uji dua sisi, sisi kiri dan sisi kanan. Masing-masing adalah sebagai berikut:

- Uji dua sisi
 $H_0 : \beta_1 = 0$
 $H_1 : \beta_1 \neq 0$ artinya variabel X (prediktor) berpengaruh terhadap variabel Y (respon)
- Uji sisi kiri
 $H_0 : \beta_1 = 0$
 $H_1 : \beta_1 < 0$ artinya ada pengaruh negatif variabel X terhadap variabel Y
- Uji sisi kanan
 $H_0 : \beta_1 = 0$
 $H_1 : \beta_1 > 0$ artinya ada pengaruh positif variabel X terhadap variabel Y

b. Menghitung statistik uji

$$t_{hit} = \frac{b_1}{s_{b1}} \text{ dengan } s_{b1}^2 = \frac{s^2}{SS_X}$$

c. Menentukan kriteria uji

Karena menggunakan uji t, maka daerah kritis ditentukan dengan menggunakan tabel distribusi student t, dengan tingkat kesalahan α dan derajat bebas, $db = n - 2$. Prosedurnya sama dengan uji rata-rata jika simpangan baku populasi tidak diketahui (lihat Bab 6 Uji Hipotesis).

d. Membuat kesimpulan

Buatlah kesimpulan berdasarkan hipotesis yang disusun dan kriteria-kriteria penolakan H_0 .

Dengan langkah-langkah yang sama, kita juga dapat menguji koefisien b_0 .

Menguji koefisien intercept (β_0).

Koefisien β_0 tidak selalu harus diuji, tergantung kepada masalah yang dihadapi. Dalam masalah hubungan antara berat badan atas tinggi badan, tidak perlu menguji β_0 . Dalam masalah regresi kita **tidak selalu harus menguji β_0** , tergantung kepada masalah yang dihadapi. Jika memang harus menguji β_0 , maka langkah kerjanya serupa dengan menguji β_1 , yang beda hanya statistik ujinya.

Untuk menguji β_0 statistik uji yang digunakan adalah :

$$t_{hit} = \frac{b_0}{S_{b_0}} .$$

dimana t_{hit} berdistribusi student t dengan derajat bebas $db = n - 2$.

Contoh 7.2.

Pandang Contoh 7.1. Jika kita akan menguji apakah koefisien regresi berpengaruh, maka yang kita peroleh seperti pada tampilan output berikut:

Predictor	Coef	SE Coef	T	P
Constant	-92,69	12,39	-7,48	0,000
X	0,93340	0,07510	12,43	0,000

Dari potongan output di atas, kita dapat menyimpulkan bahwa b_0 dan b_1 masing-masing sangat mempengaruhi persamaan regresi. Koefisien b_0 berpengaruh negatif, sedangkan b_1 berpengaruh positif. Hal ini bisa dilihat dari nilai p (p-value) yang masing-masing sangat kecil, mendekati nol.

Nilai statistik uji t_{hit} diperoleh dari :

$$t_{hit(b_0)} = \frac{b_0}{s_{b_0}} = \frac{-92,69}{12,39} = -7,48 \text{ dan}$$
$$t_{hit(b_1)} = \frac{b_1}{s_{b_1}} = \frac{0,93340}{0,07510} = 12,43$$

Sedangkan nilai s_{b_0} dan s_{b_1} diperoleh dari :

$$s_{b1}^2 = \frac{s^2}{SS_X}$$

dengan

$$SS_X = \sum_{i=1}^n X_i^2 - \frac{\left(\sum_{i=1}^n X_i\right)^2}{n} = 326676 - \frac{(1978)^2}{12} = 635,6667$$

Jadi,

$$s_{b1}^2 = \frac{3,5889}{635,6667} = 0,005646$$

$$\text{Sehingga, } s_{b1} = \sqrt{0,005646} = 0,0751$$

Dengan cara yang sama, diperoleh:

$$s_{b0} = \sqrt{153,5121} = 12,39.$$

7.2. REGRESI LINIER BERGANDA

Analisis regresi linear berganda sebenarnya sama dengan analisis regresi linear sederhana, hanya variabel bebasnya lebih dari satu buah. Persamaan umumnya adalah:

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_pX_p$$

dengan Y adalah variabel bebas, dan X_i adalah variabel-variabel bebas, b_0 adalah konstanta (intersept) dengan $i = 1, 2, \dots, k$ dimana k adalah banyaknya variabel bebas, dan b_i adalah koefisien regresi pada masing-masing variabel bebas.

Apabila ada satu variabel terikat Y dan k variabel bebas X_1, X_2, \dots, X_k sehingga terdapat hubungan semacam garis regresi Y atas X_1, X_2, \dots, X_k . Dalam bagian ini akan dijelaskan secara singkat bagaimana garis regresi yang dimaksud dapat ditentukan dan yang akan ditinjau di sini hanyalah garis regresi Y atas X_1, X_2, \dots, X_k yang paling sederhana ialah yang dikenal dengan nama *regresi linear berganda*, dimana b_0, b_1, \dots, b_k harus

ditentukan dari data hasil pengamatan. Mudah dilihat bahwa regresi di atas ini merupakan perluasan dari regresi linear sederhana.

Pertanyaan yang timbul adalah: bagaimana nilai koefisien-koefisien b_0, b_1, \dots, b_k ditentukan? Secara sama dengan regresi linear sederhana, maka dipergunakan metode kuadrat terkecil. Oleh karena ada $k+1$ parameter yang harus dicari maka diperlukan $k+1$ persamaan dengan $k+1$ nilai yang akan ditentukan (b_0, b_1, \dots, b_k). Dapat dibayangkan bahwa hal itu memerlukan metode penyelesaian yang lebih baik dan karenanya memerlukan matematika yang lebih tinggi lebih-lebih untuk variabel yang cukup banyak.

Untuk regresi linear berganda yang sederhana :

$$Y = b_0 + b_1 X_1 + b_2 X_2$$

Misalnya kita harus menyelesaikan 3 persamaan dengan 3 variabel yang tidak diketahui berbentuk :

$$\begin{aligned} \sum_{i=1}^n Y_i &= n b_0 + b_1 \sum_{i=1}^n X_{1i} + b_2 \sum_{i=1}^n X_{2i} \\ \sum_{i=1}^n Y_i X_{1i} &= b_0 \sum_{i=1}^n X_{1i} + b_1 \sum_{i=1}^n X_{1i}^2 + b_2 \sum_{i=1}^n X_{1i} X_{2i} \\ \sum_{i=1}^n Y_i X_{2i} &= b_0 \sum_{i=1}^n X_{2i} + b_1 \sum_{i=1}^n X_{1i} X_{2i} + b_2 \sum_{i=1}^n X_{2i}^2 \end{aligned}$$

Untuk menentukan nilai b_0, b_1 , dan b_2 , tiga persamaan di atas dituliskan dalam bentuk matriks seperti dibawah ini.

$$\left[\begin{array}{ccc|c} n & \sum X_{1i} & \sum X_{2i} & \sum Y_i \\ \sum X_{1i} & \sum X_{1i}^2 & \sum X_{1i} X_{2i} & \sum Y_i X_{1i} \\ \sum X_{2i} & \sum X_{1i} X_{2i} & \sum X_{2i}^2 & \sum Y_i X_{2i} \end{array} \right]$$

Selanjutnya, dengan menggunakan metode Gauss, dapat diperoleh nilai-nilai b_0, b_1 , dan b_2 .

Untuk mencari nilai parameter model regresi paling mudah dicari dengan menggunakan matriks. Persamaan matriksnya dapat ditulis dalam bentuk sebagai berikut :

$$\begin{aligned} \mathbf{Y} &= \mathbf{Xb} \\ \mathbf{Xb} &= \mathbf{Y} \end{aligned}$$

dimana:

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \\ \mathbf{M} \\ \mathbf{y}_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} \mathbf{1} & \mathbf{x}_{11} & \mathbf{x}_{21} \\ \mathbf{1} & \mathbf{x}_{12} & \mathbf{x}_{22} \\ \mathbf{1} & \mathbf{x}_{13} & \mathbf{x}_{23} \\ \mathbf{M} & \mathbf{M} & \mathbf{M} \\ \mathbf{1} & \mathbf{x}_{1n} & \mathbf{x}_{2n} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} \mathbf{b}_0 \\ \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix}$$

Persamaan di atas jika kedua ruas dikalikan dengan \mathbf{X}^T akan menjadi:

$$\mathbf{X}^T \mathbf{X} \mathbf{b} = \mathbf{X}^T \mathbf{Y}$$

Sehingga \mathbf{X}^T menjadi:

$$\mathbf{X}^T = \begin{bmatrix} \mathbf{1} & \mathbf{1} & \mathbf{1} & \dots & \mathbf{1} \\ \mathbf{x}_{11} & \mathbf{x}_{12} & \mathbf{x}_{13} & \dots & \mathbf{x}_{1n} \\ \mathbf{x}_{21} & \mathbf{x}_{22} & \mathbf{x}_{23} & \dots & \mathbf{x}_{2n} \end{bmatrix}$$

diperoleh:

$$\mathbf{X}^T \mathbf{X} = \begin{bmatrix} \mathbf{n} & \sum \mathbf{x}_{i1} & \sum \mathbf{x}_{i2} \\ \sum \mathbf{x}_{i1} & \sum \mathbf{x}_{i1}^2 & \sum \mathbf{x}_{i1} \mathbf{x}_{i2} \\ \sum \mathbf{x}_{i2} & \sum \mathbf{x}_{i1} \mathbf{x}_{i2} & \sum \mathbf{x}_{i2}^2 \end{bmatrix}$$

Sehingga persamaan

$$\mathbf{X}^T \mathbf{Y} = \begin{bmatrix} \mathbf{1} & \mathbf{1} & \mathbf{1} & \dots & \mathbf{1} \\ \mathbf{x}_{11} & \mathbf{x}_{12} & \mathbf{x}_{13} & \dots & \mathbf{x}_{1n} \\ \mathbf{x}_{21} & \mathbf{x}_{22} & \mathbf{x}_{23} & \dots & \mathbf{x}_{2n} \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \\ \dots \\ \mathbf{y}_n \end{bmatrix} = \begin{bmatrix} \sum \mathbf{y}_i \\ \sum \mathbf{x}_{i1} \mathbf{y}_i \\ \sum \mathbf{x}_{i2} \mathbf{y}_i \end{bmatrix}$$

Sehingga dapat ditulis dalam bentuk matriks menjadi:

$$\mathbf{X}^T \mathbf{X} \mathbf{b} = \mathbf{X}^T \mathbf{Y}$$

$$\begin{bmatrix} \mathbf{1} & \mathbf{1} & \mathbf{1} & \dots & \mathbf{1} \\ \mathbf{x}_{11} & \mathbf{x}_{12} & \mathbf{x}_{13} & \dots & \mathbf{x}_{1n} \\ \mathbf{x}_{21} & \mathbf{x}_{22} & \mathbf{x}_{23} & \dots & \mathbf{x}_{2n} \end{bmatrix} \begin{bmatrix} \mathbf{1} & \mathbf{x}_{11} & \mathbf{x}_{21} \\ \mathbf{1} & \mathbf{x}_{12} & \mathbf{x}_{22} \\ \mathbf{1} & \mathbf{x}_{13} & \mathbf{x}_{23} \\ \dots & \dots & \dots \\ \mathbf{1} & \mathbf{x}_{1n} & \mathbf{x}_{2n} \end{bmatrix} \begin{bmatrix} \mathbf{b}_0 \\ \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{1} & \mathbf{1} & \mathbf{1} & \dots & \mathbf{1} \\ \mathbf{x}_{11} & \mathbf{x}_{12} & \mathbf{x}_{13} & \dots & \mathbf{x}_{1n} \\ \mathbf{x}_{21} & \mathbf{x}_{22} & \mathbf{x}_{23} & \dots & \mathbf{x}_{2n} \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \\ \dots \\ \mathbf{y}_n \end{bmatrix}$$

Jika $\mathbf{X}^T \mathbf{X}$ tidak singular, maka dugaan bagi β (atau dinotasikan dengan \mathbf{b}) dapat dirumuskan sebagai berikut:

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$$

Berikut tabel Anova untuk model regresi berganda:

Tabel 2.1 Analisis Variansi (ANOVA)

Sumber	Derajat Bebas	Jumlah Kuadrat	Rata-rata Jumlah Kuadrat	F hitung
Regresi	K	$b'X'Y - n\bar{Y}^2$	$\frac{b'X'Y - n\bar{Y}^2}{k} = MS_{regr}$	$\frac{MS_{regr}}{MS_{res}}$
Residual	$n - (k+1)$	$Y'Y - b'X'Y$	$\frac{Y'Y - b'X'Y}{n - k - 1} = MS_{res}$	
Total terkoreksi	$n - 1$	$Y'Y - n\bar{Y}^2$		

Uji Normalitas

Uji normalitas diperlukan untuk menguji apakah dalam sebuah model regresi, variabel tak bebas dan variabel bebas atau keduanya berdistribusi normal atau tidak. Syarat mendapatkan model regresi yang baik adalah datanya harus berdistribusi normal atau paling tidak mendekati normal. Suatu model regresi dikatakan berdistribusi normal jika model tersebut menghasilkan grafik data yang menyebar mendekati garis normal. Hipotesis yang digunakan untuk menentukan apakah suatu data observasi mengikuti distribusi normal atau tidak adalah :

Hipotesis:

H_0 : data berdistribusi normal

H_1 : data tidak berdistribusi normal

Taraf signifikansi:

Taraf signifikansi = α

Daerah kritis

H_0 ditolak apabila $p\text{-value} < \alpha$

H_0 diterima apabila $p\text{-value} > \alpha$

Selain itu, untuk mengetahui kenormalan data dapat dilakukan dengan menggunakan beberapa uji kenormalan yaitu:

- a. *Anderson-darling*
- b. *Ryan-joiner*
- c. *Kolmogorov-smirnov*

Uji Multikolinieritas

Uji multikolinieritas digunakan untuk menunjukkan adanya indikasi multikolinieritas. Multikolinieritas menunjukkan adanya hubungan atau

korelasi antara dua variabel bebas. Untuk mengetahui adanya korelasi antar dua variabel bebas dapat dilakukan dengan:

- a. Melihat nilai *Variance Inflation Factor* (VIF) dari masing-masing variabel bebas terhadap variabel tak bebas.
 1. Nilai $VIF > 10$ mengindikasikan terdapat multikolinearitas antar variabel bebas.
 2. Nilai $VIF < 10$ mengindikasikan tidak terdapat multikolinearitas antar variabel bebas.
- b. Koefisien korelasi
 1. Jika koefisien korelasi 0 - 0,30 tidak terdapat korelasi antara dua variabel bebas.
 2. Jika koefisien korelasi 0,31 - 0,60 tidak terdapat korelasi atau terdapat korelasi yang lemah antara dua variabel bebas.
 3. Jika koefisien korelasi 0,61 - 1 terdapat kuat korelasi antara dua variabel bebas.
- c. Melihat tanda parameter model regresi dengan tanda nilai koefisien korelasi antara variabel bebas dengan variabel tak bebasnya.
 1. Jika tanda parameter model regresi berlawanan dengan tanda nilai koefisien korelasi antara variabel bebas dan variabel tak bebas mengindikasikan adanya multikolinearitas.
 2. Jika tanda parameter model regresi tidak berlawanan (sama) dengan tanda nilai koefisien korelasi antara variabel bebas dan variabel tak bebas mengindikasikan tidak adanya multikolinearitas.

Pemilihan Model Terbaik

Salah satu tujuan dalam analisis regresi adalah untuk mendapatkan model terbaik yang menjelaskan hubungan antara variabel bebas dengan variabel tak bebas. Model terbaik adalah model yang seluruh koefisien regresinya berarti (*significant*) dan mempunyai kriteria model terbaik optimum. Beberapa kriteria model terbaik adalah :

Tabel 2.2 Kriteria Model Terbaik

No	Kriteria	Formula	Optimum
1	SSE	$\sum(Y - \hat{Y})^2$	Minimum
2	MSE	$\sum(Y - \hat{Y})^2 / (n - p - 1)$	Minimum
3	R^2	$R^2 = \frac{\sum(\hat{Y} - \bar{Y})^2}{\sum(Y - \bar{Y})^2} 100\%$	Maksimum

4	Adjusted R ²	$1 - [1 - R^2] \frac{(n-1)}{(n-p)}$	Maksimum
5	Cp Mallow	$\frac{SSE}{MSE} - (n - 2p)$	Minimum

Untuk memperoleh model terbaik ada beberapa metode yang biasa digunakan antara lain yaitu :

2.1.1 *Stepwise*

Pemilihan model terbaik dengan metode *Stepwise* adalah dengan menambahkan dan membuang variabel bebas ke dalam model regresi. Metode *Stepwise* dibagi menjadi dua yaitu:

a. *Forward*

Pembuatan model regresi terbaik dilakukan dengan menambahkan variabel satu per satu. Tahap awal dengan memasukkan satu variabel bebas. Tahap selanjutnya menambahkan variabel bebas baru sehingga ada dua variabel bebas dalam model. Penambahan diulangi sampai semua variabel bebas masuk ke dalam model.

b. *Backward*

Pembuatan model regresi terbaik dilakukan dengan membuat terlebih dahulu model regresi untuk semua variabel bebas. Selanjutnya dikurangi variabel satu per satu sampai tinggal satu variabel bebas.

2.1.2 *Best Subset Regression*

Pemilihan model terbaik berdasarkan *Best Subset Regression* adalah dengan menyajikan k buah model regresi terbaik dengan 1, 2, ..., p variabel bebas. Kemudian diantara k model regresi tersebut dicari yang optimum.

Analisis regresi linear berganda memerlukan pengujian secara serempak dengan menggunakan F hitung. Signifikansi ditentukan dengan membandingkan F hitung dengan F tabel atau melihat signifikansi pada output Minitab. Dalam beberapa kasus dapat terjadi bahwa secara simultan (serempak) beberapa variabel mempunyai pengaruh yang signifikan, tetapi secara parsial tidak.

Penggunaan metode analisis regresi linear berganda memerlukan asumsi klasik yang secara statistik harus dipenuhi. Asumsi klasik tersebut meliputi asumsi normalitas, multikolinearitas, autokorelasi, heteroskedastisitas dan asumsi linearitas (akan dibahas belakangan).

Langkah-langkah yang lazim dipergunakan dalam analisis regresi linear berganda adalah 1) koefisien determinasi; 2) Uji F dan 3) uji t . Persamaan regresi sebaiknya dilakukan di akhir analisis karena interpretasi terhadap persamaan regresi akan lebih akurat jika telah diketahui signifikansinya.

Koefisien determinasi sebaiknya menggunakan *adjusted R Square* dan jika bernilai negatif maka uji F dan uji t tidak dapat dilakukan.

Contoh

Seorang guru mengamati nilai hasil belajar siswa (Y), lama belajar, X₁, (menit), dan banyaknya soal uraian yang dikerjakan (X₂). Dari 10 siswa yang diamati, hasilnya ditampilkan dalam table berikut ini.

No.	Y	X ₁	X ₂
1	65	100	3
2	70	110	4
3	75	120	4
4	80	120	5
5	90	130	5
6	95	125	7
7	100	90	6
8	80	110	6
9	75	95	4
10	85	125	5

Bagaimana model hubungan antara lama belajar dan banyaknya soal terhadap skor hasil ujian siswa?

Untuk membuat model regresi dari permasalahan di atas, kita dapat menyusun tabel seperti di bawah ini.

No.	Y	X ₁	X ₂	Y ²	X ₁ ²	X ₂ ²	Y*X ₁	Y*X ₂	X ₁ *X ₂
1	65	100	3	4225	10000	9	6500	195	300
2	70	110	4	4900	12100	16	7700	280	440
3	75	120	4	5625	14400	16	9000	300	480
4	80	120	5	6400	14400	25	9600	400	600
5	90	130	5	8100	16900	25	11700	450	650
6	95	125	7	9025	15625	49	11875	665	875
7	100	90	6	10000	8100	36	9000	600	540
8	80	110	6	6400	12100	36	8800	480	660
9	75	95	4	5625	9025	16	7125	300	380
10	85	125	5	7225	15625	25	10625	425	625
Jumlah	815	1125	49	67525	128275	253	91925	4095	5550

Dari tabel, dapat diringkaskan hasil sebagai berikut:

$$\sum Y_i = 815, \quad \sum X_{1i} = 1125, \quad \sum X_{2i} = 49, \quad \sum Y_i^2 = 67525, \\ \sum X_{1i}^2 = 128275, \quad \sum X_{2i}^2 = 253, \quad \sum Y_i X_{1i} = 91925, \quad \sum Y_i X_{2i} = 4095 \text{ dan} \\ \sum X_{1i} X_{2i} = 5550.$$

Sehingga, persamaan normal yang dapat dibentuk adalah:

1. $\sum_{i=1}^n Y_i = nb_0 + b_1 \sum_{i=1}^n X_{1i} + b_2 \sum_{i=1}^n X_{2i}$
 $815 = 10b_0 + 1125b_1 + 49b_2$
2. $\sum_{i=1}^n Y_i X_{1i} = b_0 \sum_{i=1}^n X_{1i} + b_1 \sum_{i=1}^n X_{1i}^2 + b_2 \sum_{i=1}^n X_{1i} X_{2i}$
 $91925 = 1125b_0 + 128275b_1 + 5550b_2$
3. $\sum_{i=1}^n Y_i X_{2i} = b_0 \sum_{i=1}^n X_{2i} + b_1 \sum_{i=1}^n X_{1i} X_{2i} + b_2 \sum_{i=1}^n X_{2i}^2$
 $4095 = 49b_0 + 5550b_1 + 253b_2$

Selanjutnya, matrik yang dapat dibentuk adalah:

$$\left(\begin{array}{ccc|c} 10 & 1125 & 49 & 815 \\ 1125 & 128275 & 5550 & 91925 \\ 49 & 5550 & 253 & 4095 \end{array} \right)$$

Dengan menggunakan metode Gauss, diperoleh hasil sebagai berikut:
 Bagilah baris pertama dengan 10,

$$\begin{array}{cccc} 1 & 112,5 & 4,9 & 81,5 \\ 1125 & 128275 & 5550 & 91925 \\ 49 & 5550 & 253 & 4095 \end{array}$$

Baris kedua dikurangi 1125 kali baris pertama dan baris ketiga dikurangi 49 kali baris pertama.

$$\begin{array}{cccc} 1 & 112,5 & 4,9 & 81,5 \\ 0 & 1712,5 & 37,5 & 237,5 \\ 0 & 37,5 & 12,9 & 101,5 \end{array}$$

Baris kedua dibagi 1712,5.

$$\begin{array}{cccc} 1 & 112,5 & 4,9 & 81,5 \\ 0 & 1 & 0,021898 & 0,138686 \\ 0 & 37,5 & 12,9 & 101,5 \end{array}$$

Baris ketiga dikurangi 37,5 kali baris kedua.

1	112,5	4,9	81,5
0	1	0,021898	0,138686
0	0	12,07883	96,29927

Baris ketiga dibagi 12,07883.

1	112,5	4,9	81,5
0	1	0,021898	0,138686
0	0	1	7,972566

Diperoleh:

$$b_2 = 7,972566$$

$$b_1 + 0,021898b_2 = 0,138686, \text{ sehingga}$$

$$b_1 = 0,138686 - 0,021898(7,972566) = -0,036$$

$$b_0 = 81,5 - (4,9b_2 + 112,5b_1) = 46,473$$

Persamaan regresi yang terbentuk adalah :

$$Y = 46,473 - 0,036X_1 + 7,973X_2$$

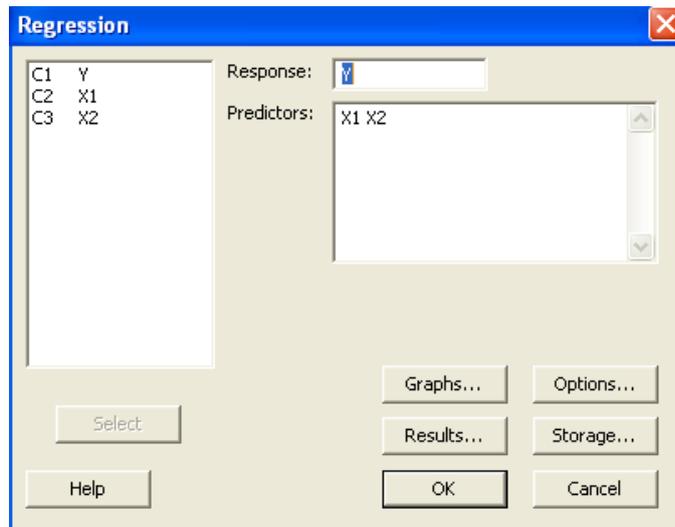
Jika dikerjakan dengan Minitab, langkah-langkahnya adalah sebagai berikut:

Masukkan data pada sheet seperti berikut ini:

↓	C1	C2	C3
	Y	X1	X2
1	65	100	3
2	70	110	4
3	75	120	4
4	80	120	5
5	90	130	5
6	95	125	7
7	100	90	6
8	80	110	6
9	75	95	4
10	85	125	5

Pilih Stat >> Regression >> Regression

Masukkan variable Y ke kotak respond an X1 dan X2 ke kotak predictors



Klik OK. Hasilnya adalah sebagai berikut.

Regression Analysis: Y versus X1, X2

The regression equation is
 $Y = 46.5 - 0.036 X1 + 7.97 X2$

Persamaan regresi yang terbentuk adalah seperti output di atas.

Predictor	Coef	SE Coef	T	P
Constant	46.47	18.55	2.51	0.041
X1	-0.0359	0.1640	-0.22	0.833
X2	7.973	1.889	4.22	0.004

S = 6.56625 R-Sq = 72.6% R-Sq(adj) = 64.8%

Potongan output di atas merupakan uji koefisien regresi, dapat diartikan sebagai berikut:

- Jika diambil kesalahan 5%, nilai intersep 46,47 berpengaruh terhadap model regresi. Kita tidak perlu mengartikan nilai intersep tersebut.
- Variabel X_1 ternyata tidak berpengaruh nyata terhadap model. Hal ini ditunjukkan dengan besarnya nilai p yang melebihi 5%.
- Sedangkan variabel X_2 secara nyata berpengaruh terhadap model, karena nilai p kurang dari 5%.

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	800.69	400.35	9.29	0.011
Residual Error	7	301.81	43.12		
Total	9	1102.50			

Secara serempak, model regresi yang dihasilkan dapat diterima, karena nilai p yang dihasilkan kurang dari 5%.

8

Analisis Korelasi

Persamaan regresi melukiskan hubungan fungsional antar variabel terikat dengan variabel bebas. Dalam model persamaan regresi terlihat bagaimana pengujian terhadap parameter-parameter dilakukan. Persoalan berikutnya yang dirasakan perlu adalah mengetahui berapa kuat hubungan antara variabel-variabel itu terjadi.

Bila bentuk hubungan antar variabel X dengan variabel Y sudah dapat kita terima maka kita ingin pula mengetahui seberapa besar keeratan hubungannya. Walaupun bentuk hubungan antara variabel X dengan variabel Y ada dalam bentuk yang benar belum tentu korelasinya besar karena banyak variabel lain yang turut mempengaruhi perubahan variabel Y. Besarnya perubahan variabel Y yang dapat diterangkan oleh variabel X dengan menggunakan persamaan garis regresi yang diperoleh disebut koefisien determinasi.

Studi yang membahas tentang derajat hubungan antara variabel-variabel dikenal dengan nama analisis korelasi. Ukuran yang dipakai untuk mengetahui derajat hubungan, terutama untuk data kuantitatif, dinamakan *koefisien korelasi*.

Perlu diketahui bahwa: dalam analisis regresi, kita anggap bahwa variabel bebas X adalah konstan, jadi bukan suatu variabel acak. Dalam analisis korelasi, variabel X dan Y keduanya merupakan variabel acak yang menyebar bersama.

8.1. KORELASI

Dalam masalah regresi kita bisa menjawab pertanyaan-pertanyaan : bagaimana bentuk hubungan (pola hubungan) antara Y atas X, dan bagaimana persamaan matematis untuk hubungan tersebut, berapa besar perubahan rata-rata Y apabila X berubah 1 unit, berapa batas-batas harga rata-rata Y apabila $X = X_0$, berapa batas-batas harga Y apabila $X = X_0$. Tetapi apabila ada pertanyaan : kalau memang ada hubungan antara Y dengan X, berapa besarnya **keeratan hubungan** tersebut, atau dengan perkataan lain kalau memang ada hubungan antara Y atas X berapa **kekuatan X untuk bisa menerangkan Y**?

Pertanyaan ini tidak bisa dijawab oleh analisis regresi yang bisa menjawab pertanyaan ini adalah **analisis korelasi**.

Apabila garis regresi yang terbaik untuk sekumpulan data berbentuk linier, maka derajat hubungannya akan dinyatakan dengan r dan biasa dinamakan *koefisien korelasi*. Apabila antara 2 variabel Y dan X ada hubungan linier korelatif, maka eratnya hubungan tersebut diperlihatkan oleh koefisien korelasi yang didefinisikan sebagai :

$$r = \frac{\sum X_i Y_i - \frac{(\sum X_i)(\sum Y_i)}{n}}{\sqrt{\left[\sum X_i^2 - \frac{(\sum X_i)^2}{n} \right] \left[\sum Y_i^2 - \frac{(\sum Y_i)^2}{n} \right]}}$$

Koefisien korelasi ini dihitung dari sampel (statistik) sebagai estimator untuk koefisien korelasi dalam populasi yang dilambangkan oleh ρ (rho).

Sifat – sifat statistik koefisien korelasi adalah :

- 1) Harganya terbatas yaitu $-1 \leq \rho \leq +1$,
 $\rho = 0$, artinya tidak ada hubungan linier antara X dan Y .
 Jadi, jika harga ρ makin mendekati $+1$, artinya hubungan yang positif itu makin erat, jika semakin mendekati -1 artinya hubungan itu makin erat juga tetapi hubungan negatif.
 Contoh : $\rho = 0,8$ dan $\rho = -0,8$. Kedua ρ tersebut menunjukkan hubungan yang sama eratnya, bedanya yang satu hubungan positif dan yang lain hubungan negatif.

2) Koefisien korelasi ρ **tidak mempunyai satuan**

Untuk memudahkan melakukan interpretasi mengenai kekuatan hubungan antara dua variabel penulis memberikan kriteria sebagai berikut (Sarwono,2006):

- 1) 0 : Tidak ada korelasi antara dua **variable**
- 2) 0 – 0,25: Korelasi sangat lemah
- 3) 0,25 – 0,5: Korelasi cukup
- 4) 0,5 – 0,75: Korelasi kuat
- 5) 0,75 – 0,99: Korelasi sangat kuat
- 6) 1: Korelasi sempurna

Jika nilai koefisien korelasi dikuadratkan, menjadi r^2 , maka r^2 dinamakan *koefisien determinasi* atau *koefisien penentu*. Dinamakan demikian oleh karena $100 r^2\%$ daripada variasi yang terjadi dalam variabel tak bebas Y dapat dijelaskan oleh variabel bebas X dengan adanya regresi linier Y atas

X. Harga $\sqrt{1-r^2}$ dinamakan *koefisien alienasi* atau *koefisien perenggangan*. Harga $1 - r^2$ sendiri dapat dinamakan *koefisien non determinasi*.

Koefisien korelasi r tentu saja didapat dengan jalan mengambil akar dari r^2 . Mudah dilihat bahwa akan berlaku $0 \leq r^2 \leq 1$ sehingga untuk koefisien korelasi didapat hubungan $-1 \leq r \leq +1$. Harga $r = -1$ menyatakan adanya *hubungan linier sempurna tak langsung* antara X dan Y. Ini berarti bahwa titik-titik yang ditentukan oleh (X_i, Y_i) seluruhnya terletak pada garis regresi linier dan harga X yang besar menyebabkan atau berpasangan dengan Y yang kecil sedangkan harga X yang kecil berpasangan dengan Y yang besar. Harga $r = +1$ menyatakan adanya *hubungan linier sempurna langsung* antara X dan Y. Letak titik-titik ada pada garis regresi linier dengan sifat bahwa harga X yang besar berpasangan dengan harga Y yang besar, sedangkan harga X yang kecil berpasangan dengan Y yang kecil pula.

Harga-harga r lainnya bergerak antara -1 dan $+1$ dengan tanda negatif menyatakan adanya *korelasi tak langsung* atau korelasi negatif dan tanda positif menyatakan *korelasi langsung* atau korelasi positif. Khusus untuk $r = 0$, maka hendaknya ini ditafsirkan bahwa *tidak terdapat hubungan linier* antara variabel-variabel X dan Y.

Bentuk-bentuk rumus koefisien korelasi yang lain adalah:

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

Bentuk lain dapat pula digunakan, ialah :

$$r = \sqrt{1 - s_{y.x}^2 / s_y^2}$$

dengan $s_{yx}^2 = \sum (Y_i - \bar{Y})(X_i - \bar{X})$ yang merupakan kuadrat kekeliruan taksiran dan $s_y^2 = \sum (Y_i - \bar{Y})^2$ yang merupakan ragam untuk variabel Y.

Contoh 8.1.

Pandang Contoh 7.1. Misalkan akan ditentukan apakah ada hubungan linier antara berat badan dengan tinggi badan. Data dari contoh, dituliskan ulang sebagai berikut:

No	X	Y	XY	X ²	Y ²
1	160	60	9600	25600	3600
2	172	65	11180	29584	4225
3	155	50	7750	24025	2500
4	162	60	9720	26244	3600
5	167	63	10521	27889	3969
6	159	54	8586	25281	2916
7	172	70	12040	29584	4900
8	181	76	13756	32761	5776
9	163	60	9780	26569	3600
10	167	64	10688	27889	4096
11	165	60	9900	27225	3600
12	155	52	8060	24025	2704
Jumlah	1978	734	121581	326676	45486

Dari tabel di atas, diperoleh :

$$\begin{aligned}
 n &= 12 \\
 \sum X_i &= 1978 \\
 \sum Y_i &= 734 \\
 \sum X_i Y_i &= 121581 \\
 \sum X_i^2 &= 326676 \\
 \sum Y_i^2 &= 45486
 \end{aligned}$$

Masukkan semua informasi yang diperlukan dalam rumus:

$$\begin{aligned}
 r &= \frac{\sum X_i Y_i - \frac{(\sum X_i)(\sum Y_i)}{n}}{\sqrt{\left[\sum X_i^2 - \frac{(\sum X_i)^2}{n} \right] \left[\sum Y_i^2 - \frac{(\sum Y_i)^2}{n} \right]}} \\
 r &= \frac{121581 - \frac{(1978)(734)}{12}}{\sqrt{\left[326676 - \frac{(1978)^2}{12} \right] \left[45486 - \frac{(734)^2}{12} \right]}} = 0,969
 \end{aligned}$$

Dari hasil ini ternyata didapat korelasi positif yang sangat kuat antara Tinggi Badan (X) dan Berat Badan (Y). Hal ini berarti, meningkatnya tinggi badan akan meningkatkan pula berat badan. Besar hubungannya

ditentukan oleh koefisien determinasi $r^2 = 0,939$ atau sebesar 93,97%. Ini berarti bahwa perubahan Berat Badan 93,9% dapat dijelaskan oleh Tinggi Badan melalui hubungan linier X dan Y, sedangkan sisanya 6,1% ditentukan oleh keadaan lain.

Jika kita menggunakan Minitab, langkah-langkah yang ditempuh sama dengan pada saat kita menghitung regresi. Output yang diperoleh adalah sebagai berikut:

Regression Analysis: Y versus X

The regression equation is
 $Y = -92,7 + 0,933 X$

Predictor	Coef	SE Coef	T	P
Constant	-92,69	12,39	-7,48	0,000
X	0,93340	0,07510	12,43	0,000

S = 1,89334 R-Sq = 93,9% R-Sq(adj) = 93,3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	553,82	553,82	154,49	0,000
Residual Error	10	35,85	3,58		
Total	11	589,67			

Dari output terlihat bahwa nilai $r^2 = 93,9\%$,sama seperti perhitungan manual kita di atas.

8.2. UJI HIPOTESIS KOEFISIEN KORELASI

Misalkan Suatu populasi yang terdiri dari dua variabel memiliki koefisien korelasi ρ . Jika $\rho = 0$, maka X dan Y independen, artinya kedua variabel saling bebas, tidak berhubungan sama sekali. Sehingga, untuk populasi berdistribusi normal $\rho = 0$ mengakibatkan bahwa X dan Y independen dan sebaliknya. Sifat ini tidak berlaku untuk populasi yang tidak berdistribusi normal.

Mengingat dalam banyak penelitian sering ingin mengetahui apakah antara dua variabel terdapat hubungan yang independen atau tidak, maka kita perlu melakukan uji independen. Dalam hal ini, maka hipotesis yang harus diujikan adalah :

$$H_0 : \rho = 0 \text{ melawan } H_1 : \rho \neq 0$$

Uji ini sebenarnya ekuivalen dengan uji $H_0 : \beta_1 = 0$ dimana β_1 menyatakan koefisien arah regresi linier untuk populasi. Untuk menguji $H_0 : \rho = 0$ melawan $H_1 : \rho \neq 0$, jika sampel acak yang diambil dari populasi normal bivariabel dua itu berukuran n memiliki koefisien korelasi r , maka dapat digunakan statistik uji t yaitu :

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

Selanjutnya untuk taraf nyata $= \alpha$, maka hipotesis kita terima jika $-t_{(1-\frac{1}{2}\alpha)} < t < t_{(1-\frac{1}{2}\alpha)}$, dimana distribusi t yang digunakan mempunyai $dk = (n-2)$. Dalam hal lainnya H_0 kita tolak.

Tentu saja bentuk alternatif untuk menguji hipotesis H_0 bisa $H_1 : \rho > 0$ atau $H_1 : \rho < 0$. Dalam hal pertama merupakan uji pihak kanan sedangkan yang kedua merupakan uji pihak kiri. Daerah kritis pengujian, seperti biasa harus disesuaikan dengan alternatif yang diambil.

Contoh 8.2.

Lihat Contoh 8.1. diatas tentang hubungan Berat Badan dengan Tinggi Badan.

1. Pemilihan hipotesis alternatif, H_1 berdasarkan kasus yang diteliti. Dalam contoh hubungan Berat Badan dan Tinggi Badan, kita menduga bahwa ada hubungan linier yang positif antara keduanya, sehingga hipotesis yang kita susun adalah:

$$H_0 : \rho = 0 \text{ vs} \\ H_1 : \rho > 0$$

Sebenarnya, pemilihan rumusan H_1 berdasarkan informasi penelitian. Jika memang hubungan belum diketahui atau belum diduga, kita bias menggunakan uji dua sisi, yang berarti $H_1: \rho \neq 0$. Sebaliknya, jika hubungan tadi sudah diduga, kita dapat menggunakan uji satu sisi, sisi kiri jika hubungannya negatif dan sisi kanan jika hubungannya positif.

2. Tingkat kesalahan $\alpha = 0,05$
3. Statistik uji yang digunakan untuk menguji $H_0 : \rho = 0$

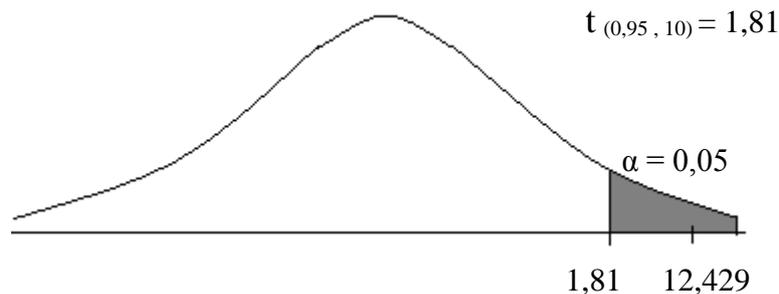
$$t_{hit} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}, \text{ dengan derajat bebas, } dk = n-2.$$

Kalau analisisnya **mulai dari regresi**, dimana korelasi hanya merupakan analisis lanjutan, setelah menghitung r tidak perlu menguji ρ karena pada regresi kita telah menguji β_1 . Alasannya adalah :

$$t = \frac{b_1}{S_{b_1}} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

$$t = \frac{0,969\sqrt{12-2}}{\sqrt{1-(0,969)^2}} = 12,429$$

4. Daerah dan titik kritis



Hasil hitungan uji statistik jatuh di daerah penolakan. Pengujian secara signifikan harus menolak significant H_0 .

5. Kesimpulan

Ada hubungan positif linier antara berat badan atas tinggi badan.

8.3. UJI RANK KORELASI SPEARMAN

Uji rank korelasi Spearman sebenarnya masuk dalam pembahasan statistika nonparametrika. Hal ini terjadi karena data yang digunakan bukan berupa skor atau nilai yang asli (skala rasio), tetapi merupakan selisih peringkat dari masing-masing pasangan subyek pengamatan (skala ordinal). Selisih peringkat yang dimaksud, dirumuskan dalam bentuk:

$$D_i = X_{(i)} - Y_{(i)}$$

dengan $X_{(i)}$ adalah peringkat data ke-i.

Sebagai contoh, misalnya ingin diteliti kesesuaian dua dosen dalam memberikan nilai kepada 20 mahasiswanya. Masing-masing dosen diminta untuk membuat penilaian yang obyektif terhadap 20 mahasiswa tadi. Kemudian dari semua mahasiswa akan diranking oleh masing-masing dosen, dari peringkat 1 sampai 20. Apabila ada peserta dengan nilai yang sama, maka peringkatnya dibagi sama. Sebagai contoh, ada 2 mahasiswa mendapat nilai sama pada peringkat 3, maka masing-masing mahasiswa diberi peringkat 3,5 yang merupakan hasil dari penghitungan: $3 + 4$ kemudian dibagi 2.

Berdasarkan pemahaman di atas, maka jelas sekali bahwa uji Spearman hanya diperuntukkan bagi uji dengan 2 subjek yang berbeda atau disebut juga sampel bebas.

Konsep dan interpretasi nilai Korelasi Spearman sama dengan konsep koefisien Korelasi pada Regresi (Linier Sederhana).

Langkah-langkah untuk menghitung korelasi Spearman adalah sebagai berikut:

- 1) Susunlah pasangan data X_i dan Y_i dalam sebuah tabel. Urutkan berdasarkan penomoran sampel.
- 2) Berikanlah peringkat (ranking) untuk masing-masing data, baik dari X_i maupun Y_i dengan cara memberikan peringkat 1 untuk data tertinggi dan seterusnya. Apabila terdapat data yang memiliki peringkat yang sama, maka peringkat data tersebut adalah rata-rata dari peringkat yang sama. Misalkan pada urutan data ke 1, 2, dan 3 memiliki nilai pengamatan yang sama, maka skor peringkatnya dihitung dari $\frac{1+2+3}{3} = 2$.
- 3) Hitung D_i , yaitu selisih peringkat data ke- i untuk variabel X_i dan Y_i .
- 4) Hitung nilai koefisien korelasi Spearman dengan rumus:

$$r_{xy} = 1 - \frac{6 \sum_{i=1}^n D_i^2}{n(n^2 - 1)}$$

Contoh 8.3.

Dua orang dosen sedang menguji 20 mahasiswa yang mengambil kuliah Kapita Selekt. Nilai selengkapnya seperti dalam tabel berikut ini.

No	Nama Mahasiswa	Dosen A	Dosen B
1	AGUS SUKOYO	85	80
2	AZWAR AMIRUDIN	70	60
3	DWI ASTUTI	75	75
4	RATNA URSILA	50	60
5	FAIKOTU SYARIFAH	55	50
6	YANGKI ADISTIYAN	80	85
7	YOPI SULISTIONO	83	79
8	WINDRIYANI	68	70
9	CITRA INNEKE	79	81
10	NOVI AMELIA	85	80
11	KAMILAH DWI SEPTIANI	90	90
12	SITI MASITOH	95	90
13	HANIFAH	88	84

14	SALSABILA	83	81
15	ATIKA SURYANINGSIH	55	50
16	TITIN ATINAH	64	68
17	AYU ROSELINA	75	80
18	RETNO SUSILOWATY	78	74
19	RURI NUR SASWURI	81	88
20	EVIALADO RAHULINA S	83	79

Berdasarkan Dosen A, nilai mahasiswa dapat diurutkan berdasarkan skor nilainya sebagai berikut:

No sampel	Nama Mahasiswa	Peringkat
12	SITI MASITOH	1
11	KAMILAH DWI SEPTIANI	2
13	HANIFAH	3
1	AGUS SUKOYO	4
10	NOVI AMELIA	5
7	YOPI SULISTIONO	6
14	SALSABILA	7
20	EVIALADO RAHULINA S	8
19	RURI NUR SASWURI	9
6	YANGKI ADISTIYAN	10
9	CITRA INNEKE	11
18	RETNO SUSILOWATY	12
3	DWI ASTUTI	13
17	AYU ROSELINA	14
2	AZWAR AMIRUDIN	15
8	WINDRIYANI	16
16	TITIN ATINAH	17
5	FAIKOTU SYARIFAH	18
15	ATIKA SURYANINGSIH	19
4	RATNA URSILA	20

Terlihat bahwa ada beberapa mahasiswa yang memiliki nilai yang sama, sehingga mereka seharusnya memiliki skor peringkat yang sama. Selanjutnya kita susun berdasarkan skor peringkat dan dikembalikan ke dalam urutan sampel yang semula.

Tabel skor dari Dosen A

No sampel	Nama Mahasiswa	Skor Peringkat
1	AGUS SUKOYO	4,5

2	AZWAR AMIRUDIN	15
3	DWI ASTUTI	13,5
4	RATNA URSILA	20
5	FAIKOTU SYARIFAH	18,5
6	YANGKI ADISTIYAN	10
7	YOPI SULISTIONO	7
8	WINDRIYANI	16
9	CITRA INNEKE	11
10	NOVI AMELIA	4,5
11	KAMILAH DWI SEPTIANI	2
12	SITI MASITOH	1
13	HANIFAH	3
14	SALSABILA	7
15	ATIKA SURYANINGSIH	18,5
16	TITIN ATINAH	17
17	AYU ROSELINA	13,5
18	RETNO SUSILOWATY	12
19	RURI NUR SASWURI	9
20	EVIALADO RAHULINA S	7

Dengan cara yang sama, kita urutkan penilaian dari Dosen B. Hasilnya sebagai berikut.

Tabel skor Dosen B

No sampel	Nama Mahasiswa	Skor Peringkat
1	AGUS SUKOYO	8,5
2	AZWAR AMIRUDIN	17,5
3	DWI ASTUTI	13
4	RATNA URSILA	17,5
5	FAIKOTU SYARIFAH	19,5
6	YANGKI ADISTIYAN	4
7	YOPI SULISTIONO	11,5
8	WINDRIYANI	15
9	CITRA INNEKE	6,5
10	NOVI AMELIA	8,5
11	KAMILAH DWI SEPTIANI	1,5
12	SITI MASITOH	1,5
13	HANIFAH	5
14	SALSABILA	6,5
15	ATIKA SURYANINGSIH	19,5
16	TITIN ATINAH	16

17	AYU ROSELINA	10
18	RETNO SUSILOWATY	14
19	RURI NUR SASWURI	3
20	EVIALADO RAHULINA S	11,5

Gabungan skor dari kedua dosen, selisih skor peringkat, dan kuadrat selisih peringkat adalah sebagai berikut:

No	Nama Mahasiswa	Dosen A	Dosen B	Di	D_i^2
1	AGUS SUKOYO	4,5	8,5	-4	16
2	AZWAR AMIRUDIN	15	17,5	-2,5	6,25
3	DWI ASTUTI	13,5	13	0,5	0,25
4	RATNA URSILA	20	17,5	2,5	6,25
5	FAIKOTU SYARIFAH	18,5	19,5	-1	1
6	YANGKI ADISTIYAN	10	4	6	36
7	YOPI SULISTIONO	7	11,5	-4,5	20,25
8	WINDRIYANI	16	15	1	1
9	CITRA INNEKE	11	6,5	4,5	20,25
10	NOVI AMELIA	4,5	8,5	-4	16
11	KAMILAH DWI SEPTIANI	2	1,5	0,5	0,25
12	SITI MASITOH	1	1,5	-0,5	0,25
13	HANIFAH	3	5	-2	4
14	SALSABILA	7	6,5	0,5	0,25
15	ATIKA SURYANINGSIH	18,5	19,5	-1	1
16	TITIN ATINAH	17	16	1	1
17	AYU ROSELINA	13,5	10	3,5	12,25
18	RETNO SUSILOWATY	12	14	-2	4
19	RURI NUR SASWURI	9	3	6	36
20	EVIALADO RAHULINA S	7	11,5	-4,5	20,25
JUMLAH					202,5

Masukkan ke dalam rumus, diperoleh nilai koefisien korelasi Spearmannya adalah:

$$r_{xy} = 1 - \frac{6 \sum_{i=1}^n D_i^2}{n(n^2 - 1)} = 1 - \frac{(6)(202,5)}{(20)(20^2 - 1)} = 1 - 0,8477 = 0,15226$$

Jika kita akan melihat apakah ada hubungan penilaian antara Dosen A dengan Dosen B, kita dapat melakukan uji hipotesis sebagai berikut:

1) Hipotesis

$H_0 : \rho = 0$ (tidak ada hubungan antara Dosen A dengan Dosen B dalam memberikan nilai kepada mahasiswanya), terhadap :

$H_1 : \rho \neq 0$ (ada hubungan antara Dosen A dengan Dosen B dalam memberikan nilai kepada mahasiswanya)

2) Tingkat kesalahan

Gunakan $\alpha = 0,05$

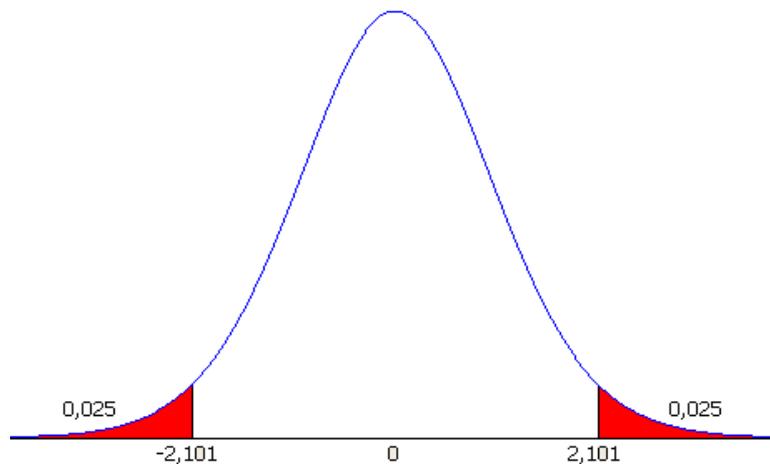
3) Statistik uji

Statistik uji yang digunakan adalah $t_{hit} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$

$$t_{hit} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0,15226\sqrt{20-2}}{\sqrt{1-0,15226^2}} = 0,6536$$

4) Daerah kritis

Karena merupakan uji dua sisi, maka daerah kritisnya untuk derajat bebas, $dk = n - 2 = 20 - 2 = 18$ adalah seperti berikut:



H_0 ditolak jika $t_{hit} < -t\left(\frac{\alpha}{2}; n-2\right)$ atau jika $t_{hit} > t\left(\frac{\alpha}{2}; n-2\right)$. Karena

$-t\left(\frac{\alpha}{2}; n-2\right) < t_{hit} < t\left(\frac{\alpha}{2}; n-2\right)$, yaitu $-2,101 < 0,6536 < 2,101$, maka

H_0 diterima.

5) Kesimpulan

Dosen A dan Dosen B dalam memberikan nilai kepada mahasiswanya tidak saling berhubungan.

Prosedur yang harus dilakukan dalam melakukan uji korelasi Spearman menggunakan Minitab adalah sebagai berikut:

- 1) Pemasukan data ke dalam lembar worksheet minitab dimulai dari kolom pertama, dimana kata Dosen A diketikkan pada kolom C1 dan Dosen B pada kolom C2.
- 2) Menu Data diklik, Rank disorot dan diklik hingga kotak dialog Rank ditampilkan. Setelah itu, pada kotak dialog Rank, kolom Rank Data diklik, C2 disorot, tombol Select diklik, kolom Store Rank diklik, C4 diketik, dan kemudian diklik OK
- 3) Menu Data diklik, Rank disorot dan diklik hingga kotak dialog Rank ditampilkan. Setelah itu, pada kotak dialog Rank, kolom Rank Data diklik, C3 disorot, tombol Select diklik, kolom Store Rank diklik, C5 diketik, dan kemudian diklik OK
- 4) Worksheet pada minitab diperhatikan, terdapat data rank pada C4 dan C5
- 5) Menu Calc diklik, calculator disorot dan diklik, C6 diketik pada bagian Store Result in Variable, kolom Expression diklik, C4 disorot, tombol Select diklik, tanda – diklik, C5 disorot, tombol Select diklik, dan kemudian diklik OK
- 6) Menu Calc diklik, calculator disorot dan diklik, C7 diketik pada bagian Store Result in Variable, kolom Expression diklik, C6 disorot, tombol Select diklik, tanda * diklik, C6 disorot, tombol Select diklik, dan kemudian diklik OK
- 7) Menu Calc diklik, Column Statistic disorot dan diklik, tombol Sum diklik pada kolom Statistics, C7 disorot pada kolom Input Variable, tombol Select diklik, tombol OK diklik pada bagian Store Result In. Setelah itu nilai K1 akan ditampilkan pada bagian Session
- 8) Menu Calc diklik, calculator disorot dan diklik, K2 diketik pada bagian Store Result in Variable, kolom Expression diklik, tombol 6 diklik, tanda * diklik, K1 disorot, tombol Select diklik, dan kemudian diklik OK
- 9) Menu Calc diklik, calculator disorot dan diklik, K3 diketik pada bagian Store Result in Variable, kolom Expression diklik, tombol 8 diklik, tanda * diklik, tombol 8 diklik, dan kemudian diklik OK
- 10) Menu Calc diklik, calculator disorot dan diklik, K4 diketik pada bagian Store Result in Variable, kolom Expression diklik, K3 disorot, tanda – diklik, tombol 1 diklik, dan kemudian diklik OK
- 11) Menu Calc diklik, calculator disorot dan diklik, K5 diketik pada bagian Store Result in Variable, kolom Expression diklik, tombol 8 diklik, tanda * diklik, K4 disorot, tombol Select diklik, dan kemudian diklik OK
- 12) Menu Calc diklik, calculator disorot dan diklik, K6 diketik pada bagian Store Result in Variable, kolom Expression diklik, K2 disorot, tombol Select diklik, tanda / diklik, K5 disorot, tombol Select diklik, dan kemudian diklik OK

- 13) Menu Calc diklik, calculator disorot dan diklik, K7 diketik pada bagian Store Result in Variable, kolom Expression diklik, tombol 1 diklik, tanda – diklik, K6 disorot, tombol Select diklik, dan kemudian diklik OK
- 14) Menu Data diklik, Display data disorot dan diklik, K2 sampai K7 disorot, tombol Select diklik, dan kemudian diklik OK
- 15) Display data akan ditampilkan pada bagian Session

Setelah langkah-langkah di atas dilakukan, maka display data hasil analisis korelasi Spearman yang telah dilakukan akan ditampilkan.

BAB 9
TEKNIK PENELUSURAN ANALISIS DATA PENELITIAN KUANTITATIF
DALAM PENELITIAN KEPENDIDIKAN

PENDAHULUAN

Salah satu tahap dalam proses penelitian adalah tahap analisis data. Tahap analisis data merupakan tahap penting di mana data yang dikumpulkan dengan menggunakan berbagai teknik pengumpulan data (misalnya; observasi, interview, angket maupun teknik pengumpulan data yang lain) diolah dan disajikan sedemikian rupa untuk membantu peneliti menjawab permasalahan yang ditelitinya.

Penelitian kependidikan pada hakekatnya tidak berbeda dengan penelitian-penelitiannya lainnya, khususnya penelitian ilmu-ilmu sosial, baik langkah-langkah maupun proserdurnya. Jika dibandingkan dengan penelitian ilmu-ilmu eksakta, memang terdapat perbedaan yang menonjol yakni pada karakteristik yang diteliti. Di mana peneliti tidak dapat melepaskan sisi humanitas dalam kegiatan penelitiannya. Aspek ini terlihat lebih jeas terutama penelitian kependidikan yang menggunakan disain eksperimental.

Kesulitan yang umumnya dijumpai dalam proses analisis data adalah memilih teknik analisis data yang paling tepat untuk permasalahan yang diteliti. Ketepatan ini berkaitan dengan jenis permasalahan, teknik pengukuran data, sampel yang diambil serta faktor-faktor yang lainnya.

Terdapat dua paradigma penelitian, yakni kuantitatif dan kualitatif. Sementara ahli ada yang memisahkan kedua paradigma tersebut, seperti Lexy

Moleong¹⁾ dan Noeng Muhadjir²⁾. Ahli lain berupaya memadukan kedua paradigma tersebut, seperti Julia Brannen³⁾. Dalam tulisan ringkas ini, penulis hanya membatasi pada penelusuran teknik analisis data kuantitatif, itupun sudah tentu mustahil untuk menjabarkan semua teknik analisis data yang cukup beragam serta kecanggihannya yang sangat bervariasi. Barangkali tulisan ini hanyalah merupakan suatu pengantar ringkas dalam pembahasan teknik analisis data kuantitatif di bidang kependidikan

Faktor-Faktor Yang Menentukan Teknik Analisis Data Kuantitatif

Dalam memilih teknik statistik untuk analisis data, paling tidak perlu dipertimbangkan empat faktor berikut:

1. Karakteristik **problem** atau permasalahan penelitian
2. Karakteristik **data** yang dikumpulkan
3. Karakteristik **sampel** atau cuplikan
4. Karakteristik **hubungan** dan **banyaknya variabel**

Penjabaran ringkas dari keempat faktor tersebut diuraikan dalam uraian berikut.

Karakteristik Permasalahan Penelitian

Ditinjau dari jenisnya, permasalahan penelitian dapat diklasifikasikan menjadi 3 (tiga) kelompok besar, yaitu: (1) deskripsi; (2) korelasi; dan (3) komparasi. Untuk masing-masing kategori permasalahan terdapat beberapa alternatif teknik analisis yang dapat dipilih, setelah mempertimbangkan faktor data, cuplikan dan variabelnya.

Penelitian yang bersifat deskriptif pada umumnya memerlukan pemerian suatu data untuk suatu variabel, menyangkut misalnya; mean (rerata), median, modus, simpang baku, varian, range, dan sejenisnya⁴⁾.

Di antara hitungan-hitungan tersebut, mean dan simpangan baku merupakan yang paling banyak dipergunakan. Terdapat beberapa rumus perhitungan mean, di antaranya :

$$\bar{X} = \frac{\sum x_i}{n} \qquad \bar{X} = \frac{\sum f_i x_i}{\sum f_i} \qquad \bar{X} = \frac{\sum n_i \bar{x}_i}{\sum n_i}$$

Rumus yang pertama digunakan untuk menghitung mean dari *data tunggal*, rumus yang kedua dipergunakan untuk data yang telah disusun dalam distribusi frekuensi, sedangkan rumus yang ke tiga dipergunakan untuk menghitung mean gabungan. Untuk *data kelompok* yang telah disusun dalam distribusi frekuensi, penghitungan X_i didasarkan pada *nilai tengah* dari panjang kelas interval. Misalkan terdapat kelas interval 61 – 70, maka X_i -nya adalah $(61 - 70)/2$ yakni 65,5⁵⁾.

Meskipun Mean banyak dipergunakan, namun kadang-kadang analisa dengan mean kurang tetap dalam menggambarkan realita yang sebenarnya. Misalnya di sebuah desa D yang berpenduduk 6000 orang, terdapat 12.000 ekor ayam. Dengan data tersebut kemudian seorang peneliti menyimpulkan bahwa rata-rata tiap penduduk memiliki 2 ekor ayam. Kesimpulan tersebut boleh jadi tidak menggambarkan realita yang sebenarnya karena yang memiliki ayam dalam desa tersebut hanya 10 orang saja sedangkan yang lainnya tidak memiliki.

Terdapat beberapa rumus simpangan baku (s), di antaranya:

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}} \qquad 3 \qquad s = \sqrt{\frac{\sum f_i (x_i - \bar{x})^2}{n - 1}}$$

Pada rumus yang pertama, terlebih dahulu dicari mean terlebih dahulu,

$$s = \sqrt{\frac{n \sum x_i^2 - (\sum x_i)^2}{n(n-1)}} \qquad s = \sqrt{\frac{n \sum f_i x_i^2 - (\sum f_i x_i)^2}{n(n-1)}}$$

sedangkan pada rumus yang kedua tidak perlu terlebih dahulu menghitung mean, hasilnya perhitungannya sama. Dua rumus yang sebelah kiri dipergunakan untuk data yang belum dibuat distribusi frekuensi, sedangkan dua rumus sebelah kanan dipergunakan untuk data yang telah disusun dalam distribusi frekuensi, yang data tunggal maupun data kelompok (terdapat panjang kelas interval). Misalnya

a. Contoh data tunggal

Nilai siswa	Frekuensi
4	4
5	8
6	12
7	7
8	5
9	2

b. Contoh data Kelompok

Nilai siswa	Frekuensi
31 – 40	2
41 - 50	7
51 – 60	9
61 – 70	14
71 – 80	8
80 - 90	3

Untuk data tunggal X_i adalah nilai siswa, sedangkan data kelompok X_i adalah tanda kelas atau nilai tengah atau mid point dari masing-masing kelas, dengan cara menjumlahkan nilai batas bawah dan nilai batas atas dibagi dua. Misalnya 31 – 40 maka nilai tengahnya adalah $(31 + 40) : 2 = 35,5$; $(41 + 50) : 2 = 45,5$ dan seterusnya. Dari segi keakuratan, data tunggal lebih akurat karena setiap data

mewakili dirinya sendiri, tidak sebagaimana data kelompok. Di sisi lain data kelompok lebih simple terutama untuk data yang memiliki rentang (jarak antara data tertinggi dengan terendah) yang panjang.

Simpangan baku sangat bermanfaat untuk mengetahui dispersi dan variasi data. Sebagai contoh sebagai berikut:

6	6	6	6	6
5	7	5	7	6
4	8	6	5	7

Ketiga data tersebut jika dihitung meannya akan diperoleh hasil yang sama, padahal penyebarannya sangat berbeda. Dengan dihitung simpangan baku dapat diketahui manakah data yang paling homogen, yakni yang memiliki simpangan baku terkecil.

Untuk penghitungan simpangan baku populasi (σ) terdapat sedikit perbedaan rumusnya yakni penyebut pada rumus simpangan baku populasi (σ) adalah N (banyaknya populasi), sedangkan rumus simpangan baku sampel (s) adalah n-1.

Pada penelitian yang berusaha untuk menemukan sejauh mana suatu variabel berkorelasi dengan variabel lainnya, ada yang bersifat *bivariate* (menyangkut dua variabel) atau *multivariate* (menyangkut lebih dari dua variabel). Dalam teknik analisis ini terdapat dua pertanyaan pokok yakni; banyaknya variabel yang dihubungkan dan tingkat pengukuran datanya, karena untuk itu masing-masing tersedia beberapa alternatif teknik analisis.

Untuk menguji hubungan 2 variabel terdapat beberapa teknik korelasi. Beberapa di antaranya yang terkenal banyak dipakai adalah korelasi *Pearson*, *Spearman* dan *Kendall's tau-b*. Teknik korelasi *Pearson* hanya digunakan untuk data

yang diukur dalam tingkat interval atau rasio. Sedangkan teknik korelasi *Spearman* dan *Kendall's tau-b* merupakan ukuran *nonparametrik* yang secara tertentu digunakan bilamana data memuat *outlier* atau bila distribusi datanya berbentuk nonnormal. Kedua teknik korelasi yang terakhir ini didasarkan pada penerapan *rank* pada kedua variabelnya.

Salah satu teknik korelasi untuk teknik korelasi *bivariate* adalah *Korelasi Product Moment*. Teknik korelasi ini merupakan salah satu teknik korelasi yang populer dipergunakan. Perlu diperhatikan bahwa teknik korelasi ini hanya tepat dipergunakan untuk data tingkat skala minimal interval, serta hubungan kedua variabel linier.

Terdapat beberapa rumus untuk menghitung koefisien korelasi product moment, salah satu di antaranya adalah:

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

Untuk uji hipotesis hubungan dua variabel yang mendasarkan analisis data sampel dapat menggunakan Uji “t” korelasi⁶⁾

$$t = \frac{r \sqrt{n-2}}{\sqrt{1-r^2}}$$

Uji “t” di atas dipergunakan untuk menguji hipotesis sebagai berikut:

Hipotesis Nihil (H_0) : $\rho = 0$ (tidak ada korelasi pada populasi), dan

Hipotesis Alternatif (H_a) : $\rho \neq 0$, (ada korelasi pada populasi) untuk uji dua pihak / dua sisi, dan

Hipotesis Alternatif (H_a) : $\rho > 0$ (ada korelasi positif pada populasi) atau $\rho < 0$ (ada korelasi negatif pada populasi), untuk uji satu pihak / satu sisi.

Manakala teori pendukung sudah ada atau telah ditemukan penelitian sebelumnya, uji satu pihak (kiri / kanan) lebih bermakna dari pada uji dua pihak. Namun jika teori pendukung belum menunjukkan kejelasan hubungan kedua variabel tersebut, maka uji dua pihak lebih baik dipergunakan.

Untuk uji dua pihak / atau dua sisi, untuk taraf nyata = α , maka hipotesis nihil diterima jika :

$$-t_{(1-\frac{1}{2}\alpha)} < t < t_{(1-\frac{1}{2}\alpha)}$$

Sedangkan untuk uji satu pihak kanan $\rho > 0$ hipotesis nihil diterima jika

$$t < t_{(1-\frac{1}{2}\alpha)}$$

Dan uji satu pihak kiri $\rho < 0$ hipotesis nihil diterima jika

$$-t_{(1-\frac{1}{2}\alpha)} < t$$

Di mana distribusi t yang digunakan mempunyai derajat kebebasan (dk) = (n-2), dalam hal lainnya H_0 ditolak.

Pada penelitian yang bersifat komparatif, biasanya peneliti membandingkan dua hal atau lebih. Teknik komparatif dapat dilakukan antara individu dengan

individu, individu dengan kelompok, maupun antar kelompok, dan untuk tiap-tiap hal tersebut terdapat teknik statistik tersendiri. Pertanyaan yang berkaitan dengan teknik analisis ini adalah berkaitan dengan jumlah kelompok yang dibandingkan serta tingkat pengukuran data yang dikumpulkan.

Untuk membandingkan dua variabel salah satu teknik yang banyak dipergunakan adalah “uji t”. Teknik uji ini diambil dari penemunya yakni “Student”, yakni kemudian terkenal dengan singkatannya “t”. Untuk membandingkan dua variabel uji yang dipergunakan di antaranya adalah uji mean (rata-rata hitung) dari kedua variabel tersebut. Uji ini dapat dikategorikan menjadi dua yakni⁷⁾:

- 1). Uji “t” *independent sample* (sampel dari kedua rata-rata yang hendak dibandingkan berasal dari sampel yang berbeda / lain sampel

$$t = \frac{\overline{X}_1 - \overline{X}_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

dengan :

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Statistik uji ini dipergunakan untuk menguji hipotesis:

Hipotesis Nihil (H_0) : $\mu_1 = \mu_2$, (rata-rata populasi keduanya sama) dan

Hipotesis Alternatif (H_a) : $\mu_1 \neq \mu_2$, (rata-rata populasi keduanya tidak sama) untuk uji dua pihak / dua sisi

Hipotesis Alternatif (H_a) : $\mu_1 > \mu_2$, atau $\mu_1 < \mu_2$, untuk uji satu pihak / satu sisi

Untuk uji dua pihak / atau dua sisi, untuk taraf nyata = α , maka hipotesis nihil diterima jika :

$$-t_{(1-\frac{1}{2}\alpha)} < t < t_{(1-\frac{1}{2}\alpha)}$$

Di mana distribusi t yang digunakan mempunyai derajat kebebasan (dk) = $(n_1 + n_2 - 2)$, dalam hal lainnya H_0 ditolak. Jika uji satu sisi (kanan atau kiri) sebagaimana uji t korelasi di atas.

Perlu ditambahkan di sini, bahwa teknik uji ini dipergunakan manakala $\sigma_1 = \sigma_2 = \sigma$ tetapi σ tidak diketahui

Jika kedua simpangan baku populasi tidak sama dan tidak diketahui, tetapi kedua populasi berdistribusi normal, hingga sekarang belum ada statistik yang tepat yang dapat digunakan. Pendekatan yang cukup memuaskan adalah dengan menggunakan statistik “ t ’ ” sebagai berikut:

Dengan kriteria pengujian: terima hipotesis nihil jika:

$$t' = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{(s_1^2 / n_1) + (s_2^2 / n_2)}}$$

Dengan :

$$-\frac{w_1 t_1 + w_2 t_2}{w_1 + w_2} < t' < \frac{w_1 t_1 + w_2 t_2}{w_1 + w_2}$$

$$w_1 = s_1^2/n_1 ; w_2 = s_2^2/n_1$$

$$t_1 = t_{(1-\frac{1}{2}\alpha)}, (n_1-1) \text{ dan}$$

$$t_2 = t_{(1-\frac{1}{2}\alpha)}, (n_2-1) \text{ dan}$$

2). Uji “t” *dependent sample* (sampel dari kedua rata-rata yang hendak dibandingkan berasal dari sampel yang sama / sampel sama).

Istilah lain uji ini adalah observasi berpasangan, dimana $\mu_B = \mu_1 - \mu_2$.

Hipotesis nihil dan alternatifnya adalah:

$$H_0 : \mu_B = 0$$

$$H_1 : \mu_B \neq 0$$

Jika $B_1 = x_1 - y_1, B_2 = x_2 - y_2, \dots, B_n = x_n - y_n$, maka data B_1, B_2, \dots, B_n

Menghasilkan rata-rata

$$\bar{B}$$

dan simpangan baku s_B . Untuk pengujian hipotesis dipergunakan statistik

uji:

$$t = \frac{\bar{B}}{s_B / \sqrt{n}}$$

Dengan kriteria keputusan hipotesis nihil diterima jika:

$$-t_{(1-1/2\alpha)} < t < t_{(1-1/2\alpha)}$$

$t_{1-1/2}$ didapat dari daftar distribusi “t” dengan peluang $1 - 1/2 \alpha$ dan dk = $(n - 1)$, di mana n adalah jumlah pasangan.

Meskipun tidak semua masalah penelitian dapat dimasukkan dalam salah satu dari ketiga klasifikasi di atas, namun dapat dikatakan sebagian besar dari penelitian kuantitatif yang banyak dilakukan dapat dikategorikan ke dalamnya.

Karakteristik Data Penelitian

Secara umum data dapat dikelompokkan menjadi diskrit (data pilah) dan kontinu (data menerus). Klasifikasi lain yang lebih konkrit sekaligus menunjukkan tingkatan pengukurannya, data dapat dikelompokkan menjadi data skala *nominal*, *ordinal*, *interval* dan *rasio*⁸⁾.

Data mempunyai level pengukuran *nominal* jika angka yang dikaitkan dengan suatu deskriptor hanya berfungsi sebagai pengganti (substitusi) deskriptor tersebut. Dengan kata lain dalam skala nominal angka-angka yang ada tidak mempunyai makna serta menunjukkan suatu besaran tertentu, sehingga tidak dapat dilakukan manipulasi secara matematis. Contoh skala ini misalnya; nomor punggung pemain sepakbola, nomor urut partai peserta pemilu, nomor presensi mahasiswa, dan sebagainya.

Skala *ordinal* adalah tingkat paling rendah dari suatu pengukuran data di mana angka yang dikaitkan dengan deskriptor suatu variabel mempunyai makna kuantitatif. Selain mempunyai karakteristik untuk membedakan antara obyek-obyek yang diukur seperti halnya skala nominal, angka ordinal ini mempunyai sifat tambahan yakni dapat memberi indikasi mana di antara dua obyek tertentu yang mempunyai kelebihan. Tingkat kejuaraan dalam suatu perlombaan atletik, nomor ranking prestasi siswa di sekolah, klasifikasi sosial ekonomi menjadi: tinggi, rendah dan sedang dan sejenisnya merupakan contoh pengukuran data ordinal.

Pada skala *interval*, selain mempunyai karakteristik sebagaimana dimiliki oleh skala nominal dan ordinal, juga mempunyai karakteristik tambahan berupa interval / jarak yang sama antara dua angka yang berturutan, sehingga jarak antara 5

dan 9 misalnya, akan sama besar dengan jarak antara 2 dan 6. Umur, ukuran suhu pada termometer, sekor tes dalam suatu ujian merupakan contoh-contoh pengukuran data yang mencapai level interval.

Pengukuran pada level *rasio* merupakan tingkat pengukuran tertinggi, yang mempunyai karakteristik tambahan berupa titik nol mutlak, sehingga memungkinkan adanya pengukuran proporsi / rasio misalnya 8 adalah dua kali lebih besar dari pada 4, dengan angka nol menandakan absennya deskriptor yang diukur. Berat, jarak, jumlah anak dan sejenisnya merupakan contoh-contoh data dengan skala rasio.

Karakteristik Sampel Penelitian

Dalam suatu studi yang sifatnya komparatif, haruslah diketahui dengan pasti berapa kelompok sampel yang akan dikomparasikan, dengan kata lain apakah peneliti hendak membandingkan dua kelompok sampel atau lebih. Komparasi antara dua kelompok sampel misalnya penelitian yang ingin mengkomparasikan sikap keberagaman antara siswa yang sekolah di sekolah “Umum” dan siswa yang sekolah di sekolah “Agama”. Apabila suatu penelitian tentang sikap keberagaman siswa dihubungkan dengan tingkat sosial ekonomi orang tua ke dalam klasifikasi; tinggi, sedang dan rendah, maka klasifikasinya adalah komparasi dengan lebih dari dua kelompok sampel. Apabila komparasi tersebut hendak digabungkan, maka akan diperoleh komparasi antara 6 kelompok sampel yang terbagai dalam dua dimensi, yaitu jenis sekolah dan status / tingkat sosial ekonomi orang tua. Ketiga skema model analisis tersebut dapat digambarkan sebagai berikut:

1. Komparasi dua kelompok sampel

Sekolah “Umum”	Sekolah “Agama”
----------------	-----------------

2. Komparasi tiga kelompok sampel

SES Tinggi	SES Sedang	SES Rendah
-------------------	-------------------	-------------------

3. Komparasi gabungan kedua hal di atas:

		STATUS SOSIAL EKONOMI		
		TINGGI	SEDANG	RENDAH
JENIS SEKOLAH	“UMUM”			
	“AGAMA”			

Meskipun ketiganya membandingkan masalah yang sama, perbedaan jumlah kelompok yang dibandingkan dan karakteristik datanya akan menyebabkan peneliti harus memilih teknik analisis yang berbeda antara kasus pertama, kedua dan ketiga.

Karakteristik dan Hubungan Antar Variabel

Secara umum, penelitian yang sifatnya asosiatif korelasional berusaha mencari sifat dan besarnya hubungan antar variabel-variabel penelitian yang nantinya diharapkan dapat memberikan penjelasan terhadap gejala yang sedang diamati. Teknik yang digunakan untuk mencari besarnya hubungan antar dua variabel berbeda dengan teknik untuk lebih dari dua variabel. Apabila hipotesis dalam suatu penelitian menyangkut tatahubung antara beberapa variabel (lebih dari dua), maka secara teoritik harus ditentukan terlebih dahulu sifat hubungan tersebut, kemudian baru dapat dipilih teknik analisis stataistik yang sesuai. Jadi perbedaan antara penggunaan **analisis regresi ganda, korelasi parsial** atau **analisis jalur** untuk memecahkan persoalan hubungan antara beberapa variabel tersebut tergantung pada kerangka berfikir teritik yang mendasari hipotesisnya serta level pengukuran datanya⁹⁾.

Jadi pada dasarnya, sebelum memilih teknik analisis, seorang peneliti terlebih dahulu harus menjawab pertanyaan-pertanyaan berikut sebagai pertimbangan dalam pemilihan teknik analisis data yang sesuai.

4 pertanyaan berkaitan dengan pemilihan teknik analisis¹⁰⁾

1. Apakah penelitian termasuk jenis deskripsi, korelasi atau komparasi ?
2. Apakah data penelitian yang diamati diukur dalam level nominal, ordinal, interval atau rasio ?
3. Berapakah jumlah variabel yang terlibat dan bagaimana tatahubung antara variabel-variabel tersebut?

4. Apakah yang akan dibandingkan, berapa jumlah kelompok yang dibandingkan, dan apakah kelompok tersebut berkaitan ataukah independen ?

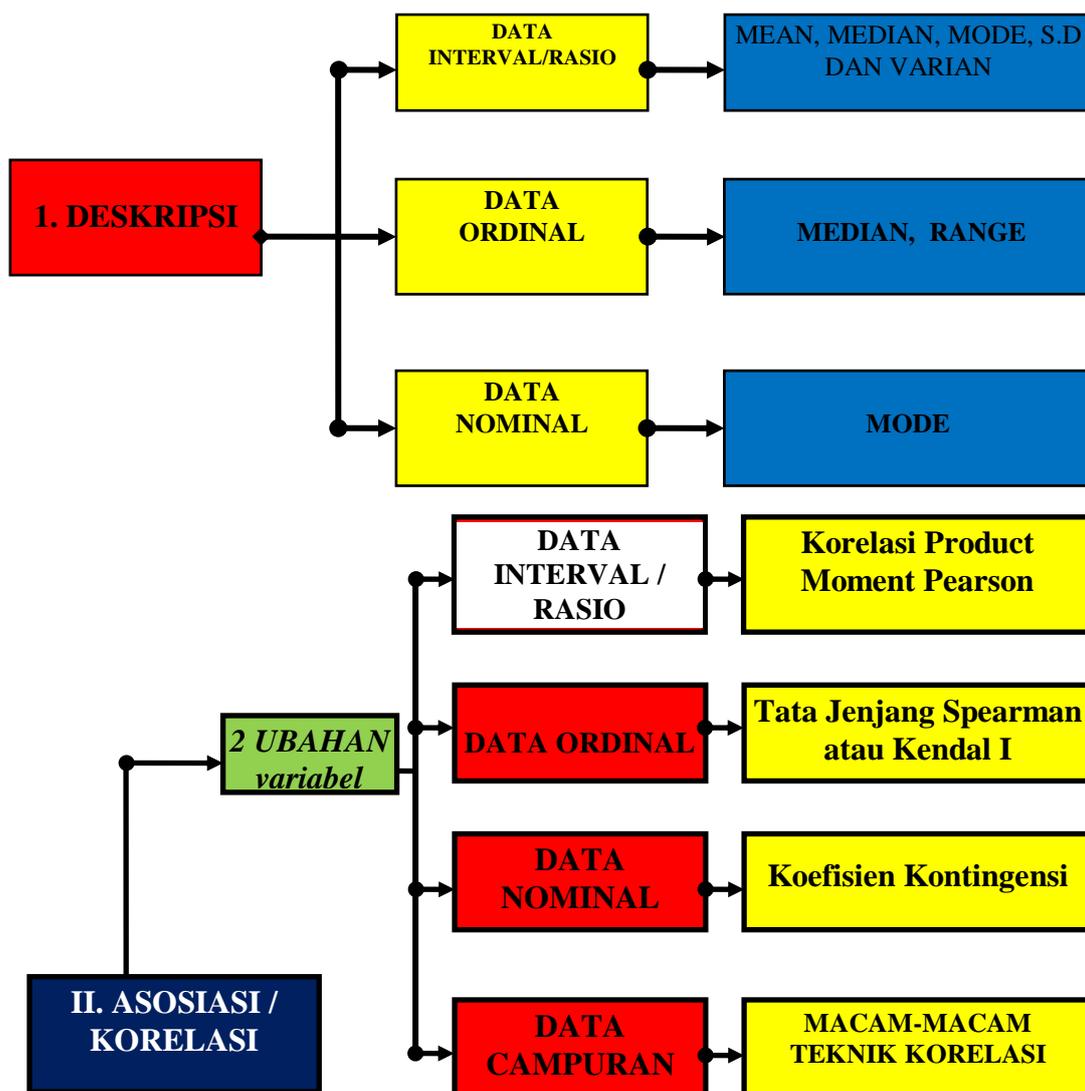
Jawaban terhadap empat pertanyaan di atas kemudian digunakan untuk memilih teknik analisis dengan melihat pada peta analisis pada bagian berikut. Dengan melihat peta berikut peneliti dapat memilih teknik analisis yang tepat untuk permasalahan yang ditelitinya. Peneliti juga dapat menciptakan secara kreatif dan komunikatif terhadap pola-pola penyajian data hasil penelitiannya, khususnya dalam teknik analisis deskriptif kuantitatif.

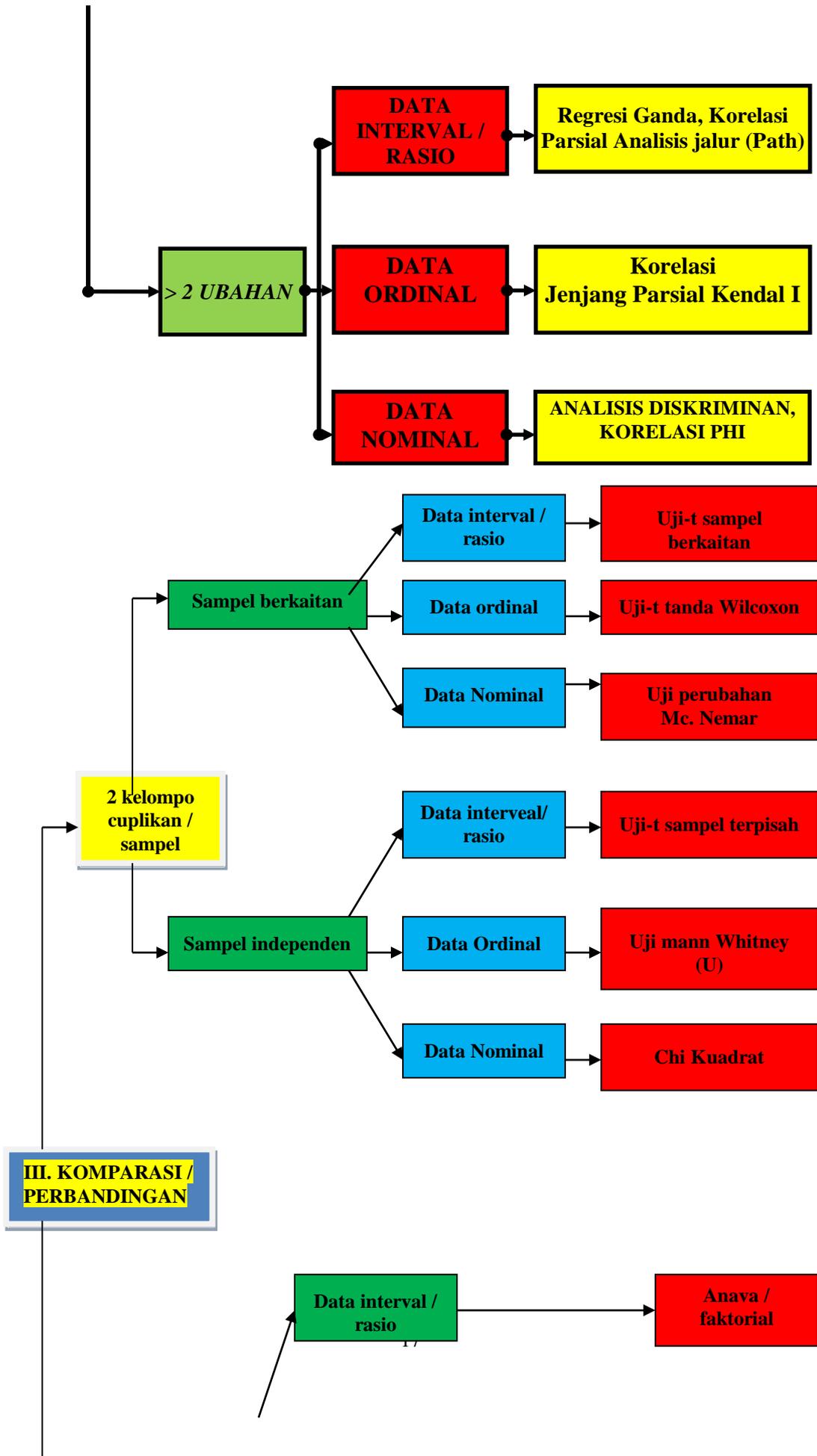
Dengan berkembangpesatnya teknologi komputer hingga saat ini peneliti dapat memanfaatkan program SPSS (*Statistical Programs for Social Sciencies*) untuk melakukan penghitungan data secara akurat, cermat dan cepat, baik untuk statistik deskriptif, korelatif maupun komparatif. Untuk program windows program SPSS telah mengalami perkembangan dari versi 6.0, 7.2, 9.0, 1.0 dan 11.0, 12, 13 dan 14.. Keuntungan lain data yang telah dimasukkan dapat dilakukan revisi (edit kembali) manakala terjadi kekeliruan, kelebihan ini yang tidak dimiliki oleh kalkulator. Di samping itu SPSS menyediakan berbagai fasilitas analisis serta penyajian hasil analisis tersebut dalam *out put*.¹¹

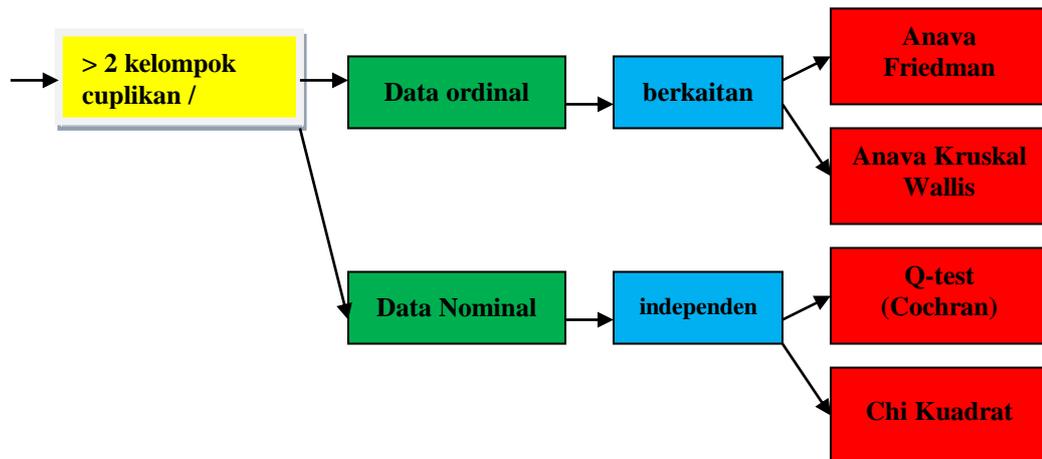
Hal lain yang perlu diingat bahwa langkah yang terpenting dalam analisis data adalah pemaknaan yang tersirat di balik yang tersurat, jadi tabel, gambar atau hasil cetak komputer yang dihasilkan oleh peneliti harus ditafsirkan dan dimaknakan agar dapat menjawab pertanyaan yang diteliti.

Untuk memudahkan dalam penelurusan dan pemilihan teknik analisis yang tepat berikut ini disajikan matrik analisis data dalam penelitian kuantitatif deskriptif, korelatif dan komparatif. Keterangan secara lebih terperinci tentang teknik-teknik analisis serta perhitungannya dapat ditelaah pada buku-buku referensi yang dicantumkan pada bagian akhir tulisan ini.

MATRIKS ANALISIS DATA KUANTITATIF¹¹⁾







-
- ¹⁾ Lexy J Moleong (1994), *Metodologi Penelitian Kualitatif*, Bandung: PT Remaja Rosdakarya.
- ²⁾ Noeng Muhadjir (2000), *Metodologi Penelitian Kualitatif*, edisi IV, Yogyakarta: Rake Serasin. Buku ini agak berbeda yang buku metodologi penelitian lainnya, di antara perbedaannya dalam buku ini diuraikan filsafat yang mendasarinya.
- ³⁾ Julia Brannen (1997), *Memadu Metode Penelitian Kualitatif & Kuantitatif*, Yogyakarta: Pustaka Pelajar.
- ⁴⁾ Glass, Gene V & Kenneth D Hopkins (1984), *Statistical Methods in Education and Psychology*, second edition, Prentice –Hall Inc. Englewood Cliffs, New Jersey 07632, hal: 31-34. Lihat juga Sudjana, *Metode Statistika*, Edisi ke.6, Bandung: Tarsito. Hal: 66-81 dan 93-99.
- ⁵⁾ Karena dalam penghitungan menggunakan nilai tengah (*mid-point*), maka dari segi akurasi, data kelompok kurang akurat, apalagi jika panjang kelas interval cukup lebar. Meskipun demikian penghitungan dengan menggunakan data kelompok dapat dibenarkan manakala *range* sangat lebar, misalnya: pendapatan penduduk.
- ⁶⁾ Sudjana, *Op Cit*, hlm. 380
- ⁷⁾ Sudjana, *Op Cit*, hal: 238 – 246.
- ⁸⁾ Glass, *Op Cit*, hl: 6-10
- ⁹⁾ Glass, *Op Cit*, hal: 459-471. Lihat juga Sudjana, *Op Cit*, 310-330 & 387-389.
- ¹⁰⁾ Sukamto (1997), “Modul 07: Teknik Analisis data” ,Technical Education Development Project ADB Loan no. 1100 INO PACKAGE ONE for The Directorate of Technical and vocational education, *Course Materials on Applied Educational Research*.

-
- ¹¹⁾ Terdapat banyak sekali buku-buku tentang SPSS antara lain: *Panduan Lengkap SPSS 6.0 for windows* (1998) Yogyakarta: Andi dan Wahana Komputer. Wahid Sulaiman (2002), *Jalan Pintas Menguasai SPSS 10*, Yogyakarta: Andi. Cornelius Trihendradi (2005) *Step by Step SPSS 13: Analisis Data Statistik*, Yogyakarta: Andi.
- ¹¹⁾ Dimodifikasi dari Sukamto, *Loc Cit.* Untuk uraian lebih lanjut dapat dilihat; Walter R. Borg & Meredith Damien Gall (1983), *Educational Research: An Introduction*, Fourth Edition, New York & London: Longman, Bab 13, 14, 14, 15, 16 halaman 530 – 731. Lihat pula Donald Ary, Lucy Cheser Jacobs, Asghar Razavieh (t.th), *Pengantar Penelitian dalam Pendidikan*, penerjemah Arief Furchan, Surabaya: Usaha Nasional, halaman: 458 – 465. Lihat pula John.W. Best (Metodologi Penelitian Pendidikan (1982) penyunting Sanapiah Faisal dan Mulyadi Guntur Waseso, Surabaya: Usaha Nasional, halaman: 247 – 382.

CURRICULUM VITAE

A. Identitas Diri

Nama : Dr. H. Rohmad, M. Pd.
Tempat / tgl. Lahir : Karanganyar, 22 Desember 1966
NIP : 19661222 199103 1 002
Pangkat/Gol. : Pembina Utama Muda (IV/c), TMT 1 Oktober 2009
Jabatan : Lektor Kepala, 735, 80 kum TMT 1 Mei 2009
Alamat Rumah : Beji RT.04/ RW.04 Kedungbanteng Banyumas
Alamat kantor : Jln. A. Yani. No. 40 a Purwokerto 53126
Nama Ayah : H. Qomari bin Setro Sarindi.
Nama Ibu : Kamsiyah binti Mustari
Nama Isteri : Rofiqoh, S. Ag.
Nama Anak : 1. Farah Zia Ulul Albab
2. Zidna Qoulan Tsaqila
3. Zidane Asyad Hubbalillah

B. Riwayat Pendidikan

1. Pendidikan Formal

- a. SD Negeri Kalioso III, Gondangrejo, Karanganyar, lulus 1977
- b. SMP Negeri 1 Gemolong, Sragen, lulus 1981
- c. PGA Negeri Surakarta, lulus tahun 1984
- d. S-1, Fakultas Tarbiyah, Jurusan Pendidikan Agama Islam (PAI), Institut Agama Islam Negeri (IAIN) Sunan Kalijaga, Yogyakarta, lulus tahun 1989
- e. S-2, Program Pascasarjana, Penelitian dan Evaluasi Pendidikan (PEP), Universitas Negeri Yogyakarta, lulus tahun 2001.
- f. S-3, Pascasarjana UIN Sunan Kalijaga, Islamic studies lulus 4 September 2014.

2. Pendidikan Non-Formal

- a. Sekolah Rakyat Islam (SRI), tahun 1975 s.d. 1981
- b. Pondok Pesantren Wahid Hasyim Gaten, Condong Catur, Depok, Sleman, Yogyakarta, tahun 1985 s.d. 1989.

C. Riwayat Pekerjaan

1. Dosen Luar Biasa Fakultas Tarbiyah IAIN Sunan Kalijaga Yogyakarta tahun 1990 s.d. 1992
2. Dosen Tetap STAIN Purwokerto (Sebelumnya bernama Fakultas Tarbiyah IAIN Sunan Kalijaga di Purwokerto 1964 s.d. 1995, Fakultas Tarbiyah IAIN Walisongo di Purwokerto tahun 1995 s.d. 1997, 1997 sampai sekarang -2014- STAIN Purwokerto) tahun 1991 s.d. sekarang (tahun 2014)

3. Kepala Pusat Pelatihan Mengajar STAIN Purwokerto tahun 1997 s.d. 1998
4. Ketua Program Diploma II STAIN Purwokerto tahun 2002 s.d. 2006
5. Ketua Program Studi Pendidikan Guru Madrasah Ibtidaiyah STAIN Purwokerto tahun 2007 s.d. 2010
6. Pembantu Ketua 1 STAIN Purwokerto 2010 s.d. 2013
7. Wakil Ketua 1 Bidang Akademik dan Pengembangan Lembaga tahun 2013 s.d. 2014.
8. Kepala UPT Bahasa STAIN Purwokerto, 1 Desember 2014 s.d. 31 Mei 2015.
9. Ketua Program Studi Pendidikan Agama Islam Pascasarjana IAIN Purwokerto, 1 Juni s.d. sekarang (14 September 2015).

D. Prestasi / Penghargaan

1. Lencana Karya Satya 10 tahun
2. Lencana Karya Satya 20 tahun
3. Peserta *Academic Recharging for Islamic Higher Education* (ARFI) di Turki tahun 2012.

E. Pengalaman Organisasi

1. Ketua Komite Sekolah SD 01, 02, 03 Beji Kedungbanteng Banyumas tahun 2010 s.d. sekarang (2014)
2. Sekretaris NU Ranting Beji 1 Kedungbanteng Banyumas tahun 2000 s.d. 2002.
3. Katib Syuriah NU Ranting Beji 1 Kedungbanteng Banyumas tahun 2002 s.d. 2004.
4. Rais Syuriah NU Ranting Beji 1 Kedungbanteng Banyumas tahun 2004 s.d. 2008.
5. Anggota Ikatan Sarjana Pendidikan Indonesia (ISPI)
6. Sekretaris BPD Desa Beji Kedungbanteng Banyumas tahun 2008 s.d. 2013.
7. Ketua RW.04 Beji Kedungbanteng tahun 2007 s.d. sekarang (2014)

F. Minat Keilmuan

1. Evaluasi Pendidikan
2. Metode Penelitian Pendidikan
3. Penelitian Tindakan Kelas
4. Statistika Pendidikan

G. Karya Ilmiah

1. Kyai Haji Ahmad Dahlan. Skripsi Fakultas Tarbiyah IAIN Sunan Kalijaga Yogyakarta tahun 1989
2. Evaluasi Kurikulum Program Studi PAI STAIN Purwokerto, Thesis Program Program Pascasarjana Universitas Negeri Yogyakarta tahun 2000
3. Model-model Evaluasi, Jurnal Insania vol 13 no. 2 tahun 2008
4. Penelusuran Teknis Analisis Data Kuantitatif bidang Kependidikan, Jurnal Insania tahun 2009
5. Relasi Gender Mantan TKW di Kabupaten Banyumas, 2009
6. Peningkatan Prestasi Belajar Mata Kuliah Statistika Pendidikan dengan menggunakan *Strategi Practice- Rehearsal Pair* di Jurusan Tarbiyah STAIN Purwokerto, Jurnal Penelitian Agama P3m STAIN Purwokerto, tahun 2010.
7. Pengembangan Kompetensi Guru Madrasah Ibtidaiyah di Kabupaten Banyumas, 2010
8. Statistika Pendidikan Menggunakan Microsoft Excel dan Minitab, Purwokerto: STAIN Press, 2012.
9. Pengembangan Instrumen Evaluasi dan Penelitian, Purwokerto: STAIN Press, 2014
10. Pengantar Statistika: Panduan Praktis bagi Pengajar dan Mahasiswa, Yogyakarta: Kalimedia, 2014
11. Pengembangan Instrumen Evaluasi Domain Afektif, Purwokerto: STAIN Press, 2015.

Purwokerto, 14 September 2015



Dr. H. Rohmad, M. Pd.



Supriyanto dilahirkan di Batang Jawa Tengah pada tanggal 25 Mei 1974. Ia menyelesaikan Pendidikan Dasar dan Menengahnya di Batang dan Yogyakarta. Setelah tamat SMA Negeri 4 Yogyakarta, ia meneruskan studinya di Program Studi Statistika Universitas Gadjah Mada pada tahun 1992. Lulus sebagai Sarjana Statistika (S.Si.) tahun 1997, mengikuti Program Karyasiswa Dikti Departemen Pendidikan dan Kebudayaan sebagai penerima Beasiswa Pascasarjana Ikatan Dinas di Universitas Jenderal Soedirman dengan mengambil S2 di Jurusan Matematika konsentrasi Statistika Universitas Gadjah Mada tahun 1998 dan lulus tahun 2001. Kini ia mendisiplin ilmu Statistika yang ter bentuk atas bimbingan (alm.) Prof. Drs. H Zanzawi Soejoeti, M.Sc., Ph.D, yang merupakan guru besar pertama bidang statistika di UGM (teriring doa, semoga pahaladari ilmu yang bermanfaat ini selalunya mengalir bagai belia). Sejak 2000 ia tercatat sebagai staf pengajar Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, FMIPA (yang sebelumnya Fakultas Sains dan Teknik) Universitas Jenderal Soedirman.

Selain sebagai staf pengajar (Lektor dalam bidang Statistika) Universitas Jenderal Soedirman ia juga sebagai Dosen Luar Biasa di STAIN Purwokerto sejak tahun 2007 dan beberapa pendidikan tinggi lainnya. Sampai saat ini penulis juga tercatat sebagai anggota Indonesian Mathematical Society (IndoMS).

DAFTAR PUSTAKA

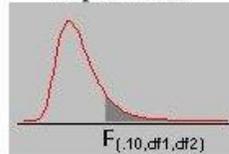
- Arikunto, Suharsimi (2000), *Manajemen Penelitian*, Edisi Baru Cetakan kelima, Jakarta: PT Rineka Cipta.
- _____, (2010), *Prosedur Penelitian: Suatu Pendekatan Praktik*, Jakarta: Rineka Cipta.
- Atmaja, SL (2009), *Statistika untuk Bisnis dan Ekonomi*, Yogyakarta: Andi
- Anderson, D.R., D.J. Sweeney dan T.A. Williams (2004), *Quantitative Methods for Business*, 9th ed. Singapore: Thomsone-Brooks/Cole.
- Bhattacharryya, G.K. dan R.A. Johnson (1977), *Statistical Concepts and Methods*, John Wiley.
- Bluman, AG., (2012), *Elementary Statistics: a step by step approach*, 8th ed. McGraw-Hill
- Borg, Walter R.& Meredith Damien Gall (1983), *Educational Research: An Introduction*, Fourth Edition, New York & London: Longman, 1983
- Brannen, Julia, (1997), *Memadu Metode Penelitian Kualitatif & Kuantitatif*, Yogyakarta: Pustaka Pelajar.
- Cochran, William G.(2005), *Teknik Penarikan Sampel*, penerjemah Rudiansyah, Jakarta: Penerbit Universitas Indonesia, edisi ketiga.
- Donald Ari., Lucy Cheser Jacobs., Asghar Razavich, *Introduction to Research in Education*, Penterjemah Arief Furchan(t.t), *Pengantar Penelitian dalam Pendidikan*, Surabaya: Usaha nasional.
- Eppen, G.D. dan E.J. Gould (1979), *Quantitative Concepts for Management: Decision Making Without Algorithms*, Englewood Cliffs, New Jersey, USA: Prentice Hall.
- Fraenkel, J. dan Wallen, N. (1993), *How to Design and Evaluate Research in Education*, 2nd , New York: McGraw-Hill Inc.
- Gay, L.R. dan Diehl, P.L. (1992), *Research Methods for Business and Management*, New York: McMillan Publishing Company.
- Glass, Gene V and Kenneth Hopkin (1984), *Statistical Method in Education and Psychology*, 2nd , Englewood Cliffs, New Jersey 07632: Prentice Hall.
- Hartono (2004), *Statistik untuk Penelitian*, Yogyakarta: Pustaka Pelajar.
- John W Best, *Research in Education*, Penyunting Sanafiah Faisah (1982), *Metodologi Penelitian Pendidikan*,: Surabaya Usaha Nasional
- Keith P.L. (2006), Statistical Power, Sample sizes and the software to Calculate them easily, *Bio Science*, Vol. 56. No. 7 (July 2006), P. 607-612.
- Krejcie, R.V. dan Morgan, D.W., (1970), *Determining sample size for research activities Educational and Psychological Measurement*, 30, 607-610.
- Kirk, Roger E. (1968), *Experimental Design: Procedures for the Behavioral Sciences*, Belmont, California: Brooks/Cole Publishing Company.

- Kleinbaum, David G. and Lawrence L. Kupper (1978), *Applied Regression?, Analysis and Other Multivariable Methods*, North Scituate, Massachusetts: Duxbury Press.
- Kuncoro, M., (2009), *Metode Riset untuk Bisnis & Ekonomi: Bagaimana Meneliti dan Menulis Tesis ?*, edisi 3, Erlangga.
- Kusmayadi, ES.(200), *Metode Penelitian dalam Bidang Kepariwisataaan*, Jakarta: PT Gramedia Pustaka Utama.
- Malhotra K.N. (1993), *Marketing Research an Applied Orientation*, 2nd edition, New Jersey: Prentice Hall International Inc.
- Mohamad Ali (1987), *Penelitian Kependidikan: Prosedur & Strategi*, Bandung: Angkasa.
- Moleong, Lexy J (1994), *Metodologi Penelitish Kualitatif*, Bandung: PT Remaja Rosdakarya..
- Noeng Muhadjir (2000)r, *Metodologi Penelitian Kualitatif*, edisi IV, Yogyakarta: Rake Serasin.
- Richard J. Harris dan Dana Quade (1992), The Minimally Important Difference Significant Criterion for Sample Size, *Journal of Educational Statistics*, vol. 17, no. 1 (Spring), p. 27-49.
- Riduwan (2005), *Belajar Mudah Penelitian Untuk Guru, Karyawan dan Peneliti Pemula*, Bandung: Alfabeta.
- Saldanha. L dan P Thomson (2003), Conceptions of Sample and Their Relationship to Statistical Inference, *Educational Studies in Mathematic*, vol. 51, no. 3, p. 257-270.
- Sekaran, U. (2006) *Metodologi Penelitian untuk Bisnis*, Jakarta: Salemba Empat.
- Singarimbun, M. (2011), *Metode Penelitian Survey*, Jakarta: LP3ES.
- Shaughnessy, John J., Eugene B. Zechmeister., Jeanne S. Zechmeister, (2007), *Metodologi Penelitian Psikologi*, Penerjemah: Helly Prajitno Soetjipto & Sri Mulyantini Soetjipto, Yogyakarta: Pustaka Pelajar.
- Soejoeti, Z. (1988), *Metode Statistika*, Universitas Terbuka.
- Sudjana, (1995), *Desain dan Analisis Eksperimen*, edisi IV, Bandung: Tarsito.
- _____, (1996), *Metoda Statistika*, edisi. 6, bandung: Tarsito.
- _____, (2003), *Teknik Analisis Regresi dan Korelasi bagi Para Peneliti*, cetakan ulang ketiga, Bandung: Tarsito.
- Sugiyono (2008), *Metode Penelitian Pendidikan: Pendekatan Kuantitatif, Kualitatif, dan R & D*, Bandung: Alfabeta.
- _____(2007), *Statistik Non Parametrik untuk Penelitian*, Bandung: Alfabeta.
- Sukamto (1997), "Modul 07: Teknik Analisis data" ,Technical Education Development Project ADB Loan no. 1100 INO PACKAGE ONE for The Directorate of Technical and vocational education, *Course Materials on Applied Educational Research*.
- Supranto, J. (2007), *Teknik Sampling untuk Survei dan Eksperimen*, Jakarta: Rineka Cipta.
- Susetyo, B. (2009), *Statistika*, Jakarta: Direktorat Jendral Pendidikan Islam Departemen Agama RI.

- Suwarno, (2013), Analisis Regresi Linier Sederhana dan Korelasi,
<http://suwarnostatistik.wordpress.com>, diunduh tanggal 20 April 2013.
- Sarwono, J. (2006), *Metode Penelitian Kuantitatif dan Kualitatif*, Yogyakarta: Graha Ilmu.
- Sarwono, J. (2013), Teori Analisis Korelasi: Mengenal Analisis Korelasi,
<http://www.jonathansarwono.info/korelasi/korelasi.htm>.
- Walpole, R.E. (1982), *Introduction to Statistic*, Mc Milan, 3th edition
- Walpole, R.E. & R.H. Myers (1982), *Probability and Statistics for Engineers and Scientist*,
McMilan 2nd edition.

Lampiran 4
Tabel Distribusi F

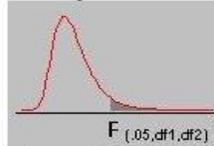
Tabel Distribusi F
alpha=0.10



Df 1 Df2	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	INF
1	39.86346	49.50000	53.59324	55.83296	57.24008	58.20442	58.90595	59.43898	59.85759	60.19498	60.70521	61.22034	61.74029	62.00205	62.26497	62.52905	62.79428	63.06064	63.32812
2	8.52632	9.00000	9.16179	9.24342	9.29263	9.32553	9.34908	9.36677	9.38054	9.39157	9.40813	9.42471	9.44131	9.44962	9.45793	9.46624	9.47456	9.48289	9.49122
3	5.53832	5.46238	5.39077	5.34264	5.30916	5.28473	5.26619	5.25167	5.24000	5.23041	5.21562	5.20031	5.18448	5.17636	5.16811	5.15972	5.15119	5.14251	5.13370
4	4.54477	4.32456	4.19086	4.10725	4.05058	4.00975	3.97897	3.95494	3.93567	3.91988	3.89553	3.87036	3.84434	3.83099	3.81742	3.80361	3.78957	3.77527	3.76073
5	4.06042	3.77972	3.61948	3.52020	3.45298	3.40451	3.36790	3.33928	3.31628	3.29740	3.26824	3.23801	3.20665	3.19052	3.17408	3.15732	3.14023	3.12279	3.10500
6	3.77595	3.46330	3.28876	3.18076	3.10751	3.05455	3.01446	2.98304	2.95774	2.93693	2.90472	2.87122	2.83634	2.81834	2.79996	2.78117	2.76195	2.74229	2.72216
7	3.58943	3.25744	3.07407	2.96053	2.88334	2.82739	2.78493	2.75158	2.72468	2.70251	2.66811	2.63223	2.59473	2.57533	2.55546	2.53510	2.51422	2.49279	2.47079
8	3.45792	3.11312	2.92380	2.80643	2.72645	2.66833	2.62413	2.58935	2.56124	2.53804	2.50196	2.46422	2.42464	2.40410	2.38302	2.36136	2.33910	2.31618	2.29257
9	3.36030	3.00645	2.81286	2.69268	2.61061	2.55086	2.50531	2.46941	2.44034	2.41632	2.37888	2.33962	2.29832	2.27683	2.25472	2.23196	2.20849	2.18427	2.15923
10	3.28502	2.92447	2.72767	2.60534	2.52164	2.46058	2.41397	2.37715	2.34731	2.32260	2.28405	2.24351	2.20074	2.17843	2.15543	2.13169	2.10716	2.08176	2.05542
11	3.22520	2.85951	2.66023	2.53619	2.45118	2.38907	2.34157	2.30400	2.27350	2.24823	2.20873	2.16709	2.12305	2.10001	2.07621	2.05161	2.02612	1.99965	1.97211
12	3.17655	2.80680	2.60552	2.48010	2.39402	2.33102	2.28278	2.24457	2.21352	2.18776	2.14744	2.10485	2.05968	2.03599	2.01149	1.98610	1.95973	1.93228	1.90361
13	3.13621	2.76317	2.56027	2.43371	2.34672	2.28298	2.23410	2.19535	2.16382	2.13763	2.09659	2.05316	2.00698	1.98272	1.95757	1.93147	1.90429	1.87591	1.84620
14	3.10221	2.72647	2.52222	2.39469	2.30694	2.24256	2.19313	2.15390	2.12195	2.09540	2.05371	2.00953	1.96245	1.93766	1.91193	1.88516	1.85723	1.82800	1.79728
15	3.07319	2.69517	2.48979	2.36143	2.27302	2.20808	2.15818	2.11853	2.08621	2.05932	2.01707	1.97222	1.92431	1.89904	1.87277	1.84539	1.81676	1.78672	1.75505
16	3.04811	2.66817	2.46181	2.33274	2.24376	2.17833	2.12800	2.08798	2.05533	2.02815	1.98539	1.93992	1.89127	1.86556	1.83879	1.81084	1.78156	1.75075	1.71817
17	3.02623	2.64464	2.43743	2.30775	2.21825	2.15239	2.10169	2.06134	2.02839	2.00094	1.95772	1.91169	1.86236	1.83624	1.80901	1.78053	1.75063	1.71909	1.68564
18	3.00698	2.62395	2.41601	2.28577	2.19583	2.12958	2.07854	2.03789	2.00467	1.97698	1.93334	1.88681	1.83685	1.81035	1.78269	1.75371	1.72322	1.69099	1.65671
19	2.98990	2.60561	2.39702	2.26630	2.17596	2.10936	2.05802	2.01710	1.98364	1.95573	1.91170	1.86471	1.81416	1.78731	1.75924	1.72979	1.69876	1.66587	1.63077
20	2.97465	2.58925	2.38009	2.24893	2.15823	2.09132	2.03970	1.99853	1.96485	1.93674	1.89236	1.84494	1.79384	1.76667	1.73822	1.70833	1.67678	1.64326	1.60738
21	2.96096	2.57457	2.36489	2.23334	2.14231	2.07512	2.02325	1.98186	1.94797	1.91967	1.87497	1.82715	1.77555	1.74807	1.71927	1.68896	1.65691	1.62278	1.58615
22	2.94858	2.56131	2.35117	2.21927	2.12794	2.06050	2.00840	1.96680	1.93273	1.90425	1.85925	1.81106	1.75899	1.73122	1.70208	1.67138	1.63885	1.60415	1.56678
23	2.93736	2.54929	2.33873	2.20651	2.11491	2.04723	1.99492	1.95312	1.91888	1.89025	1.84497	1.79643	1.74392	1.71588	1.68643	1.65535	1.62237	1.58711	1.54903
24	2.92712	2.53833	2.32739	2.19488	2.10303	2.03513	1.98263	1.94066	1.90625	1.87748	1.83194	1.78308	1.73015	1.70185	1.67210	1.64067	1.60726	1.57146	1.53270

25	2.91774	2.52831	2.31702	2.18424	2.09216	2.02406	1.97138	1.92925	1.89469	1.86578	1.82000	1.77083	1.71752	1.68898	1.65895	1.62718	1.59335	1.55703	1.51760
26	2.90913	2.51910	2.30749	2.17447	2.08218	2.01389	1.96104	1.91876	1.88407	1.85503	1.80902	1.75957	1.70589	1.67712	1.64682	1.61472	1.58050	1.54368	1.50360
27	2.90119	2.51061	2.29871	2.16546	2.07298	2.00452	1.95151	1.90909	1.87427	1.84511	1.79889	1.74917	1.69514	1.66616	1.63560	1.60320	1.56859	1.53129	1.49057
28	2.89385	2.50276	2.29060	2.15714	2.06447	1.99585	1.94270	1.90014	1.86520	1.83593	1.78951	1.73954	1.68519	1.65600	1.62519	1.59250	1.55753	1.51976	1.47841
29	2.88703	2.49548	2.28307	2.14941	2.05658	1.98781	1.93452	1.89184	1.85679	1.82741	1.78081	1.73060	1.67593	1.64655	1.61551	1.58253	1.54721	1.50899	1.46704
30	2.88069	2.48872	2.27607	2.14223	2.04925	1.98033	1.92692	1.88412	1.84896	1.81949	1.77270	1.72227	1.66731	1.63774	1.60648	1.57323	1.53757	1.49891	1.45636
40	2.83535	2.44037	2.22609	2.09095	1.99682	1.92688	1.87252	1.82886	1.79290	1.76269	1.71456	1.66241	1.60515	1.57411	1.54108	1.50562	1.46716	1.42476	1.37691
60	2.79107	2.39325	2.17741	2.04099	1.94571	1.87472	1.81939	1.77483	1.73802	1.70701	1.65743	1.60337	1.54349	1.51072	1.47554	1.43734	1.39520	1.34757	1.29146
120	2.74781	2.34734	2.12999	1.99230	1.89587	1.82381	1.76748	1.72196	1.68425	1.65238	1.60120	1.54500	1.48207	1.44723	1.40938	1.36760	1.32034	1.26457	1.19256
inf	2.70554	2.30259	2.08380	1.94486	1.84727	1.77411	1.71672	1.67020	1.63152	1.59872	1.54578	1.48714	1.42060	1.38318	1.34187	1.29513	1.23995	1.16860	1.00000

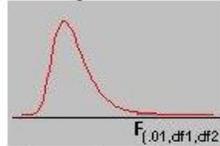
alpha=0.05



df2/df1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	INF
1	161.4476	199.5000	215.7073	224.5832	230.1619	233.9860	236.7684	238.8827	240.5433	241.8817	243.9060	245.9499	248.0131	249.0518	250.0951	251.1432	252.1957	253.2529	254.3144
2	18.5128	19.0000	19.1643	19.2468	19.2964	19.3295	19.3532	19.3710	19.3848	19.3959	19.4125	19.4291	19.4458	19.4541	19.4624	19.4707	19.4791	19.4874	19.4957
3	10.1280	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123	8.7855	8.7446	8.7029	8.6602	8.6385	8.6166	8.5944	8.5720	8.5494	8.5264
4	7.7086	6.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.0410	5.9988	5.9644	5.9117	5.8578	5.8025	5.7744	5.7459	5.7170	5.6877	5.6581	5.6281
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725	4.7351	4.6777	4.6188	4.5581	4.5272	4.4957	4.4638	4.4314	4.3985	4.3650
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.0990	4.0600	3.9999	3.9381	3.8742	3.8415	3.8082	3.7743	3.7398	3.7047	3.6689
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767	3.6365	3.5747	3.5107	3.4445	3.4105	3.3758	3.3404	3.3043	3.2674	3.2298
8	5.3177	4.4590	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881	3.3472	3.2839	3.2184	3.1503	3.1152	3.0794	3.0428	3.0053	2.9669	2.9276
9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789	3.1373	3.0729	3.0061	2.9365	2.9005	2.8637	2.8259	2.7872	2.7475	2.7067
10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204	2.9782	2.9130	2.8450	2.7740	2.7372	2.6996	2.6609	2.6211	2.5801	2.5379
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.9480	2.8962	2.8536	2.7876	2.7186	2.6464	2.6090	2.5705	2.5309	2.4901	2.4480	2.4045
12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964	2.7534	2.6866	2.6169	2.5436	2.5055	2.4663	2.4259	2.3842	2.3410	2.2962
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144	2.6710	2.6037	2.5331	2.4589	2.4202	2.3803	2.3392	2.2966	2.2524	2.2064
14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458	2.6022	2.5342	2.4630	2.3879	2.3487	2.3082	2.2664	2.2229	2.1778	2.1307
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876	2.5437	2.4753	2.4034	2.3275	2.2878	2.2468	2.2043	2.1601	2.1141	2.0658
16	4.4940	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377	2.4935	2.4247	2.3522	2.2756	2.2354	2.1938	2.1507	2.1058	2.0589	2.0096
17	4.4513	3.5915	3.1968	2.9647	2.8100	2.6987	2.6143	2.5480	2.4943	2.4499	2.3807	2.3077	2.2304	2.1898	2.1477	2.1040	2.0584	2.0107	1.9604
18	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5767	2.5102	2.4563	2.4117	2.3421	2.2686	2.1906	2.1497	2.1071	2.0629	2.0166	1.9681	1.9168
19	4.3807	3.5219	3.1274	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227	2.3779	2.3080	2.2341	2.1555	2.1141	2.0712	2.0264	1.9795	1.9302	1.8780
20	4.3512	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	2.4471	2.3928	2.3479	2.2776	2.2033	2.1242	2.0825	2.0391	1.9938	1.9464	1.8963	1.8432
21	4.3248	3.4668	3.0725	2.8401	2.6848	2.5727	2.4876	2.4205	2.3660	2.3210	2.2504	2.1757	2.0960	2.0540	2.0102	1.9645	1.9165	1.8657	1.8117
22	4.3009	3.4434	3.0491	2.8167	2.6613	2.5491	2.4638	2.3965	2.3419	2.2967	2.2258	2.1508	2.0707	2.0283	1.9842	1.9380	1.8894	1.8380	1.7831
23	4.2793	3.4221	3.0280	2.7955	2.6400	2.5277	2.4422	2.3748	2.3201	2.2747	2.2036	2.1282	2.0476	2.0050	1.9605	1.9139	1.8648	1.8128	1.7570
24	4.2597	3.4028	3.0088	2.7763	2.6207	2.5082	2.4226	2.3551	2.3002	2.2547	2.1834	2.1077	2.0267	1.9838	1.9390	1.8920	1.8424	1.7896	1.7330
25	4.2417	3.3852	2.9912	2.7587	2.6030	2.4904	2.4047	2.3371	2.2821	2.2365	2.1649	2.0889	2.0075	1.9643	1.9192	1.8718	1.8217	1.7684	1.7110
26	4.2252	3.3690	2.9752	2.7426	2.5868	2.4741	2.3883	2.3205	2.2655	2.2197	2.1479	2.0716	1.9898	1.9464	1.9010	1.8533	1.8027	1.7488	1.6906
27	4.2100	3.3541	2.9604	2.7278	2.5719	2.4591	2.3732	2.3053	2.2501	2.2043	2.1323	2.0558	1.9736	1.9299	1.8842	1.8361	1.7851	1.7306	1.6717

28	4.1960	3.3404	2.9467	2.7141	2.5581	2.4453	2.3593	2.2913	2.2360	2.1900	2.1179	2.0411	1.9586	1.9147	1.8687	1.8203	1.7689	1.7138	1.6541
29	4.1830	3.3277	2.9340	2.7014	2.5454	2.4324	2.3463	2.2783	2.2229	2.1768	2.1045	2.0275	1.9446	1.9005	1.8543	1.8055	1.7537	1.6981	1.6376
30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107	2.1646	2.0921	2.0148	1.9317	1.8874	1.8409	1.7918	1.7396	1.6835	1.6223
40	4.0847	3.2317	2.8387	2.6060	2.4495	2.3359	2.2490	2.1802	2.1240	2.0772	2.0035	1.9245	1.8389	1.7929	1.7444	1.6928	1.6373	1.5766	1.5089
60	4.0012	3.1504	2.7581	2.5252	2.3683	2.2541	2.1665	2.0970	2.0401	1.9926	1.9174	1.8364	1.7480	1.7001	1.6491	1.5943	1.5343	1.4673	1.3893
120	3.9201	3.0718	2.6802	2.4472	2.2899	2.1750	2.0868	2.0164	1.9588	1.9105	1.8337	1.7505	1.6587	1.6084	1.5543	1.4952	1.4290	1.3519	1.2539
inf	3.8415	2.9957	2.6049	2.3719	2.2141	2.0986	2.0096	1.9384	1.8799	1.8307	1.7522	1.6664	1.5705	1.5173	1.4591	1.3940	1.3180	1.2214	1.0000

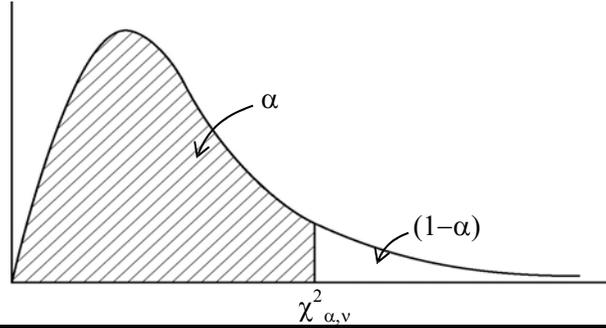
alpha=0.01



Df1/ Df2	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	INF
1	4052.18	4999.50	5403.35	5624.58	5763.65	5858.99	5928.36	5981.07	6022.47	6055.85	6106.32	6157.285	6208.73	6234.631	6260.65	6286.78	6313.03	6339.39	6365.864
2	98.503	99.000	99.166	99.249	99.299	99.333	99.356	99.374	99.388	99.399	99.416	99.433	99.449	99.458	99.466	99.474	99.482	99.491	99.499
3	34.116	30.817	29.457	28.710	28.237	27.911	27.672	27.489	27.345	27.229	27.052	26.872	26.690	26.598	26.505	26.411	26.316	26.221	26.125
4	21.198	18.000	16.694	15.977	15.522	15.207	14.976	14.799	14.659	14.546	14.374	14.198	14.020	13.929	13.838	13.745	13.652	13.558	13.463
5	16.258	13.274	12.060	11.392	10.967	10.672	10.456	10.289	10.158	10.051	9.888	9.722	9.553	9.466	9.379	9.291	9.202	9.112	9.020
6	13.745	10.925	9.780	9.148	8.746	8.466	8.260	8.102	7.976	7.874	7.718	7.559	7.396	7.313	7.229	7.143	7.057	6.969	6.880
7	12.246	9.547	8.451	7.847	7.460	7.191	6.993	6.840	6.719	6.620	6.469	6.314	6.155	6.074	5.992	5.908	5.824	5.737	5.650
8	11.259	8.649	7.591	7.006	6.632	6.371	6.178	6.029	5.911	5.814	5.667	5.515	5.359	5.279	5.198	5.116	5.032	4.946	4.859
9	10.561	8.022	6.992	6.422	6.057	5.802	5.613	5.467	5.351	5.257	5.111	4.962	4.808	4.729	4.649	4.567	4.483	4.398	4.311
10	10.044	7.559	6.552	5.994	5.636	5.386	5.200	5.057	4.942	4.849	4.706	4.558	4.405	4.327	4.247	4.165	4.082	3.996	3.909
11	9.646	7.206	6.217	5.668	5.316	5.069	4.886	4.744	4.632	4.539	4.397	4.251	4.099	4.021	3.941	3.860	3.776	3.690	3.602
12	9.330	6.927	5.953	5.412	5.064	4.821	4.640	4.499	4.388	4.296	4.155	4.010	3.858	3.780	3.701	3.619	3.535	3.449	3.361
13	9.074	6.701	5.739	5.205	4.862	4.620	4.441	4.302	4.191	4.100	3.960	3.815	3.665	3.587	3.507	3.425	3.341	3.255	3.165
14	8.862	6.515	5.564	5.035	4.695	4.456	4.278	4.140	4.030	3.939	3.800	3.656	3.505	3.427	3.348	3.266	3.181	3.094	3.004
15	8.683	6.359	5.417	4.893	4.556	4.318	4.142	4.004	3.895	3.805	3.666	3.522	3.372	3.294	3.214	3.132	3.047	2.959	2.868
16	8.531	6.226	5.292	4.773	4.437	4.202	4.026	3.890	3.780	3.691	3.553	3.409	3.259	3.181	3.101	3.018	2.933	2.845	2.753
17	8.400	6.112	5.185	4.669	4.336	4.102	3.927	3.791	3.682	3.593	3.455	3.312	3.162	3.084	3.003	2.920	2.835	2.746	2.653
18	8.285	6.013	5.092	4.579	4.248	4.015	3.841	3.705	3.597	3.508	3.371	3.227	3.077	2.999	2.919	2.835	2.749	2.660	2.566
19	8.185	5.926	5.010	4.500	4.171	3.939	3.765	3.631	3.523	3.434	3.297	3.153	3.003	2.925	2.844	2.761	2.674	2.584	2.489
20	8.096	5.849	4.938	4.431	4.103	3.871	3.699	3.564	3.457	3.368	3.231	3.088	2.938	2.859	2.778	2.695	2.608	2.517	2.421
21	8.017	5.780	4.874	4.369	4.042	3.812	3.640	3.506	3.398	3.310	3.173	3.030	2.880	2.801	2.720	2.636	2.548	2.457	2.360
22	7.945	5.719	4.817	4.313	3.988	3.758	3.587	3.453	3.346	3.258	3.121	2.978	2.827	2.749	2.667	2.583	2.495	2.403	2.305
23	7.881	5.664	4.765	4.264	3.939	3.710	3.539	3.406	3.299	3.211	3.074	2.931	2.781	2.702	2.620	2.535	2.447	2.354	2.256
24	7.823	5.614	4.718	4.218	3.895	3.667	3.496	3.363	3.256	3.168	3.032	2.889	2.738	2.659	2.577	2.492	2.403	2.310	2.211

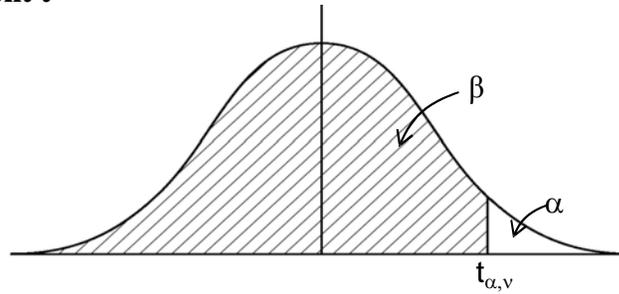
25	7.770	5.568	4.675	4.177	3.855	3.627	3.457	3.324	3.217	3.129	2.993	2.850	2.699	2.620	2.538	2.453	2.364	2.270	2.169
26	7.721	5.526	4.637	4.140	3.818	3.591	3.421	3.288	3.182	3.094	2.958	2.815	2.664	2.585	2.503	2.417	2.327	2.233	2.131
27	7.677	5.488	4.601	4.106	3.785	3.558	3.388	3.256	3.149	3.062	2.926	2.783	2.632	2.552	2.470	2.384	2.294	2.198	2.097
28	7.636	5.453	4.568	4.074	3.754	3.528	3.358	3.226	3.120	3.032	2.896	2.753	2.602	2.522	2.440	2.354	2.263	2.167	2.064
29	7.598	5.420	4.538	4.045	3.725	3.499	3.330	3.198	3.092	3.005	2.868	2.726	2.574	2.495	2.412	2.325	2.234	2.138	2.034
30	7.562	5.390	4.510	4.018	3.699	3.473	3.304	3.173	3.067	2.979	2.843	2.700	2.549	2.469	2.386	2.299	2.208	2.111	2.006
40	7.314	5.179	4.313	3.828	3.514	3.291	3.124	2.993	2.888	2.801	2.665	2.522	2.369	2.288	2.203	2.114	2.019	1.917	1.805
60	7.077	4.977	4.126	3.649	3.339	3.119	2.953	2.823	2.718	2.632	2.496	2.352	2.198	2.115	2.028	1.936	1.836	1.726	1.601
120	6.851	4.787	3.949	3.480	3.174	2.956	2.792	2.663	2.559	2.472	2.336	2.192	2.035	1.950	1.860	1.763	1.656	1.533	1.381
inf	6.635	4.605	3.782	3.319	3.017	2.802	2.639	2.511	2.407	2.321	2.185	2.039	1.878	1.791	1.696	1.592	1.473	1.325	1.000

Lampiran 3 Tabel Distribusi Khi Kuadrat



v	α												
	0.995	0.99	0.975	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.025	0.01	0.005
1	7.8794	6.6349	5.0239	3.8415	2.7055	1.3233	0.4549	0.1015	0.0158	0.0039	0.0010	0.0002	0.0000
2	10.5966	9.2103	7.3778	5.9915	4.6052	2.7726	1.3863	0.5754	0.2107	0.1026	0.0506	0.0201	0.0100
3	12.8382	11.3449	9.3484	7.8147	6.2514	4.1083	2.3660	1.2125	0.5844	0.3518	0.2158	0.1148	0.0717
4	14.8603	13.2767	11.1433	9.4877	7.7794	5.3853	3.3567	1.9226	1.0636	0.7107	0.4844	0.2971	0.2070
5	16.7496	15.0863	12.8325	11.0705	9.2364	6.6257	4.3515	2.6746	1.6103	1.1455	0.8312	0.5543	0.4117
6	18.5476	16.8119	14.4494	12.5916	10.6446	7.8408	5.3481	3.4546	2.2041	1.6354	1.2373	0.8721	0.6757
7	20.2777	18.4753	16.0128	14.0671	12.0170	9.0371	6.3458	4.2549	2.8331	2.1673	1.6899	1.2390	0.9893
8	21.9550	20.0902	17.5345	15.5073	13.3616	10.2189	7.3441	5.0706	3.4895	2.7326	2.1797	1.6465	1.3444
9	23.5894	21.6660	19.0228	16.9190	14.6837	11.3888	8.3428	5.8988	4.1682	3.3251	2.7004	2.0879	1.7349
10	25.1882	23.2093	20.4832	18.3070	15.9872	12.5489	9.3418	6.7372	4.8652	3.9403	3.2470	2.5582	2.1559
11	26.7568	24.7250	21.9200	19.6751	17.2750	13.7007	10.3410	7.5841	5.5778	4.5748	3.8157	3.0535	2.6032
12	28.2995	26.2170	23.3367	21.0261	18.5493	14.8454	11.3403	8.4384	6.3038	5.2260	4.4038	3.5706	3.0738
13	29.8195	27.6882	24.7356	22.3620	19.8119	15.9839	12.3398	9.2991	7.0415	5.8919	5.0088	4.1069	3.5650
14	31.3193	29.1412	26.1189	23.6848	21.0641	17.1169	13.3393	10.1653	7.7895	6.5706	5.6287	4.6604	4.0747
15	32.8013	30.5779	27.4884	24.9958	22.3071	18.2451	14.3389	11.0365	8.5468	7.2609	6.2621	5.2293	4.6009
16	34.2672	31.9999	28.8454	26.2962	23.5418	19.3689	15.3385	11.9122	9.3122	7.9616	6.9077	5.8122	5.1422
17	35.7185	33.4087	30.1910	27.5871	24.7690	20.4887	16.3382	12.7919	10.0852	8.6718	7.5642	6.4078	5.6972
18	37.1565	34.8053	31.5264	28.8693	25.9894	21.6049	17.3379	13.6753	10.8649	9.3905	8.2307	7.0149	6.2648
19	38.5823	36.1909	32.8523	30.1435	27.2036	22.7178	18.3377	14.5620	11.6509	10.1170	8.9065	7.6327	6.8440
20	39.9968	37.5662	34.1696	31.4104	28.4120	23.8277	19.3374	15.4518	12.4426	10.8508	9.5908	8.2604	7.4338
21	41.4011	38.9322	35.4789	32.6706	29.6151	24.9348	20.3372	16.3444	13.2396	11.5913	10.2829	8.8972	8.0337
22	42.7957	40.2894	36.7807	33.9244	30.8133	26.0393	21.3370	17.2396	14.0415	12.3380	10.9823	9.5425	8.6427
23	44.1813	41.6384	38.0756	35.1725	32.0069	27.1413	22.3369	18.1373	14.8480	13.0905	11.6886	10.1957	9.2604
24	45.5585	42.9798	39.3641	36.4150	33.1962	28.2412	23.3367	19.0373	15.6587	13.8484	12.4012	10.8564	9.8862
25	46.9279	44.3141	40.6465	37.6525	34.3816	29.3389	24.3366	19.9393	16.4734	14.6114	13.1197	11.5240	10.5197
26	48.2899	45.6417	41.9232	38.8851	35.5632	30.4346	25.3365	20.8434	17.2919	15.3792	13.8439	12.1981	11.1602
27	49.6449	46.9629	43.1945	40.1133	36.7412	31.5284	26.3363	21.7494	18.1139	16.1514	14.5734	12.8785	11.8076
28	50.9934	48.2782	44.4608	41.3371	37.9159	32.6205	27.3362	22.6572	18.9392	16.9279	15.3079	13.5647	12.4613
29	52.3356	49.5879	45.7223	42.5570	39.0875	33.7109	28.3361	23.5666	19.7677	17.7084	16.0471	14.2565	13.1211
30	53.6720	50.8922	46.9792	43.7730	40.2560	34.7997	29.3360	24.4776	20.5992	18.4927	16.7908	14.9535	13.7867
40	66.7660	63.6907	59.3417	55.7585	51.8051	45.6160	39.3353	33.6603	29.0505	26.5093	24.4330	22.1643	20.7065
50	79.4900	76.1539	71.4202	67.5048	63.1671	56.3336	49.3349	42.9421	37.6886	34.7643	32.3574	29.7067	27.9907
80	116.3211	112.3288	106.6286	101.8795	96.5782	88.1303	79.3343	71.1445	64.2778	60.3915	57.1532	53.5401	51.1719
100	140.1695	135.8067	129.5612	124.3421	118.4980	109.1412	99.3341	90.1332	82.3581	77.9295	74.2219	70.0649	67.3276

Lampiran 3
Tabel Distribusi Student t



v	β									
	0.995	0.99	0.975	0.95	0.9	0.8	0.75	0.7	0.6	0.55
1	63.6567	31.8205	12.7062	6.3138	3.0777	1.3764	1.0000	0.7265	0.3249	0.1584
2	9.9248	6.9646	4.3027	2.9200	1.8856	1.0607	0.8165	0.6172	0.2887	0.1421
3	5.8409	4.5407	3.1824	2.3534	1.6377	0.9785	0.7649	0.5844	0.2767	0.1366
4	4.6041	3.7469	2.7764	2.1318	1.5332	0.9410	0.7407	0.5686	0.2707	0.1338
5	4.0321	3.3649	2.5706	2.0150	1.4759	0.9195	0.7267	0.5594	0.2672	0.1322
6	3.7074	3.1427	2.4469	1.9432	1.4398	0.9057	0.7176	0.5534	0.2648	0.1311
7	3.4995	2.9980	2.3646	1.8946	1.4149	0.8960	0.7111	0.5491	0.2632	0.1303
8	3.3554	2.8965	2.3060	1.8595	1.3968	0.8889	0.7064	0.5459	0.2619	0.1297
9	3.2498	2.8214	2.2622	1.8331	1.3830	0.8834	0.7027	0.5435	0.2610	0.1293
10	3.1693	2.7638	2.2281	1.8125	1.3722	0.8791	0.6998	0.5415	0.2602	0.1289
11	3.1058	2.7181	2.2010	1.7959	1.3634	0.8755	0.6974	0.5399	0.2596	0.1286
12	3.0545	2.6810	2.1788	1.7823	1.3562	0.8726	0.6955	0.5386	0.2590	0.1283
13	3.0123	2.6503	2.1604	1.7709	1.3502	0.8702	0.6938	0.5375	0.2586	0.1281
14	2.9768	2.6245	2.1448	1.7613	1.3450	0.8681	0.6924	0.5366	0.2582	0.1280
15	2.9467	2.6025	2.1314	1.7531	1.3406	0.8662	0.6912	0.5357	0.2579	0.1278
16	2.9208	2.5835	2.1199	1.7459	1.3368	0.8647	0.6901	0.5350	0.2576	0.1277
17	2.8982	2.5669	2.1098	1.7396	1.3334	0.8633	0.6892	0.5344	0.2573	0.1276
18	2.8784	2.5524	2.1009	1.7341	1.3304	0.8620	0.6884	0.5338	0.2571	0.1274
19	2.8609	2.5395	2.0930	1.7291	1.3277	0.8610	0.6876	0.5333	0.2569	0.1274
20	2.8453	2.5280	2.0860	1.7247	1.3253	0.8600	0.6870	0.5329	0.2567	0.1273
21	2.8314	2.5176	2.0796	1.7207	1.3232	0.8591	0.6864	0.5325	0.2566	0.1272
22	2.8188	2.5083	2.0739	1.7171	1.3212	0.8583	0.6858	0.5321	0.2564	0.1271
23	2.8073	2.4999	2.0687	1.7139	1.3195	0.8575	0.6853	0.5317	0.2563	0.1271
24	2.7969	2.4922	2.0639	1.7109	1.3178	0.8569	0.6848	0.5314	0.2562	0.1270
25	2.7874	2.4851	2.0595	1.7081	1.3163	0.8562	0.6844	0.5312	0.2561	0.1269
26	2.7787	2.4786	2.0555	1.7056	1.3150	0.8557	0.6840	0.5309	0.2560	0.1269
27	2.7707	2.4727	2.0518	1.7033	1.3137	0.8551	0.6837	0.5306	0.2559	0.1268
28	2.7633	2.4671	2.0484	1.7011	1.3125	0.8546	0.6834	0.5304	0.2558	0.1268
29	2.7564	2.4620	2.0452	1.6991	1.3114	0.8542	0.6830	0.5302	0.2557	0.1268
30	2.7500	2.4573	2.0423	1.6973	1.3104	0.8538	0.6828	0.5300	0.2556	0.1267
40	2.7045	2.4233	2.0211	1.6839	1.3031	0.8507	0.6807	0.5286	0.2550	0.1265
50	2.6778	2.4033	2.0086	1.6759	1.2987	0.8489	0.6794	0.5278	0.2547	0.1263
80	2.6387	2.3739	1.9901	1.6641	1.2922	0.8461	0.6776	0.5265	0.2542	0.1261
100	2.6259	2.3642	1.9840	1.6602	1.2901	0.8452	0.6770	0.5261	0.2540	0.1260
200	2.6006	2.3451	1.9719	1.6525	1.2858	0.8434	0.6757	0.5252	0.2537	0.1258
300	2.5923	2.3388	1.9679	1.6499	1.2844	0.8428	0.6753	0.5250	0.2536	0.1258
400	2.5882	2.3357	1.9659	1.6487	1.2837	0.8425	0.6751	0.5248	0.2535	0.1257
600	2.5840	2.3326	1.9639	1.6474	1.2830	0.8422	0.6749	0.5247	0.2535	0.1257
1000	2.5808	2.3301	1.9623	1.6464	1.2824	0.8420	0.6747	0.5246	0.2534	0.1257

Tabel Distribusi t