

STATISTIK KONTROL KUALITAS

Teori dan Aplikasi dalam
Analisis Data Industri

Ilmu statistik bukan hanya sekadar kumpulan angka, rumus, dan grafik. Statistik ialah bahasa yang digunakan untuk memahami ketidakpastian, menafsirkan variasi, serta mengubah data menjadi informasi berharga untuk pengambilan keputusan yang bermakna. Dalam dunia industri yang bergerak cepat dan kompetitif, kemampuan untuk mengetahui kapan sesuatu berjalan baik, dan kapan sesuatu mulai menyimpang menjadi kunci keberhasilan. Di sinilah **Statistik dan Kontrol Kualitas** mengambil peran penting.

Buku ini memadukan konsep dasar statistik, statistika deskriptif, probabilitas dan distribusi probabilitas baik diskrit maupun kontinu, seperti distribusi Binomial, Poisson, dan Normal, yang kemudian menjadi landasan bagi pembahasan pendugaan interval, uji hipotesis, analisis regresi, dan korelasi. Pada bagian akhir, pembaca akan diajak memahami dan menerapkan berbagai grafik kendali untuk variabel, sebagai alat utama dalam menjaga stabilitas dan mutu proses produksi. Semoga kehadiran buku ini tidak hanya membantu mahasiswa memahami materi perkuliahan, tetapi juga menumbuhkan rasa ingin tahu dan kesadaran bahwa setiap angka memiliki cerita di baliknya.



Penerbit Haura Utama
Pusat Haura Utama
Jalan Raya
No. 100
Telp. 021-786342
Email: haura@haura.com



Sofia Debi Puspa, S.Pd., M.Si., dkk.

STATISTIK KONTROL KUALITAS: Teori dan Aplikasi dalam Analisis Data Industri



Sofia Debi Puspa, S.Pd., M.Si.
Drs. Joko Riyono, M.Si.
Dra. Christina Eni Pujiastuti, M.Si.
Aina Latifa Riyana Putri, S.Si., M.Mat.

STATISTIK KONTROL KUALITAS

Teori dan Aplikasi dalam
Analisis Data Industri



STATISTIK KONTROL KUALITAS

Teori dan Aplikasi dalam
Analisis Data Industri

Sofia Debi Puspa, S.Pd., M.Si.
Drs. Joko Riyono, M.Si.
Dra. Christina Eni Pujiastuti, M.Si.
Aina Latifa Riyana Putri, S.Si., M.Mat.



Haura Utama

PRAKATA

Ilmu statistik bukan hanya sekadar kumpulan angka, rumus, dan grafik. Statistik ialah bahasa yang digunakan untuk memahami ketidakpastian, menafsirkan variasi, serta mengubah data menjadi informasi berharga untuk pengambilan keputusan yang bermakna. Dalam dunia industri yang bergerak cepat dan kompetitif, kemampuan untuk mengetahui kapan sesuatu berjalan baik, dan kapan sesuatu mulai menyimpang menjadi kunci keberhasilan. Di sinilah **Statistik dan Kontrol Kualitas** mengambil peran penting.

Segala puji bagi Allah SWT atas rahmat dan kasih sayang-Nya, buku Statistik dan Kontrol Kualitas dapat hadir sebagai buku ajar yang menghubungkan teori statistik dengan penerapannya dalam dunia nyata. Buku ini memadukan konsep dasar statistik, statistika dekritif, probabilitas dan distribusi probabilitas baik diskrit maupun kontinu, seperti distribusi Binomial, Poisson, dan Normal, yang kemudian menjadi landasan bagi pembahasan pendugaan interval, uji hipotesis, analisis regresi, dan korelasi. Pada bagian akhir, pembaca akan diajak memahami dan menerapkan berbagai grafik kendali untuk variabel, sebagai alat utama dalam menjaga stabilitas dan mutu proses produksi. Semoga kehadiran buku ini tidak hanya membantu mahasiswa memahami materi perkuliahan, tetapi juga menumbuhkan rasa ingin tahu dan kesadaran bahwa setiap angka memiliki cerita di baliknya.

Akhir kata, penulis mengucapkan terima kasih kepada semua pihak yang telah memberikan dukungan dan kontribusinya dalam penyusunan Buku Statistik dan Kontrol Kualitas. Penulis menyadari buku ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang membangun diharapkan demi penyempurnaan edisi selanjutnya. Semoga buku ini dapat bermanfaat bagi mahasiswa, peneliti, dosen, dan seluruh pembaca yang memiliki perhatian pada bidang statistika dan pengendalian mutu.

Jakarta, Oktober 2025

Penulis

Statistik dan Kontrol Kualitas; Teori dan Aplikasi dalam Analisis data Industri,
karya Sofia Debi Puspa, S.Pd., M.Si., dkk.
diterbitkan pertama kali oleh Penerbit Haura Utama, 2025

16.8 x 23.5 cm, 245 hlm

Hak cipta dilindungi undang-undang
Dilarang mereproduksi atau memperbanyak seluruh
maupun sebagian dari buku ini dalam bentuk dan
cara apapun tanpa izin tertulis dari penerbit

Editor dan Penata isi: Salsa
Perancang sampul: Nita



CV. Haura Utama

Anggota IKAPI Nomor 375/JBA/2020
Nagrak, Benteng, Warudoyong, Sukabumi
+62877-8193-0045 haurautama@gmail.com

Cetakan I, November 2025

ISBN: 978-634-208-424-3



DAFTAR ISI

PRAKATA.....	3
DAFTAR ISI.....	4
BAB 1 Pengantar Statistik	7
A. Beberapa Istilah Dasar	8
B. Tipe Data dan Sumber Data	10
C. Level Pengukuran	13
D. Teknik Pengambilan Data (Metode Sampling).....	15
E. Visualisasi Data	18
LATIHAN 1	24
BAB 2 Statistika Deskriptif.....	27
A. Ukuran Pemusatan Data	28
B. Modus	28
C. Rata-Rata (Mean)	29
D. Median	29
E. Ukuran Penyebaran Data	33
F. Jangkauan.....	33
G. Kuartil dan Jangkauan Interkuartil.....	34
H. Varians dan Standar Deviasi.....	35
I. Pengambilan Sampel dari Populasi.....	36
J. Standar Deviasi Untuk Data Kelompok.....	38
K. Pencilan (<i>Outliers</i>)	38
L. Boxplot.....	39
M. Grafik Frekuensi Kumulatif.....	42
LATIHAN 2	45
BAB 3 Probabilitas.....	49
A. Percobaan Statistika dan Kejadian	50
B. Teori Probabilitas	50
C. Diagram Pohon	53
D. Kejadian Majemuk (<i>Compound Event</i>).....	54
E. Himpunan dan Diagram Venn	57
F. Teorema Bayes (<i>Bayes Theorem</i>)	59
LATIHAN 3	63

BAB 4 Distribusi Probabilitas Variabel Acak Diskrit dan Kontinu ...	66
A. Variabel Acak	67
B. Distribusi Probabilitas Diskrit.....	68
C. Ekspektasi Variabel Acak Diskrit.....	70
D. Varians Variabel Acak Diskrit.....	71
E. Transformasi Linear Pada Variabel Acak.....	72
F. Distribusi Probabilitas Kontinu.....	72
LATIHAN 4	74
BAB 5 Distribusi Binomial & Poisson.....	77
A. Distribusi Binomial	78
B. Mean dan Standar Deviasi Distribusi Binomial.....	81
C. Distribusi Poisson	82
LATIHAN 5	86
BAB 6 Distribusi Normal	89
A. Distribusi Normal.....	90
B. Distribusi Normal Baku (Distribusi-Z).....	94
C. Kuantil (<i>Quantiles</i>) Pada Distribusi Normal.....	96
LATIHAN 6	100
BAB 7 Pendugaan Interval.....	103
A. Pendugaan Interval Rata-Rata μ	104
B. Pendugaan Interval Proporsi (Persentase) p	107
C. Pendugaan Interval Selisih Dua Rata-Rata $\mu_1 - \mu_2$	109
D. Pendugaan Interval Selisih Dua Proporsi $P_1 - P_2$	111
LATIHAN 7	113
BAB 8 Pengujian Hipotesis	115
A. Sampel dan Populasi	116
B. Definisi Hipotesis.....	117
C. Jenis Kesalahan	119
D. Pengujian Hipotesis Tentang Rata-Rata μ	121
E. Pengujian Hipotesis Tentang Proporsi	127
F. Pengujian Hipotesis Tentang Varians	130
G. Analisis Data Tentang Pengujian Hipotesis.....	133
LATIHAN 8	147
BAB 9 Analisis Korelasi & Regresi	149
A. Regresi Linear Sederhana	150

B. Regresi Linear Berganda.....	167
C. Regresi Logistik	181
LATIHAN 9	188
BAB 10 Peta Kendali Untuk Variabel.....	192
A. Pendahuluan	193
B. Peta Kendali $\bar{x}$ dan R	194
C. Estimasi Kapabilitas Proses (C_p)	201
D. Indeks Kapabilitas Proses Kritis (C_{pk}).....	205
E. Peta Kendali $\bar{x}$ dan s	206
F. Peta Kendali Individual (I -Chart) dan Moving Range (MR -Chart)	212
LATIHAN 10	217
DAFTAR PUSTAKA	221
GLOSARIUM.....	224
LAMPIRAN.....	226
BIOGRAFI PENULIS	244

BAB 1

Pengantar Statistik

Pendahuluan

Kemampuan Akhir yang Diharapkan	Mahasiswa memahami berbagai istilah statistika, jenis data, level pengukuran, tipe data, sumber data, teknik pengambilan data, dan visualisasi data
Bahan Kajian	- Istilah dasar statistika - Tipe data dan sumber data - Teknik pengambilan data - Visualisasi data
Metode Pembelajaran	Ceramah interaktif, studi kasus
Pengalaman Belajar Mahasiswa	Mahasiswa mengidentifikasi istilah dan konsep dasar statistika melalui studi literatur dan diskusi kelas, mengklasifikasikan tipe data dan sumber data dari contoh nyata merancang teknik pengumpulan data sederhana, serta menggunakan perangkat lunak statistik untuk membuat visualisasi data
Kriteria Penilaian (Indikator)	- Mampu menjelaskan perbedaan antara populasi dan sampel, parameter dan statistik, serta variabel. - Mampu mengidentifikasi jenis data, level pengukuran, tipe data dan sumber data dengan benar. - Menyusun teknik pengambilan data sesuai jenis data yang dibutuhkan. - Membuat grafik yang sesuai dengan tipe data dan mampu menafsirkan informasi yang ditampilkan.

Tujuan Pembelajaran

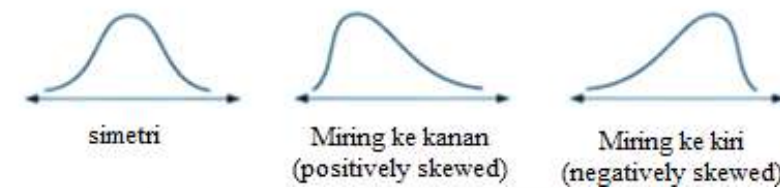
1. Mahasiswa mampu menjelaskan berbagai istilah dasar, mengidentifikasi jenis data (kualitatif dan kuantitatif) serta level pengukuran.
2. Mahasiswa mampu memahami tipe data, sumber data, serta teknik pengumpulan data.
3. Mahasiswa mampu menyajikan visualisasi dari suatu kumpulan data

A. Beberapa Istilah Dasar

Dalam kehidupan sehari-hari, istilah "statistik" menjadi bagian yang tidak asing. Bahkan, di Indonesia terdapat lembaga resmi bernama Badan Pusat Statistik (BPS). Selain itu, kita sering mendengar istilah seperti "observasi," "data," "sensus," "sampel," "populasi," dan lainnya. Selain "statistik," ada pula istilah "statistika," seperti yang digunakan dalam judul bab ini. Berikut penjelasan dari beberapa istilah tersebut:

- **Statistika** merujuk pada serangkaian metode yang digunakan untuk merancang eksperimen, mengumpulkan data, serta menyusun, meringkas, menyajikan, menganalisis, menginterpretasikan, dan mengambil kesimpulan berdasarkan data tersebut.
- **Data** adalah hasil pengamatan atau observasi yang telah dikumpulkan. Data dapat berupa hasil pengukuran, seperti tinggi dan berat badan; hasil pengelompokan, seperti jenis kelamin; atau jawaban responden terhadap kuesioner, misalnya tingkat kepuasan.
- **Populasi** merupakan kumpulan lengkap dari semua elemen yang menjadi objek penelitian. Suatu populasi dianggap lengkap jika mencakup seluruh subjek yang diselidiki.
- **Sensus** adalah proses pengumpulan data dari seluruh anggota dalam suatu populasi.
- **Sampel** adalah sebagian anggota yang dipilih dari populasi untuk dianalisis lebih lanjut.

- **Distribusi** adalah pola penyebaran data, yang bisa digambarkan sebagai berikut:



Gambar 1.1. Jenis Distribusi

- **Pencilan (Outliers)** – nilai data yang jauh lebih besar atau lebih kecil dibandingkan dengan data lainnya. Nilai-nilai ini tetap bisa dianalisis, kecuali jika terbukti berasal dari kesalahan manusia atau kesalahan teknis lainnya.
- **Statistika deskriptif** adalah cabang statistika yang berfokus pada analisis dan penyajian data suatu kelompok tanpa membuat generalisasi terhadap kelompok lain yang lebih luas.
- **Statistika Inferensial** adalah cabang statistika yang bertujuan untuk membuat kesimpulan atau deduksi tentang populasi berdasarkan analisis sampel.

Sebagai contoh, sebuah lembaga survei dapat mewawancarai 2.350 penduduk Indonesia untuk mengukur tingkat kepuasan terhadap kinerja pemerintah. Dalam hal ini, 2.350 penduduk tersebut adalah sampel, sementara seluruh penduduk Indonesia yang berjumlah sekitar 230 juta jiwa adalah populasinya. Pemerintah sendiri biasanya melakukan sensus ekonomi atau pertanian setiap lima tahun sekali, di mana semua kepala keluarga didata, dan hasilnya disebut sebagai data sensus.

Proses pengumpulan data melalui sensus memerlukan biaya, waktu, dan tenaga yang besar. Untuk efisiensi, perilaku populasi sering kali cukup dipelajari melalui sampelnya saja. Analisis yang dilakukan pada sampel ini kemudian digunakan untuk menyimpulkan karakteristik populasi. Namun, agar hasil kesimpulan dapat dipercaya, sampel yang diambil harus

representatif, yaitu benar-benar mencerminkan populasi. Jika tidak, kesimpulan yang dihasilkan bisa menyimpang dari kenyataan.

Pemilihan umum (pemilu) atau pemilihan kepala daerah (pilkada) di Indonesia merupakan proses untuk mengumpulkan aspirasi dari seluruh pemilih, sehingga dapat dianggap sebagai bentuk sensus untuk populasi pemilih. Namun, kenyataannya tidak semua data dapat diperoleh karena ada pemilih yang tidak menggunakan hak suara, atau dikenal sebagai "golput." Sebaliknya, lembaga survei yang melakukan hitung cepat atau *quick count* menggunakan pendekatan sampling, di mana data hanya diambil dari sebagian tempat pemungutan suara (TPS) yang dipilih secara strategis agar hasilnya dapat dianggap mewakili seluruh pemilih. Proses ini menghasilkan data lebih cepat karena hanya melibatkan sebagian kecil dari keseluruhan populasi. Keakuratan hasil tergantung pada kualitas sampel yang diambil dan metode analisis data yang digunakan.

Perlu diingat, dalam sistem sampling terdapat faktor kesalahan yang diperhitungkan sejak awal, seperti *sampling error*, yaitu peluang adanya perbedaan antara sampel dan populasi. Selain itu, metode analisis data dalam statistika selalu didasarkan pada teori probabilitas, sehingga tidak ada kesimpulan yang bersifat mutlak. Selalu ada kemungkinan terjadinya hasil yang berbeda. Hal ini menjelaskan mengapa beberapa lembaga survei bisa memberikan hasil *quick count* yang berbeda, terutama jika selisih suara antar kandidat sangat tipis. Contohnya, kasus pilkada Jawa Timur beberapa waktu lalu menunjukkan hal tersebut.

B. Tipe Data dan Sumber Data

Pada pembahasan sebelumnya, kita telah menjelaskan definisi dari sampel dan populasi. Perbedaan antara keduanya terletak pada cakupan observasi yang dilakukan. Dalam analisis data, untuk membedakan informasi yang berasal dari sampel dan populasi, digunakan dua istilah utama, yaitu **parameter** dan **statistik**.

- **Parameter**

Parameter adalah nilai numerik yang digunakan untuk mendeskripsikan suatu karakteristik populasi secara keseluruhan. Karena berasal dari

populasi, parameter sering dianggap sebagai nilai tetap (konstan) yang mewakili data aktual.

Contoh penggunaan parameter: Berdasarkan hasil sensus ekonomi tahun 2010, ditemukan bahwa 35% rumah tangga di Indonesia berada dalam kategori miskin. Nilai **35%** ini merupakan **parameter**, karena diperoleh dari populasi, yaitu seluruh rumah tangga di Indonesia.

- **Statistik**

Statistik adalah nilai numerik yang digunakan untuk menggambarkan karakteristik suatu sampel. Statistik diperoleh dari sebagian kecil populasi yang digunakan sebagai representasi. Oleh karena itu, statistik dapat berubah tergantung pada sampel yang diambil.

Contoh penggunaan statistik: Hasil survei terhadap 50 mahasiswa Program Studi Pendidikan Matematika Universitas Muhammadiyah Ponorogo angkatan 2008/2009 menunjukkan rata-rata nilai NEM Matematika mereka adalah 6,75. Nilai **6,75** ini merupakan **statistik**, karena diperoleh dari sampel yang terdiri dari 50 mahasiswa, bukan seluruh mahasiswa.

Selain data numerik seperti contoh di atas, data juga dapat berbentuk kategori. Data dibedakan menjadi dua jenis utama, yaitu **data kuantitatif** dan **data kualitatif**.

- **Data kuantitatif**

Data kuantitatif adalah data yang menunjukkan hasil perhitungan atau pengukuran tertentu. Jenis data ini selalu dinyatakan dalam bentuk angka dan digunakan untuk menganalisis fenomena yang bersifat numerik.

Contoh data kuantitatif adalah tinggi badan seseorang (dalam cm atau meter), nilai ujian (misalnya, NEM), temperatur dalam satuan derajat Celsius, pendapatan bulanan (dalam rupiah atau dolar), dan lain-lain.

- **Data kualitatif**

Data kualitatif juga dikenal sebagai data kategori, adalah data yang diklasifikasikan berdasarkan karakteristik tertentu yang tidak dinyatakan dalam bentuk angka. Data ini biasanya menggambarkan kelompok atau kategori tertentu.

Contoh data kualitatif adalah jenis kelamin (laki-laki atau perempuan), profesi (guru, dokter, atau insinyur), temperatur dalam deskripsi (sejuk, dingin, atau panas), dan lain-lain.

Adapun klasifikasi data kuantitatif dapat dikelompokkan lebih lanjut menjadi dua jenis, yaitu **data diskrit** dan **data kontinu**.

- **Data diskrit**

Data diskrit memiliki nilai yang terbatas dan dapat dihitung secara eksplisit. Biasanya, data diskrit terdiri dari bilangan bulat dan tidak memiliki nilai di antara dua angka tertentu. Contoh data diskrit adalah jumlah rumah yang tejual, jumlah telur yang dihasilkan oleh ayam, jumlah hari libur dalam sebulan, dan lain-lain.

- **Data kontinu**

Data kontinu memiliki nilai yang tidak terbatas dalam suatu interval tertentu dan mencakup semua nilai yang mungkin, termasuk pecahan atau desimal.

Contoh data kontinu adalah temperatur udara di berbagai lokasi (misalnya, -20°C hingga 50°C), berat badan seseorang (dalam kilogram, misalnya 65,7 kg), dan waktu yang dibutuhkan untuk menyelesaikan suatu tugas (dalam menit atau detik).

Dengan memahami perbedaan antara parameter dan statistik, serta jenis-jenis data seperti kuantitatif dan kualitatif, kita dapat memilih metode analisis yang tepat untuk mengolah data sesuai dengan sifatnya. Hal ini sangat penting untuk mendapatkan hasil yang akurat dan relevan dalam berbagai penelitian atau studi.

Selanjutnya, sebelum melakukan analisis statistik, kita perlu memahami dari mana data diperoleh dan bagaimana data tersebut dikumpulkan. Ini penting untuk menilai apakah data tersebut relevan, valid, dan dapat dipercaya. Sumber data terbagi menjadi dua jenis yaitu data primer dan data sekunder yang memiliki karakteristik, kelebihan dan kekurangan masing-masing.

- **Data Primer**

Data yang dikumpulkan langsung oleh peneliti dari sumber aslinya untuk tujuan penelitian tertentu. Contoh data primer adalah hasil wawancara atau kuesioner yang dilakukan sendiri, pengukuran langsung di lapangan (misalnya suhu mesin, waktu produksi, dll), observasi atau eksperimen langsung di laboratorium atau bengkel, dan lain-lain. Contoh data primer di bidang teknik seperti mengukur getaran mesin menggunakan sensor, mencatat suhu operasi mesin diesel dalam waktu tertentu, dan lain-lain. Keunggulan data primer adalah lebih akurat dan sesuai kebutuhan, sedangkan kelemahannya adalah memerlukan waktu dan biaya yang besar.

- **Data Sekunder**

Data yang telah dikumpulkan oleh pihak lain, dan digunakan kembali oleh peneliti. Contoh data dari laporan produksi Perusahaan, data bersumber dari Badan Pusat Statistik (BPS), data dari jurnal ilmiah, buku, atau publikasi penelitian sebelumnya, dan lain-lain. Contoh data sekunder dalam teknik yaitu menggunakan data efisiensi bahan bakar dari katalog mesin, memakai data hasil uji coba alat dari jurnal atau laboratorium lain, dan sebagainya. Keunggulan data sekunder yaitu lebih cepat dan mudah diakses, sedangkan data sekunder kelemahannya bisa kurang relevan atau tidak lengkap.

C. Level Pengukuran

Data dapat diklasifikasikan berdasarkan empat level pengukuran, yaitu nominal, ordinal, interval, dan rasio. Dalam statistika terapan, pemahaman terhadap level pengukuran data sangat penting karena memengaruhi metode analisis yang digunakan.

- **Level nominal**

Level nominal mencakup data yang berupa nama, label, atau kategori. Pada level ini, data tidak memiliki urutan atau hierarki tertentu sehingga tidak dapat disusun dari yang tertinggi ke terendah atau sebaliknya. Contoh data nominal adalah:

- a. Pilihan jawaban dalam kuesioner, seperti "Ya," "Tidak," dan "Tidak Tahu".
- b. Warna mobil yang dimiliki dosen di Universitas Muhammadiyah Ponorogo, seperti hitam, merah, putih, atau biru.

Data pada level nominal tidak bisa digunakan untuk operasi matematis seperti penjumlahan atau pengurangan. Sebagai contoh, menambahkan "Ya" dan "Tidak" atau "Merah" dan "Hitam" tidak memiliki arti matematis.

- **Level ordinal**

Level ordinal mencakup data yang dapat disusun dalam suatu urutan atau ranking. Namun, selisih antara data tidak bermakna atau tidak dapat diukur secara pasti. Contoh data ordinal adalah:

- a. Nilai akhir dalam Kartu Hasil Studi (KHS) seperti E, D, C, B-, B, A-, dan A. Nilai ini dapat diurutkan, misalnya A lebih tinggi dari B, tetapi perbedaan antara keduanya tidak dapat dihitung secara pasti.
- b. Peringkat Indeks Persepsi Korupsi (IPK) untuk 50 kota di Indonesia. Misalnya, Yogyakarta di peringkat pertama, Palangkaraya di peringkat kedua, dan Banda Aceh di peringkat ketiga. Selisih peringkat seperti $2-1=1$ tidak bermakna dalam konteks ini.

- **Level Interval**

Level interval serupa dengan level ordinal, tetapi memiliki tambahan sifat yaitu selisih antara dua data memiliki makna yang jelas. Namun,

level ini tidak memiliki titik nol alami sebagai titik awal. Contoh Data Interval:

- a. Temperatur: Suhu tubuh 36°C dan 37°C dapat diurutkan, dan selisihnya (1°C) bermakna. Namun, suhu 0°C tidak menunjukkan ketiadaan panas secara absolut.
- b. Tahun: Tahun 1000, 2000, dan 2023 dapat diurutkan, dan selisih antar tahun, seperti $2023 - 2000 = 23$, memiliki makna. Namun, tidak ada titik nol alami untuk waktu; tahun 0 adalah titik nol buatan manusia.

- **Level Rasio**

Level rasio adalah level pengukuran tertinggi. Data pada level ini memiliki sifat-sifat level interval, ditambah dengan keberadaan titik nol alami. Hal ini memungkinkan data dibandingkan dalam bentuk selisih maupun rasio (perbandingan). Contoh Data Rasio:

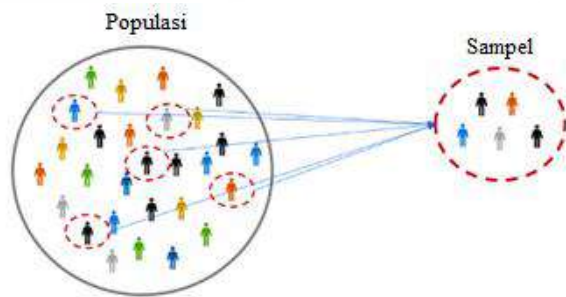
- a. Harga: Harga buku teks mahasiswa, misalnya Rp50.000 atau Rp0, di mana Rp0 berarti tidak ada harga atau gratis.
- b. Berat Badan: Berat badan seseorang, misalnya 70 kg atau 0 kg, di mana 0 kg berarti tidak ada bobot.
- c. Indeks Persepsi Korupsi (IPK): Sebelum diranking, skor IPK menggunakan skala 0-10. Misalnya, Yogyakarta memiliki skor 6,43, Palangkaraya 6,10, dan Kupang 2,97. Dalam hal ini, nilai 0 merepresentasikan kriteria terkorup.

Pemahaman terhadap keempat level pengukuran ini membantu dalam memilih metode statistik yang sesuai. Setiap level memiliki karakteristik yang menentukan jenis operasi matematis atau analisis yang dapat dilakukan pada data tersebut.

D. Teknik Pengambilan Data (Metode Sampling)

Dalam statistik ketika mengumpulkan data dari seluruh populasi tidak memungkinkan karena keterbatasan waktu, biaya, atau tenaga, maka digunakan **sampling**.

Sampling adalah proses memilih sebagian anggota populasi untuk menganalisis data dengan tujuan memperkirakan karakteristik keseluruhan populasi tersebut. Dalam praktiknya, keterbatasan waktu dan sumber daya membuat analisis tidak selalu memungkinkan untuk mengumpulkan data dari setiap individu dalam populasi yang luas. Oleh karena itu, *sampling* digunakan untuk mendapatkan sampel yang representatif, sehingga bisa memberikan informasi akurat mengenai populasi tanpa harus menyelidiki setiap anggotanya.



Gambar 1.2. Populasi dan Sampel

Sampling dibagi menjadi dua kategori besar yaitu sampling probabilitas (*probability sampling*) dan sampling non-probabilitas (*non-probability sampling*).

D.1. Sampling Probabilitas (*Probability Sampling*)

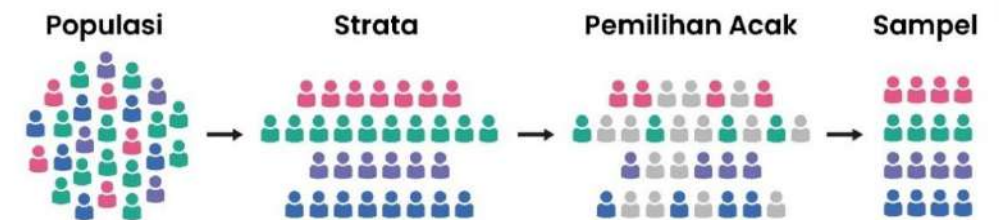
Sampling probabilitas adalah teknik pengambilan data di mana setiap anggota populasi memiliki **peluang yang sama** untuk terpilih. Sampling bisa digunakan untuk menggeneralisir hasil penelitian terhadap populasi karena berasal dari sampel yang mewakili populasi. Jenis-jenis sampling probabilitas yaitu:

a. *Simple Random Sampling* (Sampling Acak Sederhana)

Setiap individu dalam populasi memiliki peluang yang sama dan independen untuk dipilih. Contoh: mengundi nomor karyawan untuk memilih 10 orang sebagai sampel.

b. *Stratified Random Sampling* (Sampling Acak Bertingkat)

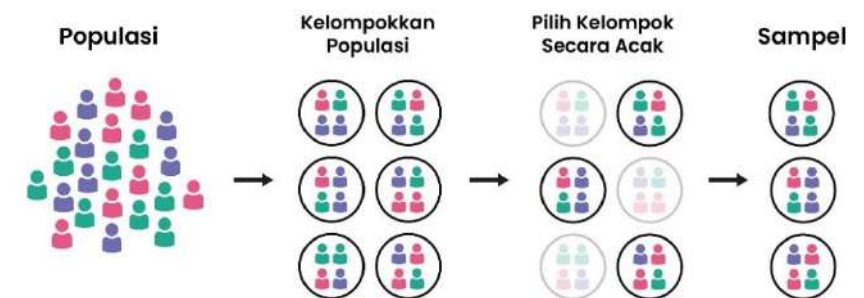
Populasi dibagi ke dalam kelompok (strata) berdasarkan karakteristik tertentu, lalu diambil sampel secara acak dari setiap strata. Contoh: membagi karyawan berdasarkan divisi, lalu ambil sampel acak dari tiap divisi, penelitian dengan sampel pasien covid-19 namun pasien covid-19 diklasifikasi terlebih dahulu misal berdasarkan riwayat penyakit, lokasi tempat tinggal, dll. Kemudian dari masing-masing strata dipilih pasien dengan teknik *simple random sampling*



Gambar 1.3. *Stratified Random Sampling*

c. *Cluster Sampling* (Sampling Klaster)

Populasi dibagi menjadi kelompok (klaster) secara alami, lalu dipilih beberapa klaster secara acak untuk dijadikan sampel. Atau populasi dibagi ke dalam satuan-satuan sampling yang besar, namun tiap cluster harus relatif heterogen. Contoh: memilih beberapa pabrik secara acak untuk mewakili seluruh jaringan pabrik; karena wilayah Indonesia luas, dipilihlah provinsi sebagai klaster ke-1 secara random. Dari tiap provinsi dipilih lagi yaitu kabupaten/kota sebagai klaster ke-2 dst. Kemudian dari masing-masing kabupaten/kota dipilih pasien covid-19 secara random



Gambar 1.4. *Cluster Sampling*

d. *Systematic Sampling* (Sampling Sistematis)

Memilih setiap elemen ke-n dari daftar populasi yang terurut. Contoh: memilih setiap karyawan ke-5 dari daftar absen.

D.2. Sampling Non-Probabilitas (*Non-Probability Sampling*).

Sampling non-probabilitas adalah teknik pengambilan data di mana tidak semua anggota populasi memiliki peluang yang sama untuk dipilih. Artinya pemilihan sampling tidak melibatkan unsur peluang. Sehingga, sampling tidak boleh dipakai untuk menggeneralisasi hasil penelitian terhadap populasi karena tidak ada unsur probabilitas. Jenis-jenis sampling non probabilitas yaitu:

a. *Convenience Sampling* (Sampling Kemudahan)

Mengambil sampel dari siapa saja yang mudah diakses. Contoh: menyurvei karyawan yang kebetulan ada di ruang makan saat istirahat.

b. *Judgmental Sampling* (Sampling Pertimbangan)

Sampel dipilih berdasarkan pertimbangan atau keahlian peneliti. Contoh memilih teknisi senior karena dianggap lebih memahami sistem produksi.

c. *Quota Sampling* (Sampling Kuota)

Menentukan kuota untuk kelompok tertentu, kemudian memilih sampel dari tiap kelompok. Contoh: menentukan akan mengambil 50 karyawan produksi dan 50 karyawan administrasi.

d. *Snowball Sampling*

Sampel pertama memilih sampel berikutnya, dan seterusnya. Contoh Dalam penelitian teknisi ahli, satu teknisi ahli merekomendasikan teknisi lain.

E. Visualisasi Data

E.1. Histogram dan Diagram Batang

Dalam penyajian data, tabel sering kali masih kurang memberikan gambaran yang jelas tentang pola distribusi data. Oleh karena itu, diperlukan

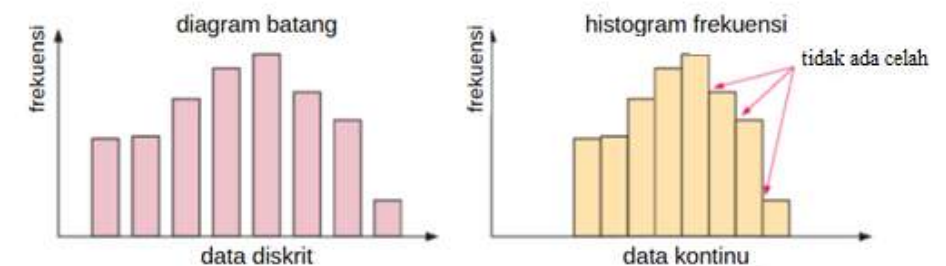
representasi grafis yang lebih mudah dipahami. Dua bentuk grafik yang umum digunakan adalah **diagram batang** dan **histogram**.

Diagram batang biasanya digunakan untuk menampilkan data kategorik atau diskrit, dengan batang-batang terpisah yang mewakili tiap kategori. Sementara itu, histogram digunakan untuk menyajikan data kuantitatif yang berbentuk interval atau rasio, di mana batang-batangnya saling berdempetan untuk menunjukkan distribusi frekuensi dalam kelas-kelas tertentu. Melalui kedua grafik ini, kita dapat dengan cepat melihat pola data, membandingkan kelompok, serta mengidentifikasi kecenderungan atau penyebaran data.

Grafik batang dan histogram frekuensi memiliki ciri-ciri berikut:

- Frekuensi kemunculan ditampilkan pada sumbu vertikal.
- Rentang skor/nilai ditampilkan pada sumbu horizontal.
- Lebar batang sama besar dan tinggi batang menunjukkan frekuensi.

Kelas modus, yaitu kelas atau kelompok nilai yang paling sering muncul, dapat dengan mudah dikenali dari batang tertinggi pada histogram frekuensi.



Gambar 1.5. Grafik Data Diskrit Dan Kontinu

Histogram adalah representasi grafis dari distribusi frekuensi suatu data dalam bentuk persegi panjang. Setiap persegi panjang dalam histogram menggambarkan frekuensi data pada suatu interval kelas tertentu.



Grafik 1.6. Histogram

Berikut adalah elemen-elemen utama dalam histogram:

- **Sumbu X (Horizontal):**

Sumbu X mewakili interval kelas, yaitu rentang nilai data yang telah dibagi ke dalam beberapa kelas. Pada sumbu X, pusat alas dari setiap persegi panjang berfungsi sebagai tanda kelas (titik tengah interval kelas). Lebar persegi panjang atau alas setiap interval kelas biasanya memiliki ukuran yang sama.

- **Sumbu Y (Vertikal):**

Sumbu Y menunjukkan frekuensi atau jumlah data yang terdapat dalam setiap interval kelas. Semakin tinggi suatu persegi panjang, semakin banyak data yang terdapat dalam kelas tersebut.

- **Alas Persegi Panjang (Interval Kelas):**

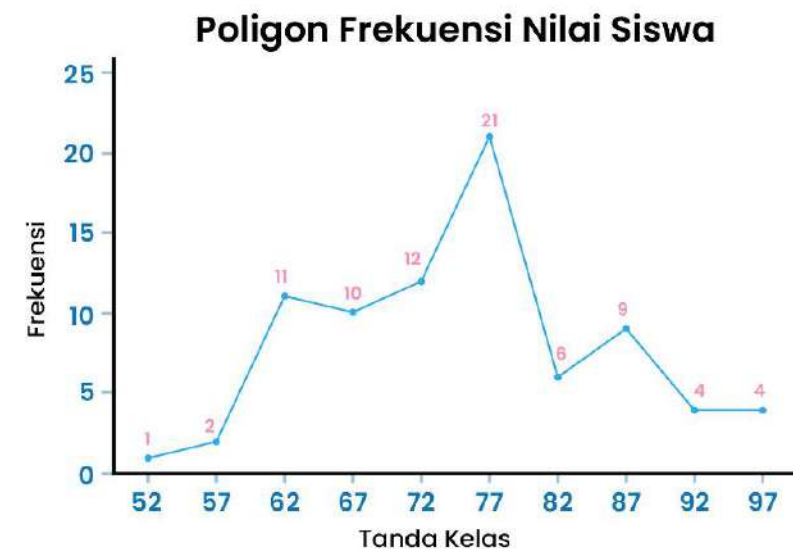
Lebar alas persegi panjang adalah lebar interval kelas yang digunakan dalam distribusi frekuensi. Lebar ini dapat bervariasi tergantung pada data dan jumlah kelas yang ditetapkan. Dalam prakteknya, lebar kelas yang digunakan biasanya seragam untuk memudahkan pembacaan dan analisis.

- **Tinggi Persegi Panjang (Frekuensi):**

Tinggi persegi panjang pada histogram menggambarkan banyaknya data yang termasuk dalam interval kelas tersebut. Dengan demikian, tinggi setiap persegi panjang secara langsung berhubungan dengan jumlah frekuensi data yang jatuh dalam rentang interval kelas tertentu.

E.2. Poligon Frekuensi

Poligon frekuensi adalah grafik garis yang menghubungkan titik-titik yang mewakili frekuensi kelas dengan tanda kelas. Poligon ini digunakan untuk menggambarkan distribusi data dalam bentuk yang lebih halus dibandingkan histogram.



Gambar 1.7. Poligon Frekuensi

Berikut adalah elemen-elemen utama dalam poligon frekuensi:

- **Tanda Kelas:**

Tanda kelas adalah titik tengah dari setiap interval kelas. Tanda kelas dihitung menggunakan rumus:

$$\text{Tanda Kelas} = \frac{\text{Limit Bawah} + \text{Limit Atas}}{2}$$

Tanda kelas ini digunakan sebagai titik di sepanjang sumbu X, yang mewakili posisi tengah dari interval kelas.

- **Pasangan Titik Frekuensi dan Tanda Kelas:**

Setiap titik dalam poligon mewakili frekuensi kelas pada tanda kelas tersebut. Titik tersebut diletakkan pada posisi tanda kelas di sumbu X dan ketinggiannya sesuai dengan frekuensi kelas pada sumbu Y.

- **Garis Lurus:**

Titik-titik pada poligon frekuensi dihubungkan oleh garis lurus. Garis ini memvisualisasikan hubungan antara frekuensi dan tanda kelas dengan cara yang lebih mulus dibandingkan dengan histogram.

- **Penutupan Garis:**

Untuk membentuk grafik yang lengkap, poligon frekuensi biasanya ditutup pada sumbu X di luar interval kelas pertama dan terakhir. Penutupan ini dilakukan untuk memberikan kesan distribusi data yang utuh dan menyeluruh.

Contoh 1.1

Sebuah sampel sebanyak 20 lobster muda dipilih secara acak dari sebuah tangki yang berisi ratusan lobster. Setiap lobster diukur panjangnya (dalam cm), dan hasilnya adalah:

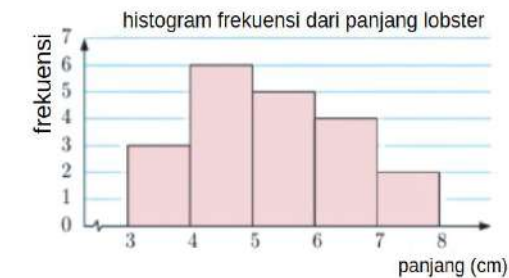
4,9, 5,6, 7,2, 6,7, 3,1, 4,6, 6,0, 5,0, 3,7, 7,3,
6,0, 5,4, 4,2, 6,6, 4,7, 5,8, 4,4, 3,6, 4,2, 5,4

- Susun data menggunakan tabel frekuensi, dan gambarkan data tersebut.
- Nyatakan kelas modus dari data tersebut.
- Jelaskan distribusi dari data tersebut.

Jawab:

- Variabel panjang lobster adalah **kontinu** meskipun panjangnya telah dibulatkan ke milimeter terdekat. Panjang terkecil adalah 3,1 cm dan yang terpanjang adalah 7,3 cm, jadi kita akan menggunakan interval kelas sebesar 1 cm.

panjang l (cm)	frekuensi
$3 \leq l < 4$	3
$4 \leq l < 5$	6
$5 \leq l < 6$	5
$6 \leq l < 7$	4
$7 \leq l < 8$	2



- Kelas modus adalah $4 \leq l < 5$ karena ini yang paling sering muncul.
- Data ini condong ke kanan (*positively skewed*).

LATIHAN 1

Pengantar Statistik

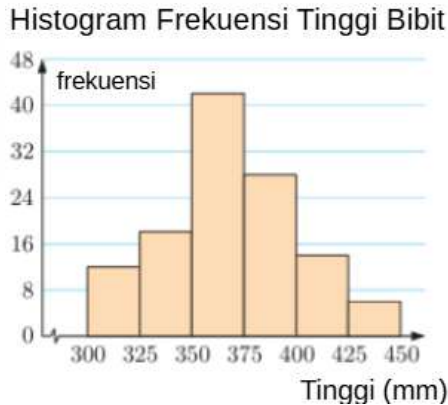
1. Identifikasi apakah nilai berikut termasuk parameter atau statistik:
 - a. DPR RI saat ini terdiri dari 150 perempuan dan 350 pria.
 - b. Rata-rata waktu belajar mandiri seminggu yang diperoleh dari sampel mahasiswa adalah 15,2 jam.
 - c. Dari total penumpang Titanic sebanyak 2223 orang, ditemukan 706 orang selamat saat kapal tenggelam.
2. Tentukan apakah nilai berikut termasuk data kontinu atau diskrit:
 - a. Gaji yang diterima pekerja Indonesia di luar negeri sebesar Rp 3.000.000 per bulan.
 - b. Dari survei terhadap 1560 pria, ditemukan bahwa 38% adalah perokok aktif.
 - c. Rata-rata berat mobil dari suatu sampel adalah 1500 kg.
3. Tentukan level pengukuran yang paling sesuai (nominal, ordinal, interval, rasio) untuk setiap kasus berikut:
 - a. Tinggi badan pemain sepak bola.
 - b. Suhu di dalam kelas saat ini.
 - c. Penilaian acara TV: "Fantastik, baik, cukup, kurang, tidak diterima."
 - d. Nomor punggung pemain basket.
 - e. Nomor telepon dalam buku telepon.
 - f. Penilaian konsumen terhadap produk: "*Best buy, recommended, not recommended.*"
4. Identifikasi mana yang termasuk sampel dan mana yang populasi. Tentukan juga apakah sampel tersebut mewakili populasinya:

- a. Seorang reporter *Newsweek* bertanya kepada 10 orang dewasa di jalan apakah mereka merasa presiden saat ini menjalankan tugasnya dengan baik.
 - b. *Nielsen Media Research* mensurvei 5000 rumah tangga secara acak dan menemukan bahwa 19% TV yang sedang menyala menonton acara *60 Minutes*.
 - c. Dalam survei Gallup terhadap 1059 orang dewasa secara acak, 39% menjawab "ya" untuk pertanyaan "Apakah Anda memiliki senjata di rumah Anda?"
 - d. Seorang mahasiswa pascasarjana di Universitas Newport mengirimkan survei kepada 500 orang dewasa yang dikenalnya, dengan pertanyaan: "Apakah Anda lebih suka menggunakan e-mail atau surat biasa?" Hanya 65 orang yang mengembalikan jawaban.
5. Tabel frekuensi tinggi badan dari sebuah tim basket disajikan di bawah.

Tinggi (cm)	Frekuensi
$170 \leq T < 175$	1
$175 \leq T < 180$	8
$180 \leq T < 185$	9
$185 \leq T < 190$	11
$190 \leq T < 195$	9
$195 \leq T < 200$	3
$200 \leq T < 205$	3

- a. Jelaskan mengapa 'tinggi badan' merupakan variabel kontinu.
- b. Buat histogram frekuensi untuk data tersebut. Sumbu-sumbu harus diberi tanda dan label dengan jelas, dan grafik harus memiliki judul.
- c. Apa kelas modusnya? Jelaskan apa artinya.
- d. Jelaskan distribusi dari data tersebut.

6. Seorang petugas inspeksi tanaman mengambil sampel acak dari bibit berumur enam bulan di sebuah pembibitan dan mengukur tinggi mereka hingga milimeter terdekat. Hasil pengukuran ditampilkan dalam bentuk histogram frekuensi.



- a. Berapa banyak bibit yang memiliki tinggi 400 mm atau lebih?
- b. Berapa persentase bibit yang memiliki tinggi antara 349 mm dan 400 mm?
- c. Jumlah seluruh bibit di pembibitan adalah 1.462 bibit. Perkirakan jumlah bibit yang memiliki tinggi:
 - i. kurang dari 400 mm
 - ii. antara 374 mm dan 425 mm

BAB 2
Statistika Deskriptif

Pendahuluan

Kemampuan yang Diharapkan	Akhir	Mahasiswa memahami ukuran pemusatan data, ukuran penyebaran data serta menginterpretasikan hasil analisis statistika deskriptif untuk mendukung pengambilan keputusan.
Bahan Kajian		- Ukuran pemusatan data (mean, median, modus) - Ukuran penyebaran data (range, varians, standar deviasi).
Metode Pembelajaran		Ceramah interaktif, studi kasus
Pengalaman Belajar Mahasiswa		Mahasiswa mampu memahami dan menjelaskan ukuran pemusatan data, ukuran penyebaran data, ukuran bentuk distribusi, serta menginterpretasikan hasil analisis statistika deskriptif untuk mendukung pengambilan keputusan.
Kriteria Penilaian (Indikator)		- Mampu memahami ukuran pemusatan data & ukuran penyebaran data - Mampu mengidentifikasi ukuran bentuk distribusi. - Menginterpretasikan hasil analisis statistika deskriptif untuk mendukung pengambilan keputusan.

Tujuan Pembelajaran

Setelah mengikuti pembelajaran, mahasiswa diharapkan mampu:

1. Menjelaskan konsep dasar ukuran pemusatan data (mean, median, modus) dan ukuran penyebaran data (range, varians, standar deviasi).
2. Menginterpretasikan hasil analisis statistika deskriptif untuk mendukung pengambilan keputusan.

A. Ukuran Pemusatan Data

Dalam memahami suatu kumpulan data adalah dengan menentukan nilai tengah atau pusat dari data tersebut, sekaligus memperoleh gambaran mengenai penyebarannya. Jika mengetahui salah satunya saja, tanpa lainnya, sering kali kurang memberikan informasi yang utuh.

Untuk menggambarkan pemusatan data, terdapat tiga ukuran statistik yang umum digunakan, yaitu modus, median, dan rata-rata.

- Modus adalah nilai yang paling sering muncul dalam kumpulan data.
- Rata-rata (mean) adalah jumlah seluruh data dibagi dengan banyaknya data.
- Median adalah nilai tengah dari data yang telah diurutkan.

Ketiga ukuran ini saling melengkapi dalam memberikan gambaran mengenai “pusat” data. Pemilihan ukuran mana yang lebih tepat digunakan bergantung pada sifat data serta tujuan analisis. Misalnya, rata-rata sering digunakan untuk data yang bersifat numerik dan terdistribusi normal, median lebih baik digunakan ketika terdapat pencilan (outlier), sedangkan modus bermanfaat untuk data kategori.

B. Modus

Pada data numerik diskrit, modus adalah nilai yang paling sering muncul dalam sekumpulan data. Sedangkan pada data numerik kontinu, kita tidak dapat menyebutkan modus dengan cara yang sama, karena tidak ada dua nilai data yang persis sama. Sebagai gantinya, kita menggunakan istilah

kelas modus, yaitu kelompok data yang memiliki frekuensi paling tinggi (paling sering muncul).

C. Rata-Rata (Mean)

Rata-rata (Mean) dari suatu kumpulan data adalah istilah statistik untuk rata-rata hitung.

$$\text{Mean} = \frac{\text{Jumlah seluruh nilai data}}{\text{Banyak data}}$$

Misalkan x adalah variabel numerik. Kita definisikan:

- x_i adalah nilai data ke- i
- n adalah jumlah nilai data dalam sampel atau populasi
- $\bar{x}$ menyatakan **rata-rata dari sebuah sampel**, sehingga:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

- μ menyatakan **rata-rata dari populasi**, sehingga: $\mu = \frac{\sum_{i=1}^n x_i}{n}$

D. Median

Median adalah nilai tengah dari kumpulan data yang telah diurutkan. Kumpulan data yang terurut diperoleh dengan menyusun data dari nilai terkecil hingga terbesar. Median membagi data menjadi dua bagian yang sama besar. Setengah dari data memiliki nilai kurang dari atau sama dengan median, dan setengah lainnya memiliki nilai lebih dari atau sama dengan median.

- Untuk jumlah data yang **ganjil**, median adalah salah satu nilai asli dari kumpulan data tersebut.
- Untuk jumlah data yang **genap**, median adalah rata-rata dari dua nilai tengah, sehingga median mungkin bukan salah satu dari nilai yang ada di data awal.

Jika ada n nilai data yang telah diurutkan dari yang terkecil ke terbesar, maka median adalah nilai data ke - $\left(\frac{n+1}{2}\right)$. Sebagai contoh:

Jika $n = 13 \rightarrow \frac{13+1}{2} = 7$, maka median adalah nilai urutan ke-7 dalam data.

Jika $n = 14 \rightarrow \frac{14+1}{2} = 7,5$, maka median adalah rata-rata dari nilai ke-7 dan ke-8 dalam urutan data.

Contoh 2.1

Data tidak dikelompokkan. Jumlah produk cacat yang dikembalikan ke toko peralatan elektronik selama periode 21 hari adalah:

3 4 4 9 8 8 6 4 7 9 1 3 5 3 5 9 8 6 3 7 1

Untuk kumpulan data ini, tentukan rata-rata (mean), median, dan modus (mode).

Jawab:

a. $mean = \frac{3 + 4 + 4 + \dots + 3 + 7 + 1}{21}$ ← jumlah dari nilai data
← banyaknya data
 $= \frac{113}{21} \approx 5,38$ produk cacat

b. Karena $n = 21, \frac{n+1}{2} = 11$

Urutan data yang telah disusun dari kecil ke besar adalah:

1 1 3 3 3 3 4 4 4 4 5 5 6 6 7 7 8 8 8 9 9 9
↑
 Nilai ke-11

median = 5 produk cacat

c. Angka 3 muncul paling banyak dibanding nilai lainnya, sehingga dapat disimpulkan bahwa modusnya adalah 3 produk cacat.

Contoh 2.2

Tabel berikut menunjukkan jumlah lemparan bebas berhasil yang dicatat oleh pemain basket dalam kuartir pertama sebuah pertandingan.

Jumlah Lemparan	1	2	3	4	5	6
Frekuensi	4	11	18	13	7	2

Dengan menggunakan data di atas, tentukan:

- Rata-rata (mean)
- Median
- Modus

Jawab:

Untuk data tersebut

Jumlah lemparan (x)	Frekuensi (f)	Perkalian ($f \cdot x$)	Frekuensi Kumulatif
1	4	4	4
2	11	22	15
3	18	54	33
4	13	52	46
5	7	35	53
6	2	12	55
Total	$\sum f = 55$	$\sum fx = 179$	

a. $\bar{x} = \frac{\sum fx}{\sum f} = \frac{179}{55} \approx 3,25$ lemparan

Dalam hal ini $\frac{\sum fx}{\sum f}$ merupakan cara singkat dari $\frac{\sum_{i=1}^6 f_i x_i}{\sum_{i=1}^6 f_i}$

- b. Jumlah data ada 55, jadi $n = 55 \cdot \frac{n+1}{2} = 28$, Artinya, nilai median adalah data ke-28. Dari kolom frekuensi kumulatif, kita lihat bahwa nilai data ke-16 hingga ke-33 berada pada kategori 3 lemparan.
- Data ke-28 adalah 3 ace
 - Median = 3 ace
- c. Dari kolom frekuensi, nilai frekuensi tertinggi adalah 18, yang terjadi pada jumlah 3 lemparan. Jadi modus sama dengan 3.

Contoh 2.3

Perkiraan rata-rata dari data usia pengemudi bus berikut, hingga tahun terdekat. Tentukan nilai rata-rata dari usia pengemudi bus tersebut.

Umur (thn)	21 - 25	26 - 30	31 - 35	36 - 40	41 - 45	46 - 50	51 - 55
Frekuensi	11	14	32	27	29	17	7

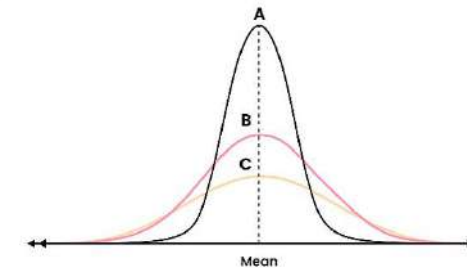
Jawab:

Umur (thn)	Frekuensi (f)	Titik Tengah (x)	Perkalian (f . x)
21 - 25	11	23	253
26 - 30	14	28	392
31 - 35	32	33	1056
36 - 40	27	38	1026
41 - 45	29	43	1247
46 - 50	17	48	816
51 - 55	7	53	371
Total	$\sum f = 137$		$\sum f.x = 5161$

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{5161}{137} = 37,7 \approx 38$$

∴ rata-rata usia pengemudi adalah sekitar 38 tahun.

E. Ukuran Penyebaran Data



Gambar 2.1. Distribusi Data

Gambar di samping adalah ilustrasi distribusi normal dengan rata-rata (mean) yang sama, tetapi memiliki penyebaran (dispersi/variabilitas) yang berbeda. Garis vertikal (mean) menunjukkan nilai rata-rata (mean) yang sama untuk ketiga distribusi (A, B, dan C) artinya pusat data sama.

Ukuran penyebaran kurva A, B, dan C berbeda yang artinya standar deviasi/varians berbeda:

- Kurva A memiliki varians/standar deviasi paling kecil, data lebih terkonsentrasi di sekitar mean (penyebaran kecil) → data homogen.
- Kurva B dengan penyebaran data lebih lebar dibanding A, tetapi lebih kecil dibanding C.
- Kurva C memiliki varians/standar deviasi paling besar, data lebih menyebar jauh dari mean (penyebaran besar) → data heterogen.

Dalam statistika deskriptif, terdapat beberapa ukuran yang umum digunakan untuk menggambarkan penyebaran data, di antaranya adalah **jangkauan (range)**, **jangkauan interkuartil (IQR)**, **variens**, dan **standar deviasi**.

F. Jangkauan

Jangkauan adalah perbedaan antara nilai data maksimum (terbesar) dan minimum (terkecil)

$$\text{Jangkauan} = \text{Maksimum} - \text{Minimum}$$

G. Kuartil dan Jangkauan Interkuartil

Median membagi kumpulan data yang di urutkan menjadi dua bagian dan kedua bagian ini dibagi lagi menjadi dua bagian oleh **kuartil**. Nilai tengah dari bagian bawah disebut dengan **kuartil bawah** atau **persentil ke 25**. Seperempat atau 25% dari data memiliki nilai kurang dari atau sama dengan kuartil bawah. 75% dari data memiliki lebih besar dari atau sama dengan kuartil bawah.

Nilai tengah dari bagian atas disebut **kuartil atas** atau **persentil ke-75**. Seperempat atau 25% dari data memiliki nilai kurang dari atau sama dengan kuartil atas. 75% dari data memiliki nilai lebih besar dari atau sama dengan kuartil atas.

Jangkauan interkuartil adalah jangkauan dari setengah bagian tengah atau 50% dari data.

$$\text{jangkauan interkuartil (IQR)} = \text{kuartil atas} - \text{kuartil bawah}$$

Kumpulan data dengan demikian dibagi menjadi kuartil oleh kuartil bawah kuartil (Q_1), median (Q_2), dan kuartil atas (Q_3). Sehingga, rumus dari jangkauan interkuartil:

$$\text{IQR} = Q_3 - Q_1.$$

Contoh 2.4

Untuk kumpulan data : 7, 3, 1, 7, 6, 9, 3, 8, 5, 8, 6, 3, 7, 1, 9. Tentukan median, kuartil bawah, kuartil atas, dan jangkauan interkuartil.

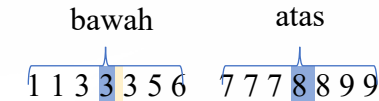
Jawab:

Kumpulan data yang diurutkan adalah: 1, 1, 3, 3, 3, 5, 6, 6, 7, 7, 7, 8, 8, 9, 9

a. karena $n = 15$, $\frac{n+1}{2} = 8$

$\therefore$ median nilai data ke-8 adalah 6

b. Karena median adalah nilai data, kita sekarang akan mengabaikannya dan membagi sisa data menjadi dua:



$$Q_1 (\text{median dari bagian bawah}) = 3 \text{ dan } Q_2 (\text{median dari bagian atas}) = 8$$

c. $\text{IQR} = Q_3 - Q_1 = 8 - 3 = 5$

H. Varians dan Standar Deviasi.

Masalah utama dalam menggunakan rentang dan jangkauan interkuartil (IQR) sebagai ukuran penyebaran adalah keduanya hanya mempertimbangkan dua titik data dalam perhitungannya. Akibatnya, informasi penting mengenai variasi data di antara nilai-nilai lain bisa terabaikan. Untuk memperoleh gambaran yang lebih lengkap tentang seberapa jauh data menyebar dari rata-ratanya, diperlukan ukuran yang mempertimbangkan seluruh data dalam perhitungannya. Di sinilah varians dan standar deviasi (simpangan baku) berperan sebagai ukuran penyebaran yang lebih informatif.

Perhatikan himpunan data dengan n nilai: $x_1, x_2, x_3, x_4, \dots, x_n$, dengan mean $\bar{x}$

$x_i - \bar{x}$ mengukur seberapa jauh x_i menyimpang dari rata-rata, sehingga kita mungkin menduga bahwa rata-rata deviasi $\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})$ akan memberikan ukuran penyebaran yang baik. Namun, nilai ini ternyata selalu nol. Sehingga, varians didefinisikan sebagai:

$$\text{Varians sampel } n \text{ nilai data adalah } s_n^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

Untuk kumpulan data n nilai, $s_n = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$ disebut **standar deviasi**.

Contoh 2.5

Sebuah perpustakaan men survei 20 peminjam setiap hari, dari Senin hingga Jumat, dan mencatat jumlah peminjam yang tidak puas dengan ragam bahan bacaan. Hasilnya adalah: 3 7 6 8 11. Tahun berikutnya, perpustakaan

menerima hibah yang memungkinkan pembelian sejumlah besar buku. Survei kemudian diulang dan hasilnya adalah: 2 3 5 4 6.

Tentukan rata-rata dan deviasi standar untuk setiap survei. Informasi apa yang dapat kita pahami dari statistik ini?

Jawab:

Survei 1

x	$x - \bar{x}$	$(x - \bar{x})^2$
3	- 4	16
7	0	0
6	- 1	1
8	1	1
11	4	16
35	Total	34

Survei 2

x	$x - \bar{x}$	$(x - \bar{x})^2$
2	- 2	4
3	- 1	1
4	1	1
5	0	0
6	2	4
20	Total	10

$$\begin{aligned} \bar{x} &= \frac{35}{5} & S_n &= \sqrt{\frac{\sum(x-\bar{x})^2}{n}} \\ &= 7 & &= \sqrt{\frac{34}{5}} \\ & & &= 2.61 \end{aligned}$$

$$\begin{aligned} \bar{x} &= \frac{20}{5} & S_n &= \sqrt{\frac{\sum(x-\bar{x})^2}{n}} \\ &= 4 & &= \sqrt{\frac{10}{5}} \\ & & &= 1.41 \end{aligned}$$

I. Pengambilan Sampel dari Populasi

Populasi sering kali berukuran sangat besar, sehingga pengumpulan data dari setiap individu tidak memungkinkan karena keterbatasan waktu dan biaya. Oleh karena itu, diambil sampel acak dari populasi dengan harapan dapat mewakili karakteristik populasi secara akurat. Agar representatif, sampel harus cukup besar dan dipilih sedemikian rupa sehingga hasilnya tidak bias. Untuk membedakan antara ukuran yang dihitung dari sampel dan dari populasi, digunakan notasi yang berbeda untuk

rata-rata, varians, dan standar deviasi, sebagaimana ditunjukkan pada tabel berikut.

	Sampel	Populasi
Rata - rata	$\bar{x}$	μ
Varians	s_n^2	σ^2
Simpangan baku	s_n	σ

Secara umum, rata-rata populasi μ dan standar deviasi σ tidak akan diketahui.

Namun, dengan statistik dari suatu sampel, kita dapat membuat kesimpulan tentang populasi menggunakan hasil yang diasumsikan tanpa bukti. Ketika sampel berukuran n digunakan untuk menarik kesimpulan tentang suatu populasi:

- rata-rata $\bar{x}$ sampel adalah perkiraan tak bias estimasi μ
- s_n adalah perkiraan untuk simpangan baku σ .

Contoh 2.6

Sampel acak sebanyak 48 domba diambil dari kawanan yang jumlahnya lebih dari 2000 domba. Rata-rata sampel berat badan mereka adalah 48,6 kg dengan varians 17,5 kg^2 .

- Hitung simpangan baku sampel. Kemudian, perkirakan simpangan baku populasi tempat sampel tersebut diambil
- Temukan perkiraan yang tidak bias untuk rata-rata berat domba di kawanan

Jawab:

- $s_n = \sqrt{variance} = \sqrt{17.5} = 4.18$ kg
 σ diperkirakan oleh s_n , jadi kami memperkirakan simpangan baku sebesar 4,18 kg
- μ diperkirakan oleh $\bar{x}$ sebesar 48,6 kg.

J. Standar Deviasi Untuk Data Kelompok.

Misalkan ada n skor dalam satu set data berkelompok dengan rata-rata $\bar{x}$.

Jika nilai data yang berbeda adalah $x_1, x_2, \dots, x_k$, dengan frekuensi yang sesuai $f_1, f_2, \dots, f_k$, maka simpangan baku diberikan oleh semua

$$s_n = \sqrt{\frac{\sum f_i(x_i - \bar{x})^2}{n}}$$

Contoh 2.7

Tentukan standar deviasi dari distribusi menggunakan data berikut:

Skor	1	2	3	4	5
Frekuensi	1	2	4	2	1

Jawab:

x	f	$f \cdot x$	$x - \bar{x}$	$(x - \bar{x})^2$	$f(x - \bar{x})^2$
1	1	1	-2	4	4
2	2	4	-1	1	2
3	4	12	1	1	0
4	2	8	0	0	2
5	1	5	2	4	4
Total	10	30			12

$$\bar{x} = \frac{\sum fx}{n} = \frac{30}{10} = 3$$

$$s_n = \sqrt{\frac{\sum f(x - \bar{x})^2}{n}} = \sqrt{\frac{12}{10}} = 1,10$$

K. Pencilan (Outliers)

Dalam analisis data, tidak jarang kita menemukan nilai-nilai yang letaknya jauh dari mayoritas data lainnya. Nilai-nilai yang menyimpang secara ekstrem ini disebut **pencilan** (*outliers*). Outliers dapat muncul karena

berbagai alasan, seperti kesalahan pengukuran, kesalahan pencatatan, atau memang merupakan karakteristik unik dari data. Kehadiran outliers penting diperhatikan, karena dapat memengaruhi nilai rata-rata, varians, maupun hasil analisis statistik lainnya. Oleh karena itu, sebelum melakukan analisis lebih lanjut, peneliti perlu mengidentifikasi apakah suatu nilai ekstrim merupakan kesalahan data yang harus diperbaiki/dihapus, atau justru mengandung informasi penting yang tidak boleh diabaikan.

• Batas Atas = Kuartil Atas + $1.5 \times \text{IQR}$

Setiap data yang lebih besar dari batas atas adalah pencilan.

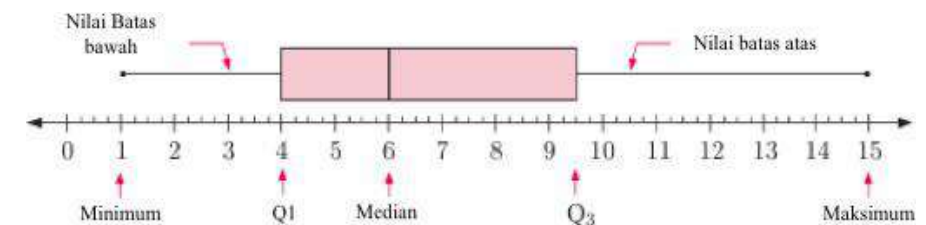
• Batas Bawah = Kuartil Bawah - $1.5 \times \text{IQR}$

Setiap data yang lebih kecil dari batas bawah adalah pencilan.

L. Boxplot

Selain tabel, diagram batang, maupun histogram, terdapat cara lain yang sangat efektif untuk menggambarkan penyebaran data, yaitu **boxplot** (atau diagram kotak). Boxplot memberikan ringkasan lima angka utama dari suatu data: nilai minimum, kuartil pertama (Q_1), median (Q_2), kuartil ketiga (Q_3), dan nilai maksimum.

- Nilai Minimum
 - Kuartil Bawah (Q_1)
 - Median (Q_2)
 - Kuartil Atas (Q_3)
 - Nilai Maksimum
- Ringkasan lima angka utama pada data



Gambar 2.2. Boxplot (*Box and Whisker Plot*)

Perhatikan boxplot pada Gambar di atas. **Kelima angka tersebut** membentuk 5 ringkasan dari kumpulan data dimana nilai minimum 1; nilai kuartil bawah (Q_1) 4; median 6; nilai kuartil atas (Q_3) 9,5; dan nilai maksimum 15.

Dengan boxplot, kita dapat dengan mudah melihat sebaran data, posisi nilai tengah, serta derajat simetri atau kecondongan distribusi. Boxplot juga sangat berguna untuk mengidentifikasi adanya pencilan (*outlier*) yang mungkin tidak terlihat jelas pada grafik lainnya. Karena kepraktisannya, boxplot sering digunakan untuk membandingkan beberapa kelompok data secara berdampingan, sehingga memudahkan analisis dalam melihat perbedaan variasi dan kecenderungan antarkelompok.

Contoh 2.8

Uji data berikut apakah terdapat pencilan dan kemudian buatlah boxplot untuk data tersebut:

3, 7, 8, 8, 5, 9, 10, 12, 14, 7, 1, 3, 8, 16, 8, 6, 9, 10, 13, 7

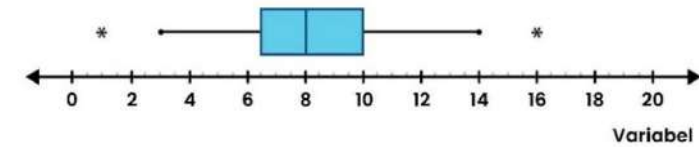
Jawab:

Kumpulan data yang diurutkan adalah:

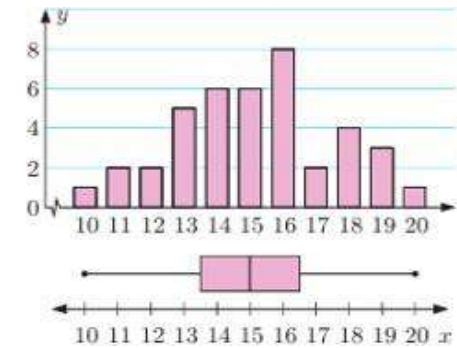
1	3	3	5	6	7	7	8	8	8	9	10	10	12	13	14	16	{ n = 20 }
↓				↓			↓				↓					↓	
Min = 1				Q1 = 6.5			Median = 8				Q3 = 10					Maks = 16	

Uji Pencilan:	Batas Atas	Batas Bawah
	= Kuartil Atas + $1.5 \times \text{IQR}$	= Kuartil Bawah - $1.5 \times \text{IQR}$
	= $10 + 1.5 \times 3.5$	= $6.5 - 1.5 \times 3.5$
	= 15.25	= 1.25

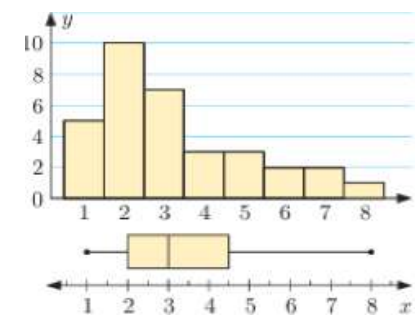
16 berada di atas batas atas, jadi 16 adalah pencilan. Begitu juga, angka 1 berada di bawah pada batas bawah, jadi 1 adalah pencilan. Gambar boxplot sebagai berikut:



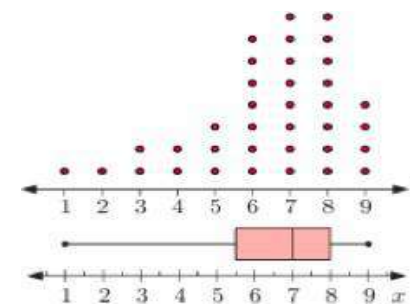
INTERPRETASI BOXPLOT



Gambar 2.3. Boxplot Simetris



Gambar 2.4. Boxplot Miring ke Kanan
(Positively Skewed)



Gambar 2.4. Boxplot Miring ke Kiri
(Negatively Skewed)

Kumpulan data dengan **distribusi simetris** akan menghasilkan diagram kotak (boxplot) yang juga tampak simetris. Pada kondisi ini, garis *whisker* di kedua sisi memiliki panjang yang hampir sama, dan garis median terletak di tengah kotak, menunjukkan bahwa data tersebar secara seimbang di sekitar nilai tengahnya.

Jika **ekor kanan (whisker kanan)** lebih panjang daripada ekor kiri dan median berada lebih dekat ke sisi kiri kotak, maka distribusi dikatakan **miring ke kanan**. Artinya, sebagian besar nilai berada di sisi bawah, namun terdapat beberapa nilai yang sangat tinggi (*outlier*) yang menarik rata-rata ke arah kanan.

Jika **ekor kiri (whisker kiri)** lebih panjang daripada ekor kanan dan median berada lebih dekat ke sisi kanan kotak, maka distribusi dikatakan **miring ke kiri**. Dalam hal ini, sebagian besar nilai cenderung tinggi, namun terdapat beberapa nilai rendah yang menyebabkan sebaran memanjang ke arah kiri.

M. Grafik Frekuensi Kumulatif

Selain mencari median, sering kali kita juga ingin mengetahui berapa banyak atau berapa proporsi data yang berada di atas atau di bawah suatu nilai tertentu. Untuk keperluan ini, kita dapat menyusun *tabel distribusi frekuensi kumulatif* dan merepresentasikannya dalam bentuk grafik yang disebut *grafik frekuensi kumulatif*.

Salah satu bentuk grafik frekuensi kumulatif adalah **ogive**. Ogive digunakan dalam statistika untuk menampilkan data berdasarkan distribusi frekuensi kumulatif, dengan menunjukkan jumlah atau proporsi kumulatif data hingga batas tertentu pada setiap interval kelas. Grafik ini memudahkan kita dalam membaca informasi mengenai sebaran data serta dalam memperkirakan ukuran pemusatan maupun penyebaran.

Contoh 2.9

Data berikut menunjukkan catatan waktu maraton putri pada Olimpiade untuk seluruh peserta yang berhasil menyelesaikan lomba.

Waktu selesai (t)	Frekuensi
2 jam 26 menit $\leq t < 2$ jam 28 menit	8
2 jam 28 menit $\leq t < 2$ jam 30 menit	3
2 jam 30 menit $\leq t < 2$ jam 32 menit	9
2 jam 32 menit $\leq t < 2$ jam 34 menit	11
2 jam 34 menit $\leq t < 2$ jam 36 menit	12
2 jam 36 menit $\leq t < 2$ jam 38 menit	7
2 jam 38 menit $\leq t < 2$ jam 40 menit	5
2 jam 40 menit $\leq t < 2$ jam 48 menit	8
2 jam 48 menit $\leq t < 2$ jam 56 menit	6

- Buatlah tabel distribusi frekuensi kumulatif.
- Gambarkan grafik frekuensi kumulatif.
- Gunakan grafik frekuensi kumulatif untuk memperkirakan:
 - nilai median pada waktu penyelesaian

- jumlah peserta yang menyelesaikan dalam waktu kurang dari 2 jam 35 menit
- persentase peserta yang membutuhkan waktu lebih dari 2 jam 39 menit untuk menyelesaikannya
- waktu yang dicapai oleh seorang peserta yang finis di 20% teratas dari pelari yang menyelesaikan maraton

Jawab:

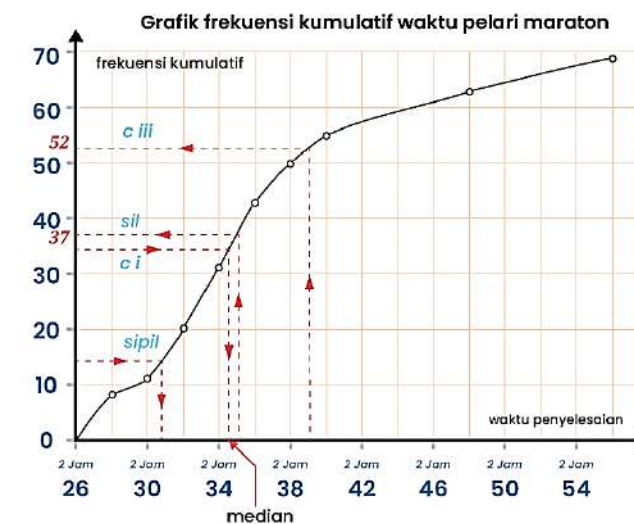
- Tabel distribusi frekuensi kumulatif

Waktu Selesai (t)	Frekuensi	Frekuensi Kumulatif
2 jam 26 menit $\leq t < 2$ jam 28 menit	8	8
2 jam 28 menit $\leq t < 2$ jam 30 menit	3	11
2 jam 30 menit $\leq t < 2$ jam 32 menit	9	20
2 jam 32 menit $\leq t < 2$ jam 34 menit	11	31
2 jam 34 menit $\leq t < 2$ jam 36 menit	12	43
2 jam 36 menit $\leq t < 2$ jam 38 menit	7	50
2 jam 38 menit $\leq t < 2$ jam 40 menit	5	55
2 jam 40 menit $\leq t < 2$ jam 48 menit	8	63
2 jam 48 menit $\leq t < 2$ jam 56 menit	6	69

8+3=11 peserta menyelesaikan maraton dalam waktu kurang dari 2 jam 30 menit.

50 peserta menyelesaikan maraton dalam waktu kurang dari 2 jam 38 menit.

- Grafik frekuensi kumulatif.



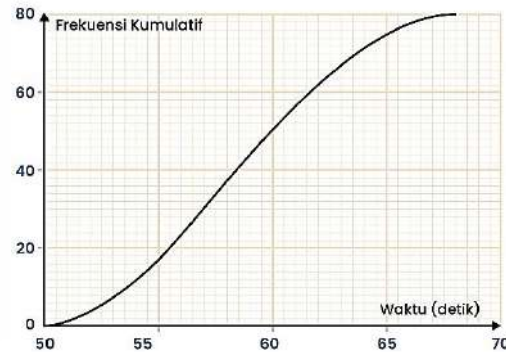
- c. Dengan menggunakan grafik frekuensi kumulatif pada no b
- Median diperkirakan menggunakan persentil ke-50. Karena 50% dari 69 adalah 34,5; kita mulai dengan frekuensi kumulatif 34,5, dan mencari waktu yang sesuai. Mediannya sekitar 2 jam 34,5 menit.
 - Ada sekitar 37 peserta yang membutuhkan waktu kurang dari 2 jam 35 menit untuk menyelesaikan lomba.
 - Ada $69 - 52 = 17$ peserta yang membutuhkan waktu lebih dari 2 jam 39 menit. Jadi, $\frac{17}{69} \approx 26,4\%$ membutuhkan waktu lebih dari 2 jam 39 menit.
 - Waktu yang dibutuhkan diperkirakan menggunakan persentil ke-20. Karena 20% dari 69 adalah 13,8; kami menemukan waktu yang sesuai dengan frekuensi kumulatif sekitar 14. 20% pesaing teratas membutuhkan waktu kurang dari 2 jam 31 menit.

LATIHAN 2

Statistika Deskriptif

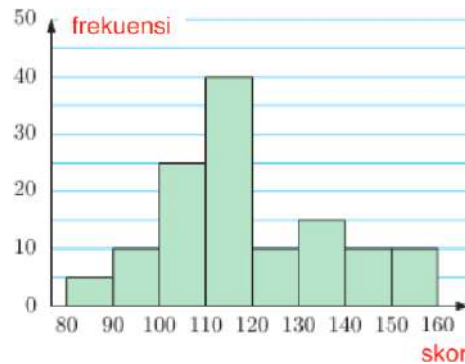
- Seorang teknisi produksi mencatat waktu (dalam menit) yang dibutuhkan oleh mesin CNC untuk menyelesaikan proses pemotongan pada sejumlah komponen logam.
0,4 2,1 3,4 3,9 4,7 3,7 0,8 3,6 4,1 4,9 2,5 3,1 1,5 2,6 4,0
1,3 3,5 0,9 1,5 4,2 3,5 2,1 3,0 1,7 3,6 2,8 3,7 2,8 3,2 3,3
 - Tentukan median dan range dari data
 - Kelompokkan data menjadi 5 grup dan tampilkan menggunakan histogram frekuensi
 - Berikan komentar mengenai kemiringan (*skewness*) data.
- Himpunan data 4, 6, 9, a , 3, b memiliki rata-rata adalah a dan nilai modulus 6. Tentukan nilai dari a dan b jika $a > b$.
- Diketahui data yaitu 11, 12, 12, 13, 14, 14, 15, 15, 15, 16, 17, 17, 18. Gambarkan boxplot dari data tersebut dan tentukan apakah terdapat pencilan (*outlier*)?
- Diketahui data sebagai berikut:
 - Tentukan nilai x jika mean pada dataset tersebut adalah 9,2.
 - Tentukan standard deviasi dan varians (gunakan *R Studio*)
 - Tentukan nilai Q_1 , Q_2 , dan Q_3 (gunakan *R Studio*)
 - Buatlah box-plot dengan menggunakan data tersebut (gunakan *R Studio*)
 - Apakah terdapat pencilan (*outlier*)?
- Sebanyak 80 mahasiswa mengerjakan ujian statistik dan tercatat lama waktu mengerjakan ujian tersebut seperti ditampilkan pada gambar. Tentukan:

- a. Median
- b. Jangkauan interkuartil (interquartile range)
- c. Waktu yang sesuai dengan 10% peserta ujian tercepat



9. Histogram frekuensi ini mengilustrasikan hasil tes bakat yang diberikan kepada sekelompok orang yang mencari posisi di sebuah perusahaan.

- a. Berapa banyak orang yang mengikuti tes?
- b. Tentukan skor rata-rata untuk tes tersebut.
- c. Berapa persentase orang yang mendapat skor kurang dari 100 untuk tes tersebut?
- d. Berapa persentase orang yang mendapat skor lebih dari 130 untuk tes tersebut?



10. Jumlah tusuk gigi dalam sampel acak dari 48 kotak dihitung dan hasilnya ditabulasikan

Jumlah tusuk gigi	33	35	36	37	38	39	40
Frekuensi	1	5	7	13	12	8	2

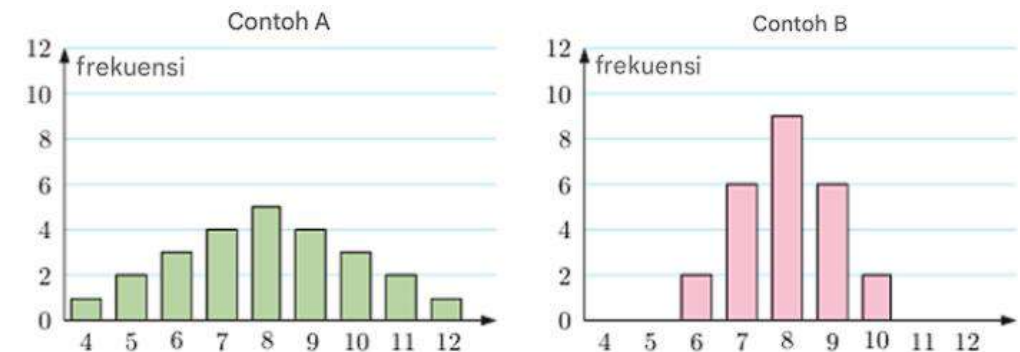
- a. Tentukan rata-rata dan standar deviasi dari tusuk gigi dalam kotak.
- b. Tentukan rata-rata dan standar deviasi dari populasi sampel diambil.

11. Panjang dari suatu benda yang berjumlah 30 buah dipilih secara acak dan diperoleh data:

Panjang	Frekuensi
$40\text{ cm} \leq P < 42\text{ cm}$	1
$42\text{ cm} \leq P < 44\text{ cm}$	1
$44\text{ cm} \leq P < 46\text{ cm}$	3
$46\text{ cm} \leq P < 48\text{ cm}$	7
$48\text{ cm} \leq P < 50\text{ cm}$	11
$50\text{ cm} \leq P < 52\text{ cm}$	5
$52\text{ cm} \leq P < 54\text{ cm}$	2

- a. Perkirakan panjang rata-rata dan standar deviasi dari benda tersebut
- b. Perkirakan panjang rata-rata dan standar deviasi populasi dari mana sampel tersebut diambil

12. Gambar berikut ini menunjukkan grafik kolom dengan masing-masing distribusinya:



- a. Dengan melihat grafik, distribusi mana yang memiliki sebaran lebih luas?
- b. Temukan rata-rata dari setiap sampel
- c. Temukan simpangan baku untuk setiap sampel. Berikan komentar pada jawaban anda

13. Jumlah poin yang dicetak oleh Arshaka dan Sena dalam 8 pertandingan basket terakhir ditabulasikan di bawah.

Poin Arshaka	23	17	31	25	25	19	28	32
Poin Sena	9	29	41	26	14	44	38	43

- Temukan rata-rata dan simpangan baku dari jumlah poin yang dicetak oleh setiap pemain.
 - Manakah dari kedua pemain tersebut yang lebih konsisten?
14. Arshaka pergi mengamati burung selama 25 hari. Jumlah burung yang dia lihat setiap hari adalah:
- 12, 5, 13, 16, 8, 10, 12, 18, 9, 11, 14, 14, 22, 9, 10, 7, 9, 11, 13, 7, 10, 6, 13, 3, 8.
- Tentukan median, kuartil bawah, dan kuartil atas dari kumpulan data.
 - Tentukan jangkauan interkuartil dari kumpulan data.
 - Tentukan batas bawah dan atas untuk pencilan.
 - Apakah ada pencilan?
 - Gambarlah boxplot dari kumpulan data.

BAB 3

Probabilitas

Pendahuluan

Kemampuan Akhir yang Diharapkan	Mahasiswa memahami teori dasar probabilitas, peluang dalam kejadian majemuk, himpunan dan diagram venn, serta kaidah bayes.
Bahan Kajian	- Teori probabilitas - Kejadian majemuk - Himpunan dan diagram venn dalam peluang - Kaidah Bayes
Metode Pembelajaran	Ceramah interaktif, studi kasus
Pengalaman Belajar Mahasiswa	Mahasiswa menganalisis konsep dasar probabilitas melalui studi kasus sederhana, melakukan simulasi kejadian majemuk, dan menerapkan kaidah Bayes pada permasalahan nyata
Kriteria Penilaian (Indikator)	- Menjelaskan definisi peluang klasik, empiris, dan subjektif dengan tepat. - Menyelesaikan soal yang melibatkan kejadian saling lepas dan tidak saling lepas secara benar. - Menyusun model probabilistik berdasarkan data, dan menghitung peluang bersyarat dengan menggunakan kaidah Bayes secara tepat.

Tujuan Pembelajaran

1. Mahasiswa mampu menyelesaikan masalah berkaitan dengan aturan peluang
2. Mahasiswa mampu menyelesaikan masalah berkaitan dengan peluang kejadian majemuk
3. Mahasiswa mampu menyelesaikan masalah berkaitan dengan Kaidah Bayes

A. Percobaan Statistika dan Kejadian

Percobaan statistika merupakan suatu proses pengamatan atau tindakan yang dilakukan dalam kondisi yang terkontrol atau terencana, dengan tujuan untuk mengumpulkan data tentang suatu fenomena acak. Sebagai contoh, melempar dadu dengan hasil yang mungkin adalah angka 1 sampai 6, menguji mesin dengan hasil yang mungkin adalah mesin gagal atau berfungsi, dan lain lain.

Himpunan yang memuat semua kemungkinan dari tindakan statistika disebut sebagai ruang sampel (*sample space*) yang dinotasikan dengan S . Contoh dari ruang sampel:

- Percobaan melempar satu dadu berisi enam

Ruang sampel: $S = \{1, 2, 3, 4, 5, 6\}$.

- Percobaan melempar satu koin

Ruang sampel: $S = \{\text{Angka, Gambar}\}$ atau bisa ditulis $S = \{A, G\}$

- Percobaan melempar dua koin

Ruang sampel: $S = \{(A, A), (A, G), (G, A), (G, G)\}$

B. Teori Probabilitas

Jika sebuah ruang contoh terdiri atas n kemungkinan kejadian yang peluang masing-masing sama besar (*equally likely*), maka peluang terjadinya kejadian A adalah:

$$P(A) = \frac{\text{Banyaknya anggota kejadian } A}{\text{Banyaknya anggota ruang sampel } S} = \frac{n(A)}{n(S)}$$

- Nilai probabilitas dinyatakan dalam rentang 0 hingga 1, atau 0% hingga 100%
- Peluang 0 artinya peristiwa tidak mungkin terjadi
- Peluang 1 artinya peristiwa pasti terjadi

Kejadian komplemen (dilambangkan dengan A^c atau A') adalah kejadian yang mencakup semua hasil yang **tidak termasuk dalam A**, tetapi masih berada di dalam ruang sampel S atau peluang terjadinya suatu kejadian yang berlawanan dengan suatu kejadian. Kejadian komplemen dirumuskan sebagai:

$$P(A) + P(A^c) = 1$$

Misalnya, peluang esok hari hujan adalah 85%, maka peluang esok hari tidak hujan adalah $100\% - 85\% = 15\%$.

Contoh 3.1

Sebuah tiket dipih secara random dari suatu box yang berisi 3 hijau, 4 kuning, dan 5 biru. Tentukan probabilitas mendapatkan:

- | | |
|--|---|
| a. Tiket berwarna hijau | d. Tiket warna jingga |
| b. Tiket warna hijau atau kuning | e. Tiket warna hijau, kuning, atau biru |
| c. Tiket warna bukan hijau atau kuning | |

Jawab:

- | | | |
|--------------------------|--|--|
| a. $P(H) = \frac{3}{12}$ | b. $P(H \text{ atau } K) = \frac{3+4}{12}$ | c. $P(H \text{ atau } K)^c = 1 - \frac{7}{12}$ |
| | $= \frac{7}{12}$ | $= \frac{5}{12}$ |
| d. $P(J) = \frac{0}{12}$ | e. $P(H \text{ atau } K \text{ atau } B)$ | |
| $= 0$ | $= \frac{3+4+5}{12} = 1$ | |

Artinya mendapatkan warna jingga tidak mungkin terjadi
Artinya mendapatkan warna hijau atau kuning atau biru pasti terjadi

Contoh 3.2

Pada suatu survei sebanyak 55 pria dan 28 wanita diminta untuk memilih jenis *smartphone*. Hasil survei ditunjukkan seperti tabel di bawah ini:

	Wanita	Pria
<i>Iphone</i>	55	28
<i>Samsung</i>	17	30

Gunakan tabel tersebut untuk menghitung probabilitas bahwa seseorang dipilih secara acak:

- a. Memilih *smartphone* tipe Iphone
- b. Wanita dan menyukai Samsung
- c. Pria atau *smartphone* tipe Samsung
- d. Menyukai *iphone*, dia berasal (given) dari wanita
- e. Dia seorang pria, berasal (given) dari menyukai tipe *samsung*

Jawab:

Dari tabel di atas kita bisa totalkan

	Wanita	Pria	Total
<i>Iphone</i>	55	28	83
<i>Samsung</i>	17	30	47
Total	72	58	130

- a. Memilih *smartphone* tipe Iphone
 $= \frac{83}{130} \approx 0,638$
- b. Wanita dan menyukai Samsung
 $= \frac{17}{130} \approx 0,131$
- c. Pria atau *smartphone* tipe *samsung*: $28 + 30 + 17 = 75$
 $= \frac{75}{130} \approx 0,577$
- d. Dari 72 wanita, 55 menyukai *iphone*
 $= \frac{55}{72} \approx 0,764$
- e. Dari 47 orang menyukai *samsung*, sebanyak 30 adalah pria
 $= \frac{30}{47} \approx 0,638$

C. Diagram Pohon

Diagram pohon adalah representasi visual yang digunakan untuk menunjukkan semua kemungkinan hasil dari suatu percobaan bertahap dengan peluang setiap kejadian. Diagram ini sangat membantu untuk menghitung peluang gabungan dari dua atau lebih kejadian. Untuk peluang gabungan, tinggal mengalikan peluang dari cabang yang dilewati. Misalnya:

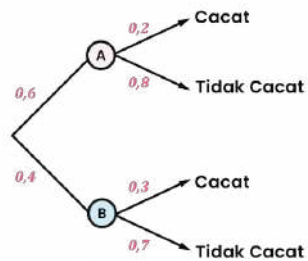
Contoh 3.3

Mesin produksi A dan B menghasilkan komponen dengan probabilitas mesin A sebesar 60%. Diketahui 20% probabilitas mesin A dan 30% mesin B beroperasi tidak baik (cacat).

- a. Buatlah diagram pohon dari kejadian di atas
- b. Tentukan probabilitas terpilih mesin B yang tidak cacat
- c. Tentukan probabilitas mesin yang cacat.

Jawab:

a. Diagram pohon dari ilustrasi di atas:



b. Probabilitas terpilih mesin B yang tidak cacat

$$= 0,4 \times 0,7$$

$$= 0,28$$

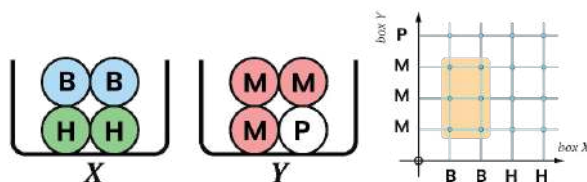
c. Probabilitas mesin yang cacat adalah mesin A cacat atau mesin B cacat, sehingga:

$$= 0,6 \times 0,2 + 0,4 \times 0,3$$

$$= 0,24$$

D. Kejadian Majemuk (Compound Event)

Kejadian majemuk adalah kejadian yang terdiri dari dua atau lebih kejadian sederhana dalam suatu ruang sampel. Artinya, kejadian majemuk merupakan gabungan dari beberapa kemungkinan hasil percobaan. Sebagai contoh, keranjang X berisi 2 bola biru dan 2 hijau. Keranjang Y berisi 1 bola putih dan 3 merah. Sebuah bola diambil secara acak dari masing-masing keranjang. Tentukan probabilitas diperoleh bola warna biru dari keranjang X dan bola warna merah dari keranjang Y.



Gambar 3.1. Ilustrasi Kejadian Majemuk

Dengan ilustrasi ruang sampel pada kisi dua dimensi yang ditunjukkan pada Gambar, kita bisa lihat bahwa terdapat 6 dari 16 kemungkinan di mana biru diambil dari X dan merah dari Y. Sehingga diperoleh:

$$P(\text{biru dari X dan merah dari Y}) = \frac{6}{16}$$

Di bagian ini kita mencari metode yang lebih cepat untuk menemukan probabilitas dalam situasi seperti di atas.

D.1. Kejadian Saling Bebas (Independent Event)

Kejadian dikatakan bebas (*independent*) jika terjadinya salah satu peristiwa tidak mempengaruhi peluang terjadinya kejadian lain.

Jika A dan B adalah kejadian bebas maka:

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

Aturan ini dapat diperluas untuk beberapa kejadian. Misal, jika A, B, dan C adalah kejadian bebas maka:

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

Contoh 3.4

Sebuah koin dan dadu dilempar bersamaan. Tentukan probabilitas dari mendapatkan gambar dan angka 3.

Jawab:

$$P(\text{Gambar} \cap 3) = P(\text{gambar}) \times P(3)$$

$$= \frac{1}{2} \times \frac{1}{6} = \frac{1}{12}$$

D.2. Kejadian Tak Bebas (Dependent Event)

Kejadian tak bebas apabila terjadinya salah satu kejadian mempengaruhi kejadian yang lain.

Dalam bentuk peluang gabungan:

$$P(A \text{ dan } B) \neq P(A) \times P(B)$$

Atau secara umum jika A dan B adalah kejadian tak bebas maka

$$P(B|A) \neq P(B)$$

Artinya, peluang B bergantung pada A.

Contoh 3.5

Sebuah keranjang berisi 4 permen warna merah dan 2 warna kuning. Dua buah permen diambil secara acak satu per satu tanpa pengembalian. Tentukan probabilitas dari:

- Terambil permen keduanya merah
- Pengambilan pertama merah dan kedua kuning

Jawab:

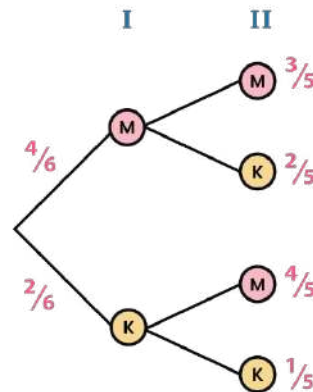
Berikut diagram pohon sebagai ilustrasi dari kejadian di atas di mana peluang kejadian kedua bergantung dari terjadinya kejadian pertama.

- $P(\text{keduanya merah})$

$$\begin{aligned} &= \frac{4}{6} \times \frac{3}{5} \\ &= \frac{2}{5} \end{aligned}$$

- $P(\text{pertama merah} \cap \text{kedua kuning})$

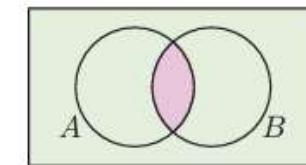
$$\begin{aligned} &= \frac{4}{6} \times \frac{2}{5} \\ &= \frac{4}{15} \end{aligned}$$



E. Himpunan dan Diagram Venn

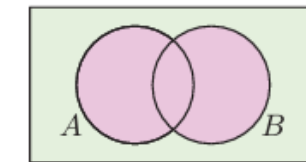
Diagram Venn berguna untuk merepresentasikan kejadian dalam ruang sampel. Diagram ini biasanya terdiri dari persegi panjang yang merepresentasikan ruang sampel lengkap U , dan lingkaran di dalamnya yang merepresentasikan kejadian tertentu. Selain itu, Diagram Venn dapat digunakan untuk menyelesaikan jenis soal probabilitas tertentu dan menetapkan sejumlah hukum probabilitas.

- $A \cap B$ menunjukkan **irisan** himpunan A dan B. Kumpulan ini berisi semua elemen yang sama untuk kedua himpunan. $A \cap B$ diarsir warna ungu pada gambar di bawah ini:



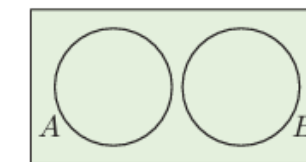
Gambar 3.2. Irisan himpunan A dan B

- $A \cup B$ menunjukkan **gabungan** himpunan A dan B. Himpunan ini berisi semua elemen yang termasuk dalam A atau B atau keduanya. $A \cup B$ diarsir warna ungu pada gambar di bawah ini:



Gambar 3.3. Gabungan himpunan A dan B

- Kejadian saling lepas (*mutually exclusive*) adalah himpunan yang tidak memiliki elemen yang sama. Sehingga $A \cap B = \emptyset$ di mana $\emptyset$ adalah himpunan kosong atau $P(A \cap B) = 0$.



Gambar 3.4. Kejadian Saling Lepas

- Hukum penjumlahan probabilitas adalah sebagai berikut:

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

- Peluang bersyarat (*Conditional Probability*) merupakan menentukan peluang kejadian A dengan syarat kejadian B sudah terjadi.

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

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

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

$P(A|B)$ dibaca: “Peluang A terjadi dengan syarat B sudah terjadi” atau “Peluang bersyarat A terhadap B.”

Contoh 3.6

Jika $P(A) = 0,6$, $P(A \cup B) = 0,7$ dan $P(A \cap B) = 0,3$.

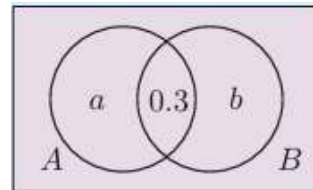
- Hitunglah $P(B)$.
- Tentukan apakah kejadian A dan B adalah kejadian saling bebas (*independent event*)?

Jawab:

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

$$0,7 = 0,6 + P(B) - 0,3$$

$$P(B) = 0,4$$



- Jika A dan B adalah kejadian saling bebas maka berlaku $P(A \cap B) = P(A) \times P(B)$.

$$\text{Kita hitung } P(A) \times P(B) = 0,6 \times 0,4 = 0,24$$

Karena nilai $P(A \cap B) = 0,3$ & tidak sama dengan nilai $P(A) \times P(B) = 0,24$, maka A dan B bukan kejadian saling bebas.

Contoh 3.7

Diketahui $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{3}$, dan $P(A \cup B) = p$. Tentukan nilai p jika:

- Kejadian A dan B adalah kejadian saling lepas (*mutually exclusive*)
- Kejadian A dan B adalah kejadian saling bebas (*independent event*)

Jawab:

- Jika A dan B adalah kejadian saling lepas maka $P(A \cap B) = 0$

$$\text{Sehingga, } P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$p = \frac{1}{2} + \frac{1}{3} - 0$$

$$p = \frac{5}{6}$$

- Jika A dan B adalah kejadian saling bebas maka $P(A \cap B) = P(A) \times P(B)$

$$\text{Sehingga, } P(A \cap B) = P(A) \times P(B)$$

$$= \frac{1}{2} \times \frac{1}{3} = \frac{1}{6}$$

$$\text{Maka, } P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$p = \frac{1}{2} + \frac{1}{3} - \frac{1}{6}$$

$$p = \frac{2}{3}$$

F. Teorema Bayes (*Bayes Theorem*)

Teorema Bayes adalah suatu aturan dalam probabilitas yang digunakan untuk menghitung peluang suatu kejadian berdasarkan informasi sebelumnya yang sudah diketahui. Jika $A_1, A_2, \dots, A_n$ adalah pembagian dari ruang sampel (partisi) dan B adalah suatu kejadian, maka:

$$P(A_i | B) = \frac{P(A_i) \cdot P(B|A_i)}{\sum_{j=1}^n P(A_j) \cdot P(B|A_j)}$$

Untuk dua kejadian A dan B :

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

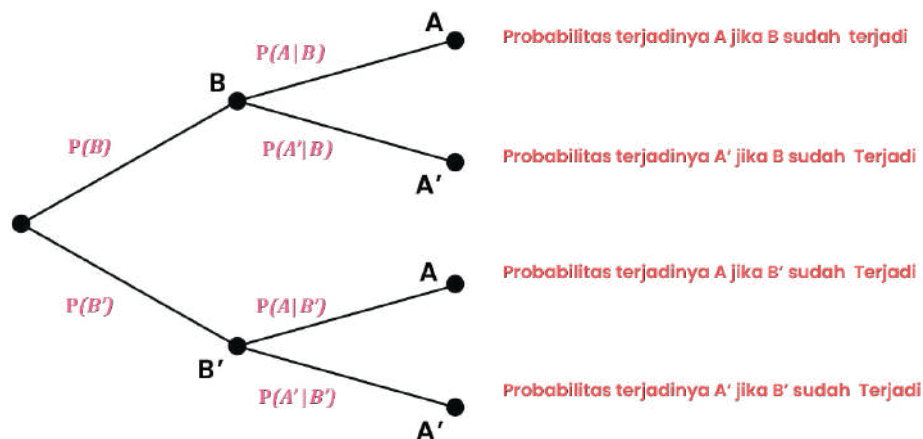
Karena $P(A \cap B) = P(B|A) \times P(A)$, maka:

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

Keterangan:

- $P(A|B)$: peluang A terjadi dengan syarat B telah terjadi (peluang bersyarat)
- $P(A)$: Peluang awal A (sebelum mengetahui B)
- $P(B|A)$: Peluang B terjadi jika A telah terjadi
- $P(B)$: Total peluang B

Untuk menggunakan rumus tersebut, kita perlu mengetahui peluang sebelumnya. Hal ini dapat ditemukan dengan menggunakan diagram pohon. Perhatikan diagram berikut:

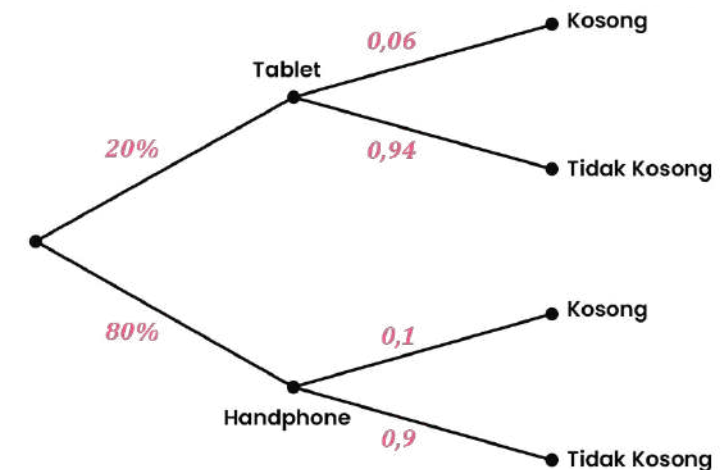


Gambar 3.5. Diagram Pohon Teorema Bayes

Contoh 3.8

Semua mahasiswa di kelas menggunakan perangkat pribadi untuk melakukan penelitian. 20% mahasiswa menggunakan tablet dan sisanya menggunakan handphone. Peluang tablet kehabisan daya selama pelajaran adalah 0,06 dan peluang handphone kehabisan daya adalah 0,1. Jika perangkat mahasiswa kehabisan daya selama pelajaran, cari peluang mahasiswa tersebut menggunakan tablet.

Jawab:



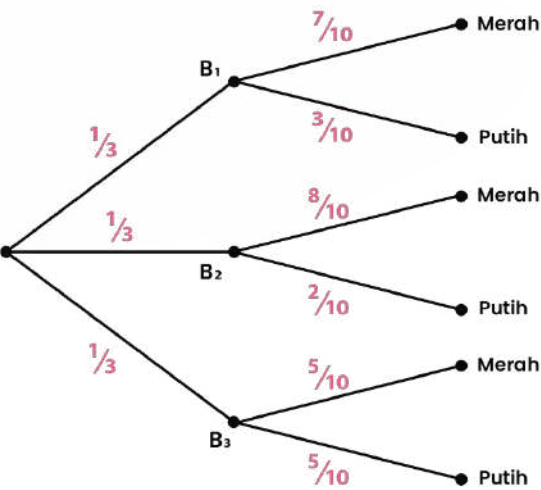
$$\begin{aligned} P(\text{Tablet}|\text{Kosong}) &= \frac{P(\text{Tablet} \cap \text{Kosong})}{P(\text{Kosong})} \\ &= \frac{0,2 \times 0,06}{0,2 \times 0,06 + 0,8 \times 0,1} \\ &= 0,13 \end{aligned}$$

Contoh 3.9

Sebanyak tiga box, B_1 , B_2 dan B_3 berisi bola merah dan putih. Box B_1 berisi tujuh bola merah dan tiga bola putih. Box B_2 berisi delapan bola merah dan dua bola putih. Box B_3 berisi lima bola merah dan lima bola putih. Salah satu box dipilih secara acak, lalu sebuah bola dipilih dari box tersebut. Jika bola tersebut berwarna merah, carilah peluang bola tersebut berasal dari box B_1 .

Jawab:

Diagram pohon dari ilustrasi di atas



$$\begin{aligned} P(B_1 | \text{Merah}) &= \frac{P(B_1 \cap \text{Merah})}{P(\text{Merah})} \\ &= \frac{\frac{1}{3} \times \frac{7}{10}}{\left(\frac{1}{3} \times \frac{7}{10}\right) + \left(\frac{1}{3} \times \frac{8}{10}\right) + \left(\frac{1}{3} \times \frac{5}{10}\right)} \\ &= \frac{7}{20} \end{aligned}$$

LATIHAN 3

Probabilitas

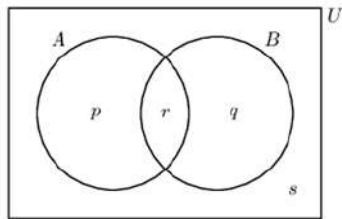
- Sebuah dadu dilempar dan pemutar berbentuk lingkaran yang diberi label A, B, C, dan D diputar. Tentukan peluang untuk memperoleh:
 - Angka 2 dan huruf konsonan
 - Angka genap dan huruf C
 - Angka ganjil atau huruf vocal
- Sebuah keranjang berisi kelereng 3 berwarna merah, 4 kuning, dan 5 biru. Dua buah kelereng akan diambil secara acak dari keranjang tanpa pengembalian. Tentukan peluang bahwa:
 - Keduanya biru
 - Keduanya memiliki warna yang sama
 - Paling sedikit satu berwarna merah
 - Tepat satu berwarna kuning
- Peluang seseorang akan hidup dalam 25 tahun kedepan adalah $\frac{3}{5}$ dan peluang bahwa istrinya akan hidup adalah $\frac{2}{3}$. Tentukan probabilitas dalam 25 tahun:
 - Keduanya akan hidup
 - Setidaknya satu akan hidup
 - Hanya istrinya yang akan hidup
- Sebuah survei terhadap 200 orang yang melibatkan 90 orang wanita ditemukan bahwa 60 orang merokok dan 40 diantaranya adalah pria.
 - Gunakan informasi tersebut untuk melengkapi tabel berikut



	Wanita	Pria	Total
Perokok			
Bukan Perokok			
Total			

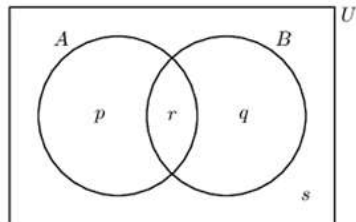
- b. Seseorang dipilih secara acak. Tentukan probabilitas bahwa orang tersebut
 - i. Wanita bukan perokok
 - ii. seorang laki-laki yang berasal dari bukan perokok.
- c. Jika dua orang dari survei dipilih secara acak, hitunglah probabilitas bahwa:
 - i. Kedua dari mereka adalah Wanita bukan perokok
 - ii. Salah satu adalah perokok dan lainnya bukan perokok

5. Diagram Venn berikut ini menunjukkan dua kejadian A dan B , di mana $P(A) = 0,3$, $P(A \cap B) = 0,2$, dan $P(A \cup B) = 0,7$. Nilai p, q, r dan s merupakan probabilitas.



- a. Tentukan nilai r
- b. Tentukan nilai p, q , dan s
- c. Hitunglah $P(B)$

6. Diketahui Diagram Venn menunjukkan dua kejadian A dan B , di mana $P(A) = 0,3$, $P(B) = 0,8$, dan $P(A \cap B) = 0,2$. Nilai p, q, r dan s merupakan probabilitas



- a. Tentukan nilai p, q, r dan s
- b. Hitunglah $P(A|B^c)$

7. Diketahui ada 3 kota produksi beras menyuplai ke ibu kota. Karawang menyuplai 40%, Cianjur menyuplai 50%, dan Subang menyuplai sisanya. Dari pengalaman diketahui produksi beras rusak dari Karawang 5%, Cianjur 6%, dan Subang 4%. Jika diambil secara acak suatu produk beras, hitunglah:

- a. Probabilitas produk tersebut rusak
 - b. Bila ternyata diambil produk rusak, tentukan probabilitas berasal dari Kota Cianjur
8. Setiap pagi Rudi berjalan kaki, bersepeda atau berkendara ke tempat kerja. Probabilitas Rudi berjalan adalah 0,2 dan Rudi memiliki kemungkinan yang sama untuk bersepeda atau berkendara. Ketika Rudi berjalan, peluang Rudi terlambat adalah 0,05. Ketika Rudi bersepeda, peluang Rudi terlambat adalah 0,1. Ketika Rudi mengemudi, peluang Rudi terlambat adalah 0,2. Tentukan:
 - a. Gambarkan diagram pohon yang merepresentasikan situasi di atas
 - b. Tentukan probabilitas Rudi telat bekerja
 - c. Bila ternyata (*Given*) Rudi terlambat, tentukan peluang Rudi berjalan ke tempat kerja

BAB 4

Distribusi Probabilitas Variabel Acak Diskrit dan Kontinu

Pendahuluan

Kemampuan Akhir yang Diharapkan	Mahasiswa memahami masalah distribusi probabilitas variabel acak diskrit dan kontinu
Bahan Kajian	- Distribusi probabilitas diskrit - Ekspektasi dan varians variabel acak diskrit - Transformasi linear variabel acak - Distribusi probabilitas kontinu
Metode Pembelajaran	Ceramah interaktif, studi kasus
Pengalaman Belajar Mahasiswa	- Menganalisis tabel distribusi peluang untuk variabel acak diskrit - Mengevaluasi distribusi probabilitas kontinu melalui interpretasi kurva - Mengerjakan studi kasus terkait distribusi peluang dalam konteks nyata
Kriteria Penilaian (Indikator)	- Mampu membedakan variabel acak diskrit dan kontinu dengan contoh nyata. - Menggunakan distribusi diskrit untuk menghitung probabilitas, ekspektasi, dan varians data diskrit - Menghitung dan menginterpretasikan peluang dari distribusi kontinu

Tujuan Pembelajaran

1. Mahasiswa memahami masalah distribusi probabilitas variabel acak diskrit
2. Mahasiswa memahami masalah distribusi probabilitas variabel acak kontinu

A. Variabel Acak

Variabel acak dinyatakan dalam bentuk angka hasil-hasil yang mungkin terjadi pada suatu percobaan acak. Variabel acak digunakan untuk mengukur kejadian acak sehingga memiliki banyak nilai. Variabel acak dapat bersifat diskrit atau kontinu (nilai dalam interval kontinu).

Suatu variabel acak diskrit X memiliki sekumpulan kemungkinan nilai yang berbeda. Data diskrit diperoleh dari hasil menghitung (bukan mengukur) sehingga antara satu bilangan dengan bilangan lain tidak dapat ditempatkan bilangan desimal atau rasional. Pada variabel acak diskrit, X bisa memiliki arti misalnya:

- jumlah bola lampu rusak dalam suatu pembelian di sebuah toko
- jumlah rumah di kota A yang memiliki saklar pengaman listrik

Variabel acak kontinu berbeda dengan variabel acak diskrit. Suatu variabel acak kontinu X dapat memiliki nilai apa pun dalam interval bilang riil. Data kontinu diperoleh dari hasil pengukuran sehingga data kontinu dapat berupa bilangan desimal atau rasional. Variabel acak kontinu X bisa memiliki arti, misalnya:

- tinggi badan pada interval $130 \leq X \leq 190$ cm
- volume air dalam tangka pada interval $20 \leq X \leq 100$ m³

B. Distribusi Probabilitas Diskrit

Jika X adalah variabel acak dengan ruang sampel $\{x_1, x_2, x_3, \dots, x_n\}$ dan probabilitas yang sesuai $\{p_1, p_2, p_3, \dots, p_n\}$ sehingga $P(X = x_i) = p_i$, $i = 1, \dots, n$ maka:

- $0 \leq p_i \leq 1$ untuk seluruh $i = 1$ sampai n
- $\sum_{i=1}^n p_i = p_1 + p_2 + p_3 + \dots + p_n = 1$
- $\{p_1, p_2, p_3, \dots, p_n\}$ menggambarkan distribusi probabilitas dari X

Contoh 4.1

Diketahui $P(x) = \frac{x^2+1}{34}$ untuk $x = 1, 2, 3, 4$

- Tunjukkan $P(x)$ merupakan fungsi distribusi peluang
- Buatlah tabel fungsi distribusi peluang tersebut

Jawab:

a. $P(1) = \frac{1^2+1}{34} = \frac{2}{34}$

$$P(2) = \frac{2^2+1}{34} = \frac{5}{34}$$

$$P(3) = \frac{3^2+1}{34} = \frac{10}{34}$$

$$P(4) = \frac{4^2+1}{34} = \frac{17}{34}$$

$$\text{Total : } \frac{2}{34} + \frac{5}{34} + \frac{10}{34} + \frac{17}{34} = 1$$

Karena seluruh probabilitas bernilai $0 \leq P(x_i) \leq 1$ dan $\sum P(x_i) = 1$

$\therefore P(x)$ adalah fungsi distribusi probabilitas

- Tabel fungsi distribusi probabilitas:

x	1	2	3	4
$P(X = x)$	$\frac{2}{34}$	$\frac{5}{34}$	$\frac{10}{34}$	$\frac{17}{34}$

Contoh 4.2

Sebuah toko buku mencatat jumlah buku yang dibeli pelanggannya per hari. Diperoleh data sebanyak 23% membeli 1 buku, 38% membeli 2 buku, 21% membeli 3 buku, 13% membeli 4 buku, dan 5% membeli 5 buku.

- Buatlah tabel probabilitas untuk variabel acak di atas
- Tentukan $P(X < 3)$

Jawab:

- Berikut tabel probabilitas:

x	1	2	3	4	5
$P(X = x)$	0,23	0,38	0,21	0,13	0,05

- $P(X > 3)$ memiliki arti yaitu probabilitas pembelian buku lebih dari 3 yaitu pembelian 4 buku atau pembelian 5 buku, sehingga:

$$\begin{aligned} P(X > 3) &= P(4) + P(5) \\ &= 0,13 + 0,05 = 0,18 \end{aligned}$$

Contoh 4.3

Tabel di bawah ini menampilkan distribusi probabilitas X , jumlah retakan pada pipa saluran pembuangan.

x	0	1	2	3
$P(X = x)$	0,45	0,3	0,15	k

- Tentukan nilai k
- Tentukan $P(X \leq 1)$
- Tentukan $P(X = 0 | X \leq 1)$

Jawab:

- $0,45 + 0,3 + 0,15 + k = 1$
 $k = 0,1$

$$\begin{aligned} \text{b. } P(X \leq 1) &= P(0) + P(1) \\ &= 0,45 + 0,3 \\ &= 0,75 \end{aligned}$$

$$\begin{aligned} \text{c. } P(X = 0 | X \leq 1) &= \frac{0,45}{0,75} \\ &= 0,6 \end{aligned}$$

C. Ekspektasi Variabel Acak Diskrit

Ekspektasi (nilai harapan) $E(X)$ dari variabel acak diskrit (juga disebut mean, μ) adalah nilai rata-rata yang akan diperoleh pada pengulangan percobaan yang dilakukan tak terhingga.

- Jika terdapat n percobaan dan suatu kejadian memiliki peluang p untuk terjadi pada setiap percobaan, maka berapa kali kejadian tersebut diharapkan terjadi adalah np
- Secara umum, **ekspektasi** dari variabel acak X adalah

$$E(X) = \sum_{i=1}^n x_i p_i$$

Contoh 4.4

Setiap kali pemain bola menendang bola untuk mencetak gol, Rudi memiliki peluang $\frac{3}{4}$ untuk berhasil. Pada suatu pertandingan tertentu, Rudi melakukan 12 tendangan ke arah gawang. Berapa banyak gol yang diharapkan pada pertandingan tersebut?

Jawab:

Ekspektasi banyaknya gol adalah

$$\begin{aligned} E(X) &= np \\ &= 12 \times \frac{3}{4} \\ &= 9 \end{aligned}$$

Contoh 4.5

Tentukan nilai rata-rata data penjualan toko buku pada Contoh 1 dan jelaskan interpretasi hasil tersebut.

x	1	2	3	4	5
$P(X = x)$	0,23	0,38	0,21	0,13	0,05

Jawab:

$$\begin{aligned} E(X) = \mu &= \sum_{i=1}^n x_i p_i \\ &= 1(0,23) + 2(0,38) + 3(0,21) + 4(0,13) + 5(0,05) \\ &= 2,39 \end{aligned}$$

Dalam jangka panjang, rata-rata banyaknya buku yang dibeli per pelanggan adalah 2,39.

D. Varians Variabel Acak Diskrit

Varians dari variabel acak X adalah:

- $Var(X) = E(X^2) - [E(X)]^2$
- Standar deviasi yaitu $\sqrt{Var(X)}$

Contoh 4.6

Suatu variabel acak diskrit X memiliki distribusi peluang seperti yang ditunjukkan pada tabel berikut.

x	0	1	2	3
$P(X = x)$	0,2	0,3	0,4	0,1

Tentukan $Var(X)$.

Jawab:

$$\begin{aligned} E(X) &= 0(0,2) + 1(0,3) + 2(0,4) + 3(0,1) \\ &= 1,4 \end{aligned}$$

$$E(X^2) = 0(0,2) + 1(0,3) + 4(0,4) + 9(0,1) \\ = 2,8$$

$$Var(X) = E(X^2) - [E(X)]^2 \\ = 2,8 - 1,4^2 \\ = 0,84$$

E. Transformasi Linear Pada Variabel Acak

- $E(aX + b) = aE(X) + b$
- $Var(aX + b) = a^2Var(X)$

Contoh 4.7

Suatu variabel acak X memiliki nilai ekspektasi 12,5 dan varians 4,8. Jika variabel acak Y dinyatakan sebagai $Y = 3X - 2$, maka tentukan nilai ekspektasi (mean) dan varians dari Y .

Jawab:

$$\begin{aligned} E(Y) &= E(3X - 2) & Var(Y) &= Var(3X - 2) \\ &= 3E(X) - 2 & &= 3^2 \times 4,8 = 43,2 \\ &= 3 \times 12,5 - 2 = 35,5 & & \end{aligned}$$

F. Distribusi Probabilitas Kontinu

Fungsi peluang pada variabel acak kontinu tidak dapat disajikan dalam bentuk tabel, namun ditampilkan dalam bentuk rumus/fungsi. Peluang suatu variabel acak kontinu dengan mengambil nilai dalam interval tertentu yang dinyatakan sebagai luas daerah dibawah grafik fungsi. Fungsi peluang $f(x)$ untuk variabel acak kontinu X disebut fungsi kepadatan peluang atau *probability density function (pdf)*.

Untuk variabel acak kontinu X dengan fungsi kepadatan peluang $f(x)$, maka

- $f(x) \geq 0$ untuk seluruh x dan $\int_{-\infty}^{\infty} f(x)dx = 1$
- $P(a < X < b) = \int_a^b f(x)dx$
- Ekspektasi variabel acak kontinu $E(X) = \int_{-\infty}^{\infty} x f(x)dx$

Contoh 4.8

Suatu variabel acak kontinu X memiliki fungsi kepadatan peluang sebagai berikut:

$$f(x) = \begin{cases} k(5x - x^2), & \text{untuk } 0 \leq x \leq 5 \\ 0, & \text{untuk lainnya} \end{cases}$$

- Tentukan nilai k
- Tentukan $P(1 < X < 3)$
- Tentukan $E(X)$

Jawab:

- $$\int_0^5 k(5x - x^2)dx = 1$$

$$k \times \left[\frac{5}{2}x^2 - \frac{1}{3}x^3 \right]_0^5 = 1$$

$$k \times \left\{ \left[\frac{5}{2}(5)^2 - \frac{1}{3}(5)^3 \right] - \left[\frac{5}{2}(0)^2 - \frac{1}{3}(0)^3 \right] \right\} = 1$$

$$k = \frac{6}{125}$$
- $$P(1 < X < 3) = \int_1^3 \frac{6}{125}(5x - x^2)dx$$

$$= \frac{6}{125} \times \left[\frac{5}{2}x^2 - \frac{1}{3}x^3 \right]_1^3$$

$$= \frac{6}{125} \left\{ \left[\frac{5}{2}(3)^2 - \frac{1}{3}(3)^3 \right] - \left[\frac{5}{2}(1)^2 - \frac{1}{3}(1)^3 \right] \right\}$$

$$= 0,544$$

$$\begin{aligned}
c. E(X) &= \int_0^5 x f(x) dx \\
&= \int_0^5 x \left[\frac{6}{125} (5x - x^2) \right] dx \\
&= \frac{6}{125} \int_0^5 5x^2 - x^3 dx \\
&= \frac{6}{125} \times \left[\frac{5}{3} x^3 - \frac{1}{4} x^4 \right]_0^5 \\
&= \frac{6}{125} \left\{ \left[\frac{5}{3} (5)^3 - \frac{1}{4} (5)^4 \right] - \left[\frac{5}{3} (0)^3 - \frac{1}{4} (0)^4 \right] \right\} = \frac{5}{2}
\end{aligned}$$

LATIHAN 4

Distribusi Probabilitas Variabel Acak Diskrit dan Kontinu

1. Suatu variabel acak diskrit X memiliki distribusi probabilitas $P(X)$ yang dideskripsikan pada tabel berikut:

x	0	1	2	3	4
$P(X = x)$	0,1	0,3	0,45	0,1	k

Tentukan:

- Nilai k
 - $P(2 \leq X < 4)$
 - $P(X \geq 3)$
2. Diketahui $P(X = x) = \frac{a}{x^2 + 1}$ dimana $x = 0, 1, 2, 3$ adalah fungsi distribusi probabilitas
- Hitunglah nilai a
 - Tentukan $P(X < 2)$
3. Tabel di bawah ini menunjukkan distribusi probabilitas X

x	0	1	2	3	4
$P(X = x)$	0,1	k	0,3	0,3	0,1

Tentukan:

- $P(X > 2)$
- $P(X = 3 | X > 2)$

4. Tabel di bawah ini menunjukkan distribusi probabilitas Y

y	1	4	9	16	25
$P(Y = y)$	0,1	k	0,2	$2k$	0,1

Tentukan:

- $P(Y \leq 16)$
- $P(Y > 4 | Y \leq 16)$

5. Suatu variabel acak X memiliki distribusi probabilitas yang diketahui adalah

$$P(X = x) = \frac{x^2 + x}{20} \text{ untuk } x = 1, 2, 3$$

Tentukan mean μ untuk distribusi tersebut.

6. Dalam survei acak terhadap pemilihan calon walikota di suatu wilayah ditampilkan pada tabel berikut

Calon A	Calon B	Calon C
165	87	48

- Tentukan masing-masing probabilitas bahwa seorang pemilih yang dipilih secara acak akan memilih calon A, B, dan C?
 - Jika terdapat 7500 orang daftar pemilih, tentukan berapa banyak masing-masing dari daftar pemilih yang berekspektasi memilih calon A, B, dan C?
7. Tentukan mean dan varians dari variabel acak Y , jika:
- $E(X) = 11, Var(X) = 5, Y = 4X - 3$
 - $E(X) = 9, Var(X) = 2, Y = \frac{1}{3}X + 1$

c. $E(X) = 14, Var(X) = 6, Y = \frac{X-2}{6}$

8. Fungsi kepadatan peluang dari variabel acak X diberikan sebagai berikut:

$$f(x) = \begin{cases} kx^2, & \text{untuk } 0 \leq x \leq 3 \\ 0, & \text{untuk lainnya} \end{cases}$$

- a. Tentukan nilai k
- b. Tentukan $P(1 \leq X \leq 2)$
- c. Tentukan $E(X)$
9. Tentukan nilai k dan $P\left(X < \frac{k}{2}\right)$ dari fungsi kepadatan peluang X atau *probability density function* (pdf) berikut:

$$f(x) = \begin{cases} x^3, & \text{untuk } 0 \leq x \leq k \\ 0, & \text{untuk lainnya} \end{cases}$$

10. Diberikan distribusi probabilitas dari varaibel acak X dalam bentuk tabel berikut.

x	0	1	2	3
$P(X = x)$	p	q	0,2	0,2

Jika $Var(X) = 0,85$, tentukan nilai $E(X)$.

BAB 5

Distribusi Binomial & Poisson

Pendahuluan

Kemampuan Akhir yang Diharapkan	Mahasiswa memahami masalah yang berkaitan dengan distribusi binomial dan poisson serta memahami rata-rata varians dari distribusi tersebut.
Bahan Kajian	- Distribusi binomial - Distribusi poisson
Metode Pembelajaran	Demonstrasi <i>software R & Problem Based Learning</i>
Pengalaman Belajar Mahasiswa	- Mengamati fenomena yang sesuai dengan distribusi binomial dan Poisson melalui studi kasus - Menghitung peluang distribusi binomial dan Poisson menggunakan rumus secara manual & bantuan <i>software</i> statistik - Menentukan nilai harapan (mean) dan varians dari distribusi
Kriteria Penilaian (Indikator)	- Dapat menentukan situasi yang sesuai dengan distribusi binomial atau poisson. - Menghitung peluang dengan benar distribusi binomial/poisson menggunakan rumus secara manual & bantuan <i>software</i> statistik - Menghitung rata-rata dan varians dari distribusi binomial dan Poisson secara tepat.

Tujuan Pembelajaran

1. Mahasiswa memahami masalah yang berkaitan dengan distribusi binomial & poisson
2. Mahasiswa memahami rata-rata dan varians dari distribusi binomial & poisson

A. Distribusi Binomial

Distribusi binomial adalah distribusi probabilitas diskrit dengan mempertimbangkan sebuah eksperimen memiliki dua kemungkinan hasil yaitu **sukses** atau **gagal** (dalam artian **ya/tidak**), misalnya:

- suatu tes yang diikuti, hanya ada dua hasil yaitu lulus atau gagal
- perkiraan cuaca hujan, hanya ada dua hasil yaitu hujan atau tidak hujan
- pelemparan sebuah koin, hanya ada dua hasil yaitu kepala atau ekor

Jika kita mengulangi percobaan dalam beberapa percobaan yang bersifat bebas (*independent*) dan memiliki dua kemungkinan, kita menyebutnya percobaan binomial. Distribusi binomial memiliki parameter n dan p dimana parameter p merupakan probabilitas kejadian sukses, sedangkan $(1 - p)$ probabilitas kejadian gagal. Parameter n merupakan jumlah percobaan.

Jika X adalah variabel acak dari percobaan binomial dengan parameter n dan p , maka $X \sim B(n, p)$ dimana simbol tilde ($\sim$) berarti “didistribusikan sebagai”. Rumus distribusi binomial yaitu:

$$P(X = r) = \binom{n}{r} p^r (1 - p)^{n-r} \text{ untuk } r = 0, 1, 2, 3, 4, \dots, n.$$
dimana r merupakan jumlah keberhasilan dengan n percobaan.

Pada distribusi binomial dapat dibagi menjadi dua kasus yaitu:

- Fungsi distribusi probabilitas binomial atau *binomial probability distribution function* (BPD) untuk mencari $P(X = r)$. Pada Contoh 5.1 akan ditunjukkan contoh kasus BPD.

- Fungsi distribusi probabilitas kumulatif binomial atau *binomial cumulative distribution function* (BCD) untuk mencari $P(X \leq r)$ atau $P(X \geq r)$. Contoh Kasus BCD ditunjukkan pada Contoh 5.2.

Contoh 5.1

Seorang penjual lampu mengatakan bahwa di antara seluruh barang dagangannya, probabilitas lampu yang **rusak** adalah **3%**. Terdapat pelanggan yang membeli lampu sebanyak 5 buah dan memilih secara acak. Jika data mengikuti distribusi binomial, hitunglah probabilitas:

- a. lampu rusak tepat sama dengan 3
- b. lampu tidak rusak tepat sama dengan 1

Jawab:

- a. Misal X = lampu rusak $\rightarrow p = 3\%$

Lampu rusak tepat sama dengan 3
artinya $P(X = 3)$

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

$$P(X = 3) = \binom{5}{3} (0,03)^3 (0,97)^{5-3} = 0,000254$$



```
> dbinom(3,5,0.03)
[1] 0.000254043
```

- b. Misal X = lampu tidak rusak $\rightarrow p = 97\%$

Lampu tidak rusak sama dengan 1
artinya $P(X = 1)$

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

$$P(X = 1) = \binom{5}{1} (0,97)^1 (0,03)^4 = 3,9285 \times 10^{-6}$$

```
> dbinom(1,5,0.97)
[1] 3.9285e-06
```



Contoh 5.2

Seorang penjual toko buah apel mengatakan bahwa probabilitas apel yang busuk adalah 5%. Rani membeli apel di toko tsb sebanyak 10 buah dan memilih secara acak. Jika data mengikuti distribusi binomial maka hitunglah probabilitas:

- Apel busuk kurang dari sama dengan 2
- Apel tidak busuk paling sedikit 7

Jawab:

- Misal X = apel busuk $\rightarrow p = 5\%$

$$\begin{aligned}P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\&= \binom{10}{0} (0,05)^0 (0,95)^{10} + \binom{10}{1} (0,05)^1 (0,95)^9 \\&\quad + \binom{10}{2} (0,05)^2 (0,95)^8 \\&= 0,9884964\end{aligned}$$

```
> pbinom(2,10,0.05)
[1] 0.9884964
```



- Misal X = apel tidak busuk $\rightarrow p = 5\%$

$$\begin{aligned}P(X \geq 7) &= P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10) \\&= \binom{10}{7} (0,95)^7 (0,05)^3 + \binom{10}{8} (0,95)^8 (0,05)^2 + \binom{10}{9} (0,95)^9 (0,05)^1 \\&\quad + \binom{10}{10} (0,95)^{10} (0,05)^0 \\&= 0,9989715\end{aligned}$$

```
> 1-pbinom(6,10,0.05)
[1] 0.9989715
```



B. Mean dan Standar Deviasi Distribusi Binomial

Jika X adalah variabel acak binomial dengan parameter n dan p , maka $X \sim B(n, p)$.

- Mean** dari X adalah $\mu = E(X) = np$
 - Standar Deviasi** dari X adalah $\sigma = \sqrt{np(1-p)}$
 - Varians** dari X adalah $\sigma^2 = np(1-p)$
- σ dibaca Sigma kecil, σ^2 = Sigma kuadrat, Σ = Sigma besar

Contoh 5.3

Diketahui 5% dari sekumpulan baterai adalah rusak. Sebanyak 80 baterai diambil secara acak sebagai sampel. Carilah rata-rata dan simpangan baku dari jumlah baterai rusak dalam sampel tersebut.

Jawab:

Diketahui: $n = 80$ & $p = 5\% = \frac{1}{20}$

Jika X adalah variabel acak untuk banyaknya baterai yang rusak maka $X \sim B\left(80, \frac{1}{20}\right)$.

Jadi, $\mu = np$

$$\begin{aligned}&= 80 \times \frac{1}{20} \\&= 4\end{aligned}$$

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

$$\begin{aligned}&= \sqrt{80 \times \frac{1}{20} \times \left(1 - \frac{1}{20}\right)} \\&= 1,949\end{aligned}$$

Contoh 5.4

Suatu koin dengan permukaan kepala atau ekor dilemparkan ke atas sebanyak 4 kali, di mana probabilitas munculnya kepala sama dengan probabilitas munculnya ekor yaitu $\frac{1}{2}$. Jika X adalah banyaknya kepala yang muncul, carilah nilai rata-rata $E(X)$ dan simpangan baku dengan menggunakan cara:

- Perhitungan secara langsung.
- Dengan menggunakan rumus $E(X) = np$ dan $\sigma = \sqrt{np(1-p)}$.

Jawab:

- Jika X adalah banyaknya kepala yang muncul maka tabel distribusi probabilitas yaitu:

x	0	1	2	3	4
$P(X=x)$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

$$E(X) = \mu = \sum_{i=1}^n x_i p_i$$

$$= (0) \left(\frac{1}{16}\right) + (1) \left(\frac{4}{16}\right) + (2) \left(\frac{6}{16}\right) + (3) \left(\frac{4}{16}\right) + (4) \left(\frac{1}{16}\right) = 1,964 \approx 2$$

$$E(X^2) = (0) \left(\frac{1}{16}\right) + (1) \left(\frac{4}{16}\right) + (4) \left(\frac{6}{16}\right) + (9) \left(\frac{4}{16}\right) + (16) \left(\frac{1}{16}\right) = 5$$

$$Var(X) = E(X^2) - [E(X)]^2$$

$$Var(X) = 5 - 2^2 = 1 \text{ maka } \sigma = \sqrt{1} = 1$$

- Dengan menggunakan rumus:

$$E(X) = np$$

$$= 4 \times \frac{1}{2}$$

$$= 2$$

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

$$= \sqrt{4 \times \frac{1}{2} \times \frac{1}{2}}$$

$$= 1$$

C. Distribusi Poisson

Distribusi Poisson biasanya melibatkan jumlah n yang besar dengan p yang kecil. Distribusi ini digunakan untuk menghitung nilai probabilitas suatu kejadian dalam suatu selang waktu dan daerah tertentu. Misalnya, banyaknya dering telpon dalam 1 hari pada *customer service*, banyaknya

rumah terbakar selama bulan Januari, atau banyaknya pasien yang berobat per minggu.

Rumus untuk menyelesaikan distribusi Poisson adalah sebagai berikut:

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

Di mana λ = rata-rata distribusi (lambda)

$x = 0, 1, 2, 3, \dots$ (menuju tak hingga)

$E(X) = \lambda$ dan $Var(X) = \lambda$

Contoh 5.5

Banyaknya telpon berdering pada *customer service* suatu Bank pada pukul 09.00 – 10.00 mengikuti distribusi Poisson dengan banyaknya telpon berdering rata-rata 5 dering. Tentukan probabilitas bahwa terdapat 10 dering.

Jawab:

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

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

$$= 0,0181327$$

```
> dpois(10, lambda=5)
[1] 0.01813279
```



Contoh 5.6

Suatu produk minuman bersoda melakukan promosi dengan memberikan hadiah menarik pada 5 botol dari 1000 botol yang diproduksi. Jika X adalah banyaknya botol yang memiliki tulisan “Hadiah” dan pabrik memproduksi 20 botol, hitunglah probabilitas:

- Tidak mendapatkan botol hadiah
- Tepat mendapatkan 2 botol hadiah
- Tepat mendapatkan 4 botol hadiah
- Mendapatkan paling banyak 2 botol hadiah
- Paling sedikit mendapat 3 botol hadiah

Jawab:

Pertama, hitung probabilitas botol hadiah yaitu $p = \frac{5}{1000} = 0,005$
 karena $n = 20$ maka $\lambda = np = 20(0,005) = 0,1$

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

$$P(0) = \frac{(0,1)^0 e^{-0,1}}{0!} = 0,9048$$

$$P(1) = \frac{(0,1)^1 e^{-0,1}}{1!} = 0,09048$$

$$P(2) = \frac{(0,1)^2 e^{-0,1}}{2!} = 0,0045$$

$$P(3) = \frac{(0,1)^3 e^{-0,1}}{3!} = 0,0001508$$

$$P(4) = \frac{(0,1)^4 e^{-0,1}}{4!} = 3,77 \times 10^{-6} \approx 0$$

Maka tabel distribusi probabilitas:

x	0	1	2	3	4	5
$P(X = x)$	0,9048	0,09048	0,0045	0,0001508	0,0000	0,0000

- Tidak mendapatkan botol hadiah
 $P(X = 0) = 0,9048$
- Tepat mendapatkan 2 botol hadiah
 $P(X = 2) = 0,0045$

- Tepat mendapatkan 4 botol hadiah
 $P(X = 4) = 0,0000$ (tidak mungkin terjadi)
- Mendapatkan paling banyak 2 botol hadiah:

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

$$= 0,9048 + 0,09048 + 0,0045$$

$$= 0,9998$$

- Paling sedikit mendapat 3 botol hadiah:

$$P(X \geq 3) = P(3) + P(4) + P(5)$$

$$= 0,0001508 + 0,0000 + 0,0000$$

$$= 0,00015$$

```
> dpois(0,lambda=0.1) #no a
[1] 0.9048374
> dpois(2,lambda=0.1) #no b
[1] 0.004524187
> dpois(4,lambda=0.1) #no c
[1] 3.770156e-06
> ppois(2,lambda=0.1) #no d
[1] 0.9998453
> 1-ppois(2,lambda=0.1) #no d
[1] 0.0001546531
```



LATIHAN 5

Distribusi Binomial & Poisson

1. Seorang produsen keramik lantai menemukan bahwa 18% barang yang diproduksi cacat. Selama pemeriksaan, produsen secara acak memilih 7 barang secara acak. Carilah probabilitas bahwa produsen menemukan:
 - a. produk cacat tepat sama dengan 1
 - b. produk cacat paling sedikit 2
 - c. produk cacat antara 4 sampai dengan 7
 - d. produk tidak cacat sama dengan 4
 - e. tidak ada produk cacat
 - f. ekspektasi & varians produk cacat
2. Sebanyak 10 mahasiswa yang akan mengikuti ujian skripsi dan probabilitas kelulusannya adalah 84%. Tentukan probabilitas:
 - a. lulus tepat sama dengan 2 orang
 - b. paling banyak 6 orang lulus
 - c. paling sedikit 4 orang lulus
 - d. yang akan lulus antara 2 sampai 5 orang
3. Pada suatu ujian pilihan ganda terdiri atas 20 pertanyaan. Setiap pertanyaan memiliki empat jawaban dengan hanya satu jawaban benar. Untuk setiap pertanyaan, Jerry secara acak memilih satu dari empat jawaban.
 - a. Tentukan jumlah pertanyaan yang diharapkan dijawab dengan benar oleh Jerry
 - b. Hitung probabilitas bahwa Jerry menjawab:
 - i. tepat empat pertanyaan dengan benar
 - ii. lebih dari empat pertanyaan dengan benar
4. Suatu distribusi binomial $X \sim B(n, p)$ diketahui memiliki mean 30 dan varians 22,5. Tentukan:
 - a. Nilai n dan p
 - b. $P(X = 25)$
 - c. $P(X \geq 25)$
 - d. $P(15 \leq X \leq 25)$
5. PT. AB Teknologi memproduksi motor DC di mana terdapat 100 motor rusak dari 10.000 produksi. Pada suatu waktu, PT. AB Teknologi memproduksi 500 motor DC. Hitunglah probabilitas:
 - a. Tepat mendapat 5 motor DC rusak
 - b. Paling banyak 2 motor DC rusak
 - c. Paling sedikit mendapat 2 motor DC rusak
6. Emma membeli ponsel baru sebagai hadiah ulang tahun dan menerima pesan teks dari teman-temannya. Diasumsikan bahwa jumlah pesan teks yang diterima Emma setiap hari mengikuti distribusi Poisson dengan rata-rata 5 pesan. Hitunglah:
 - a. Probabilitas bahwa pada hari tertentu Emma menerima lebih dari 7 pesan
 - b. Perkiraan atau ekspektasi jumlah hari dalam seminggu di mana Emma menerima lebih dari 7 pesan
 - c. Probabilitas bahwa Emma menerima kurang dari 30 pesan teks selama seminggu
7. Diketahui jumlah kucing yang mengunjungi rumah Helena setiap minggu mengikuti distribusi Poisson dengan rata-rata 0,6. Tentukan peluang:
 - a. Pada minggu tertentu tidak ada kucing mengunjungi rumah Helena
 - b. Pada suatu minggu tertentu setidaknya ada tiga kucing yang mengunjungi rumah Helena
 - c. Selama periode empat minggu, tidak lebih dari lima kucing akan mengunjungi rumah Helena
 - d. Selama periode 12 minggu akan ada tepat empat minggu di mana setidaknya satu kucing akan mengunjungi rumah Helena

8. Jumlah pesan teks yang diterima seorang Staff administrasi Rumah Sakit selama interval waktu dapat dimodelkan dengan distribusi Poisson dengan rata-rata 6 pesan per jam.
- Tentukan peluang bahwa Staff akan menerima 8 pesan antara pukul 16:00 dan 18:00 pada hari yang acak
 - Ternyata (*given*) Staff tersebut telah menerima sedikitnya 10 pesan antara pukul 16:00 dan 18:00, tentukan probabilitas dia menerima 13 pesan pada waktu tersebut
 - Selama 5 hari dalam seminggu, tentukan probabilitas tepat 3 hari ketika Staff tidak menerima pesan antara pukul 17:45 dan 18:00.

BAB 6

Distribusi Normal

Pendahuluan

Kemampuan Akhir yang Diharapkan	Mahasiswa memahami masalah yang berkaitan dengan distribusi normal dan distribusi normal baku (distribusi Z).
Bahan Kajian	- Distribusi normal - Distribusi normal baku (distribusi Z)
Metode Pembelajaran	Demonstrasi <i>software R & Problem Based Learning</i>
Pengalaman Belajar Mahasiswa	- Mengamati fenomena dunia nyata yang mengikuti distribusi normal - Menggunakan software R untuk menghitung probabilitas dan nilai Z - Mengerjakan soal aplikasi distribusi normal dalam konteks nyata
Kriteria Penilaian (Indikator)	- Mampu menjelaskan sifat-sifat distribusi normal dan perbedaannya dengan distribusi lain. - Dapat mengubah data menjadi skor Z dan menginterpretasikan artinya. - Menggunakan tabel Z atau fungsi <code>pnorm()</code> di R untuk menghitung probabilitas. - Menyelesaikan soal terapan yang memerlukan pemodelan data dengan distribusi normal.

Tujuan Pembelajaran

1. Mahasiswa memahami masalah yang berkaitan dengan distribusi normal
2. Mahasiswa memahami masalah yang berkaitan dengan distribusi normal baku (distribusi Z)

A. Distribusi Normal

Distribusi normal merupakan distribusi variabel acak kontinu yang secara luas digunakan dalam berbagai aplikasi. Distribusi normal mensyaratkan variabel yang diukur dalam bentuk data kontinu, seperti tinggi badan, skor IQ, volume air, dan lainnya. Variabel acak normal memiliki distribusi berbentuk lonceng dan terdapat dua parameter yaitu mean μ dan standar deviasi σ .

Fungsi kepadatan probabilitas normal adalah sebagai berikut:

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

di mana μ = rata-rata atau ekspektasi dari distribusi

σ = simpangan baku atau standar deviasi

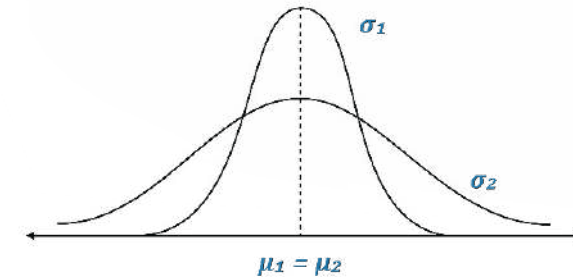
σ^2 = varians

$X \sim N(\mu, \sigma^2)$ dibaca X didistribusikan sebagai distribusi normal dengan mean μ dan standar deviasi σ .

Karakteristik distribusi normal diantaranya yaitu:

1. Distribusi normal memiliki dua parameter yaitu μ dan σ di mana mean menentukan lokasi dan standar deviasi menentukan bentuk distribusi.
2. Kurva normal memiliki bentuk yang simetri dengan nilai tengah/titik tertinggi kurva berada pada nilai mean.
3. Standar deviasi σ menentukan lebarnya kurva. Makin kecil σ , bentuk kurva makin runcing. Sebaliknya, makin besar σ , bentuk kurva lebih

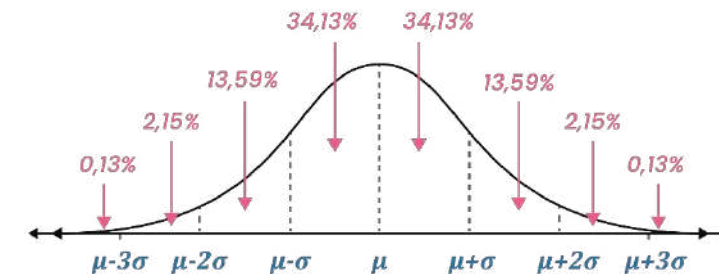
rendah dan lebih menyebar ke samping. Perhatikan Gambar 6.1 berikut di mana $\sigma_2 > \sigma_1$ dengan $\mu_2 = \mu_1$ sehingga kedua kurva normal memiliki nilai tengah yang sama.



Gambar 6.1. Dua Kurva Normal dengan $\mu_2 = \mu_1$ tetapi $\sigma_2 > \sigma_1$

4. Total luas daerah di bawah kurva normal adalah 1.
5. Jarak dari masing-masing nilai X terhadap rata-rata μ diukur dengan standar deviasi σ , maka diperoleh:
 - $\approx 68,26\%$ nilai berada antara $\mu - \sigma$ dan $\mu + \sigma$
 - $\approx 95,44\%$ nilai berada antara $\mu - 2\sigma$ dan $\mu + 2\sigma$
 - $\approx 99,74\%$ nilai berada antara $\mu - 3\sigma$ dan $\mu + 3\sigma$.

Perhatikan Gambar 6.2 berikut.



Gambar 6.2. Persentase Kurva Normal Berdasarkan Jarak μ dan σ

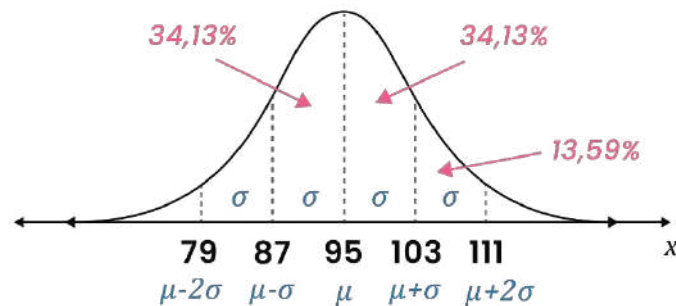
Contoh 6.1

Pengukuran dada pemain sepak bola mengikuti distribusi normal dengan mean 95 cm dan simpangan baku 8 cm.

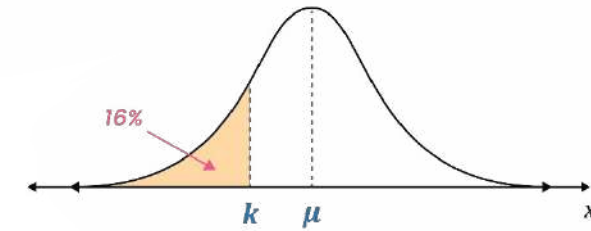
- Tentukan probabilitas dengan ukuran dada pemain sepak bola antara:
 - 87 cm dan 103 cm
 - 103 cm dan 111 cm
- Tentukan nilai k sedemikian hingga aproksimasi probabilitas 16% pada ukuran dada di bawah k cm.

Jawab:

- Diketahui $\mu = 95$ dan $\sigma = 8$, sehingga kurva normal dapat digambarkan sebagai berikut



- Karena $87 = \mu - \sigma$ dan $103 = \mu + \sigma$ maka persentase ukuran dada pemain sepak bola antara 87 cm dan 103 cm yaitu probabilitas antara $\mu - \sigma$ dan $\mu + \sigma$ yaitu $34,13\% + 34,13\% = 68,26\%$.
 - Karena $103 = \mu + \sigma$ dan $111 = \mu + 2\sigma$ maka persentase ukuran dada pemain sepak bola antara 103 cm dan 111 cm yaitu probabilitas antara $\mu + \sigma$ dan $\mu + 2\sigma$ yaitu 13,59%.
- Aproksimasi 16% dari data yaitu berada pada satu standar deviasi dibawah mean atau $\mu - \sigma$, maka $k = 95 - 8 = 87$ cm.



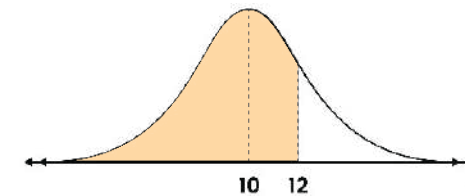
Contoh 6.2

Diketahui $X \sim N(10, 2.3^2)$ maka X terdistribusi normal dengan mean 10 dan simpangan baku 2,3. Tentukan:

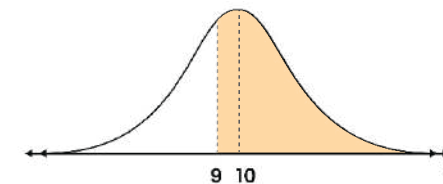
- $P(X \leq 12)$
- $P(X > 9)$
- $P(8 < X < 11)$

Jawab:

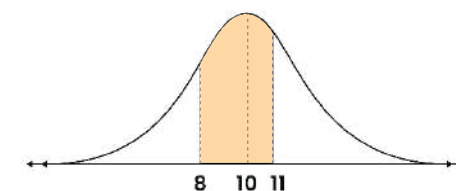
- $P(X \leq 12) \approx 0,808$



- $P(X > 9) \approx 0,668$



- $P(8 < X < 11) \approx 0,476$



```
>
pnorm(12,mean=10,sd=2.3)
#no a
[1] 0.807731
> 1-pnorm(9,10,2.3) #no b
[1] 0.6681399
> pnorm(11,10,2.3)-
pnorm(8,10,2.3) #no c
[1] 0.4758709
```

Catatan: Untuk distribusi kontinu

$$P(X \leq 12) = P(X < 12)$$

$$P(X > 9) = P(X \geq 9)$$

$$P(8 < X < 11) = P(8 \leq X \leq 11)$$

Contoh 6.3

Pada suatu kompetisi, waktu para peserta dalam melengkapi puzzle mengikuti distribusi normal dengan rata-rata 28,3 menit dan standar deviasi 3,6 menit.

- Hitunglah probabilitas peserta menyelesaikan puzzle paling sedikit 30 menit
- Dari 10 peserta yang terpilih secara acak, tentukan probabilitas paling banyak 5 peserta yang menyelesaikan puzzle paling sedikit 30 menit.

Jawab:

- Misal X adalah waktu para peserta dalam menyelesaikan puzzle sehingga dapat dituliskan sebagai $X \sim N(28.3, 3.6^2)$
 $\therefore P(X \geq 30) \approx 0.31838$
- Misal Y adalah banyaknya peserta yang menyelesaikan puzzle paling sedikit 30 menit. Sehingga probabilitas mengikuti **distribusi binomial** dan dapat dituliskan sebagai $Y \sim B(10, 0.31838)$
 $\therefore P(Y \leq 5) \approx 0.938$

```
> 1-pnorm(30,28.3,3.6) #no a  
[1] 0.31838  
> pbinom(5,10,0.31838) #no b  
[1] 0.937738
```



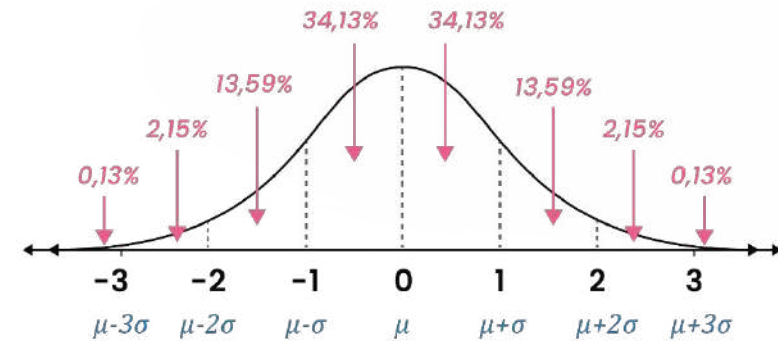
B. Distribusi Normal Baku (Distribusi-Z)

Setiap distribusi X normal dapat diubah menjadi distribusi normal standar atau distribusi-Z dengan menggunakan transformasi:

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

Distribusi-Z memiliki mean 0 dan standar deviasi 1, sehingga $Z \sim N(0, 1^2)$

Kurva pada Gambar 6.3 menunjukkan bagaimana z -score terkait dengan kurva normal secara umum:



Gambar 6.3. Kurva Distribusi-Z

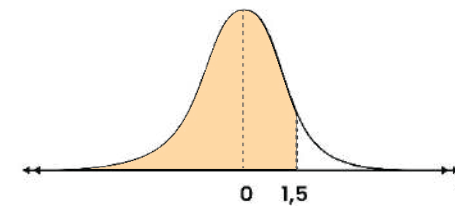
Contoh 6.4

Tentukan probabilitas berikut dan sketsalah kurva normalnya:

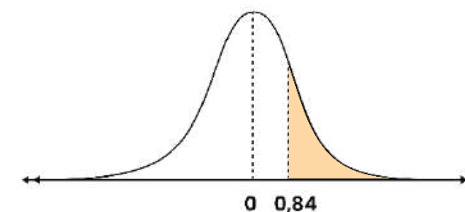
- $P(Z \leq 1.5)$
- $P(Z > 0.84)$
- $P(-0.41 < Z < 0.67)$

Jawab:

- $P(Z \leq 1.5) \approx 0.933$

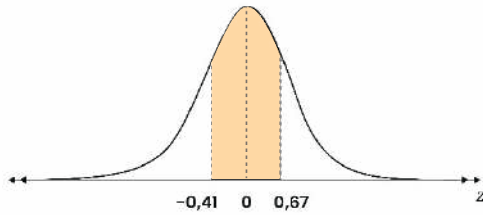


- $P(Z > 0.84) \approx 0.200$



```
> pnorm(1.5) #no a  
[1] 0.9331928  
> 1-pnorm(0.84) #no b  
[1] 0.2004542  
> pnorm(0.67)-pnorm(-  
0.41) #no c  
[1] 0.4076681
```

e. $P(-0,41 < Z < 0,67) \approx 0,408$



C. Kuantil (*Quantiles*) Pada Distribusi Normal

Kuantil dalam distribusi normal adalah nilai pada sumbu horizontal (sumbu x) yang membagi distribusi menjadi bagian-bagian dengan probabilitas tertentu. Kuantil dapat dihitung menggunakan fungsi kuantil yang merupakan **kebalikan** (*invers*) dari fungsi distribusi normal (*inverse normal function*). Kuantil sangat penting dalam analisis statistik, terutama dalam pendugaan parameter, kontrol kualitas dan pengujian hipotesis. Perhatikan contoh berikut ini.

Contoh 6.5

Dalam sebuah proses produksi, suhu kerja mesin mengikuti distribusi normal dengan rata-rata 75°C dan simpangan baku 5°C . Berapa suhu maksimum yang berada dalam 95% dari suhu distribusi normal tersebut?

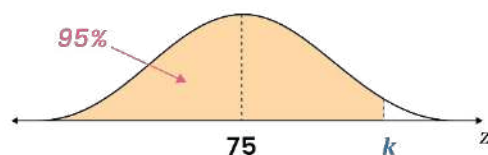
Jawab:

Untuk menjawab pertanyaan tersebut, kita perlu menentukan nilai k sedemikian hingga:

$$P(X \leq k) = 0.95$$

Nilai k dikenal sebagai **kuantil** & pada kasus ini 95% dari kuantil.

Dengan menggunakan fungsi “`qnorm(p, mean, sd)`” pada R, kita akan memperoleh nilai kuantil dari distribusi normal.



```
> qnorm(0.95, mean=75, sd=5)
[1] 83.22427
```



Jadi, 95% dari suhu kerja mesin berada di bawah $83,22^{\circ}\text{C}$

Contoh 6.6

- Diketahui $X \sim N(23.6, 3.1^2)$, tentukan nilai k sehingga $P(X < k) = 0,95$. Sketsalah kurvanya.
- Jika Z merupakan distribusi normal baku, tentukan nilai k sehingga $P(Z > k) = 0,73$. Sketsalah kurvanya.

Jawab:

- $X \sim N(23.6, 3.1^2)$: X mengikuti distribusi normal dengan mean 23,6 dan simpangan baku 3,1.

Jika $P(X < k) = 0.95$ maka

$$\therefore k \approx 28,7$$

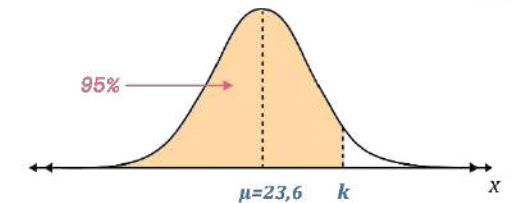
- Z mengikuti distribusi normal dengan mean 0 dan simpangan baku 1.

Jika $P(Z > k) = 0,73$ dan

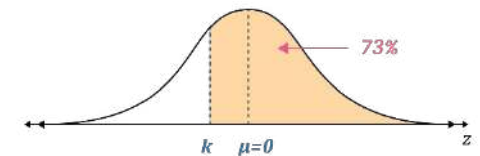
$$P(Z \leq k) = 0,27$$

$$\therefore k$$

$$\approx -0,613$$



b.



```
> qnorm(0.95, mean=23.6, sd=3.1) #no a
[1] 28.53456
> qnorm(0.27, mean=0, sd=1) #no b
[1] -0.612813
```



Contoh 6.7

PT. XYZ memproduksi baut dan diketahui populasi berat baut mengikuti distribusi normal dengan simpangan baku 5,9 gram. Jika probabilitas berat baut kurang dari 58,2 gram yaitu 15%. Tentukan rata-rata berat baut.

Jawab:

Misal X gram dinotasikan sebagai berat baut sehingga $X \sim N(\mu, 5.9^2)$

Karena mean tidak diketahui maka kita tidak bisa menggunakan fungsi invers normal secara langsung, sehingga kita perlu melakukan transformasi ke z -score terlebih dahulu.

$$P(X \leq 58,2) = 0,15$$

$$P\left(Z \leq \frac{58,2 - \mu}{5,9}\right) = \text{inversNorm}(0,15)$$

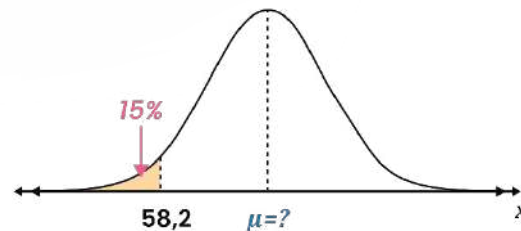
$$\frac{58,2 - \mu}{5,9}$$

$$\approx -1,0364$$

$$58,2 - \mu \approx -6,1$$

$$\therefore \mu \approx 64,3$$

Jadi, rata-rata berat baut adalah 64,3 gram



```
> qnorm(0.15)
[1] -1.036433
```

Contoh 6.8

Tentukan mean dan standar deviasi dari variabel acak X yang berdistribusi normal di mana $P(X \leq 20) = 0,1$ dan $P(X \geq 29) = 0,15$

Jawab:

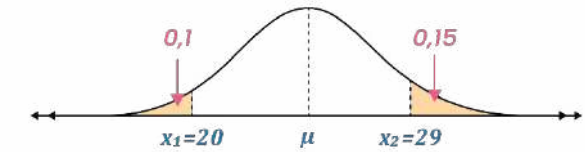
Karena mean dan standar deviasi tidak diketahui maka kita tidak bisa menggunakan fungsi invers normal secara langsung, sehingga kita perlu melakukan transformasi ke z -score terlebih dahulu.

$$P(X \leq 20) = 0,1$$

$$P\left(Z \leq \frac{20 - \mu}{\sigma}\right) = \text{inversNorm}(0,1)$$

$$\frac{20 - \mu}{\sigma} \approx -1,282$$

$$20 - \mu \approx -1,282\sigma \quad \dots (1)$$



$$P(X \geq 29) = 0,15$$

$$P(X < 29) = 0,85$$

$$P\left(Z \leq \frac{29 - \mu}{\sigma}\right) = \text{inversNorm}(0,85)$$

$$\frac{29 - \mu}{\sigma} \approx 1,036$$

$$29 - \mu \approx 1,036\sigma \quad \dots (2)$$



```
> qnorm(0.1)
[1] -1.281552
> qnorm(0.85)
[1] 1.036433
```

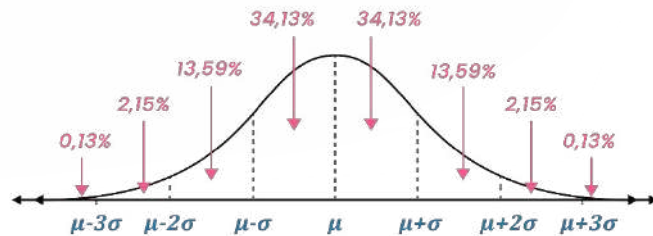
Selanjutnya lakukan metode substitusi atau eliminasi pada persamaan (1) dan (2)

Diperoleh $\mu \approx 25$ dan $\sigma \approx 3,88$

LATIHAN 6

Distribusi Normal

1. Diberikan kurva normal sebagai berikut:



Jawablah pertanyaan No 1 tanpa menggunakan bantuan software statistik & kalkulator

- Tinggi badan pria mahasiswa USAKTI terdistribusikan sebagai distribusi normal dengan mean 170 cm dan simpangan baku 8 cm. Tentukan probabilitas mahasiswa pria:
 - kurang dari 154 cm
 - lebih dari 162 cm
 - antara 170 cm dan 186 cm
 - Tentukan nilai k sedemikian hingga mahasiswa yang memiliki tinggi kurang dari k cm memiliki probabilitas 84,13%
 - Mesin pengisian botol mengisi rata-rata 20.000 botol sehari dengan standar deviasi 2000. Dengan asumsi produksi adalah berdistribusi normal dan satu tahun terdiri dari 260 hari kerja, hitunglah probabilitas jumlah hari kerja di mana:
 - kurang dari 18.000 botol terisi
 - lebih dari 16.000 botol terisi
 - antara 18.000 dan 24.000 botol terisi
2. X adalah variabel acak berdistribusi normal dengan $X \sim N(60, 5^2)$. Sketsalah daerah dan tentukan peluang dari:

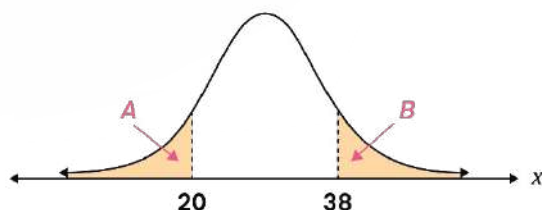
a. $P(X \leq 61)$ b. $P(X > 64)$ c. $P(60 < X < 65)$

3. Jika Z adalah distribusi normal baku. Sketsalah daerah dan tentukan peluang dari:

d. $P(Z \leq -0,53)$ e. $P(Z \geq 4)$ f. $P(-0,86 \leq Z \leq 0,32)$

4. Sebuah mesin menghasilkan baut logam. Panjangnya baut tersebut berdistribusi normal dengan mean 19,8 cm dan simpangan baku 0,3 cm. Jika sebuah baut dipilih secara acak dari mesin, tentukan peluang panjangnya antara 19,7 cm dan 20 cm.
5. Diketahui apel dari tanaman petani berdistribusi normal dengan rata-rata 173 gram dan standar deviasi 34 gram. Apel dengan berat kurang dari 130 gram terlalu kecil untuk dijual.
- Tentukan peluang apel dari tanaman tersebut yang terlalu kecil untuk dijual
 - Tentukan peluang bahwa dalam keranjang pemetik yang berisi 20 apel, terdapat paling banyak 5 apel yang terlalu kecil untuk dijual.
6. Diketahui X adalah variabel acak berdistribusi normal dengan $X \sim N(20, 3^2)$. Tentukan nilai k dan ilustrasikan daerahnya.
- $P(X \leq k) = 0,348$
 - $P(X > k) = 0,876$
7. Jika Z adalah distribusi normal baku. Tentukan nilai k dan ilustrasikan daerahnya.
- $P(Z \leq k) = 0,81$
 - $P(X \geq k) = 0,17$
8. Panjang sekrup yang dihasilkan suatu mesin berdistribusi normal dengan rata-rata 75mm dan standar deviasi 0,1 mm. Sekrup yang terlalu panjang maka secara otomatis akan ditolak. Jika 1% sekrup ditolak, berapa panjang sekrup terkecil yang akan ditolak?

9. Hitunglah nilai rata-rata dan simpangan baku dari suatu distribusi normal variabel acak X , jika diketahui $P(X \geq 80) = 0,1$ dan $P(X \leq 30) = 0,15$
10. Suatu kurva distribusi normal ditampilkan pada gambar berikut. Jika $X \sim N(\mu, \sigma^2)$ di mana luas daerah A = luas daerah B = 0,2.
- Tentukan nilai μ dan σ
 - Hitunglah:
 - $P(X \leq 35)$
 - $P(23 \leq X \leq 30)$



BAB 7

Pendugaan Interval

Pendahuluan

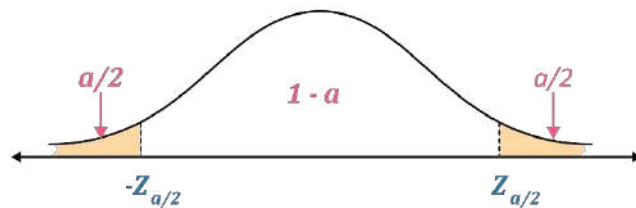
Kemampuan Akhir yang Diharapkan	Mahasiswa memahami masalah yang berkaitan dengan pendugaan interval tentang rata-rata, proporsi, serta perbedaan dua rata-rata dan proporsi
Bahan Kajian	- Pendugaan interval rata-rata - Pendugaan interval proporsi - Pendugaan interval selisih dua rata-rata dan proporsi
Metode Pembelajaran	Ceramah interaktif, studi kasus
Pengalaman Belajar Mahasiswa	- Menganalisis studi kasus nyata terkait estimasi rata-rata dan proporsi - Menghitung interval kepercayaan menggunakan rumus dan software R - Melakukan simulasi dan eksperimen data untuk mengeksplorasi pengaruh ukuran sampel dan tingkat kepercayaan terhadap lebar interval
Kriteria Penilaian (Indikator)	- Menghitung interval kepercayaan untuk rata-rata dan proporsi dengan menggunakan rumus dan software R. - Menafsirkan makna interval kepercayaan dalam konteks permasalahan dengan jelas. - Menunjukkan pengaruh ukuran sampel dan tingkat kepercayaan terhadap hasil estimasi.

Tujuan Pembelajaran

1. Mahasiswa memahami masalah yang berkaitan dengan pendugaan interval rata-rata
2. Mahasiswa memahami masalah yang berkaitan dengan pendugaan interval proporsi
3. Mahasiswa memahami masalah yang berkaitan dengan pendugaan interval perbedaan dua rata-rata dan proporsi

A. Pendugaan Interval Rata-Rata μ

Dalam membuat pendugaan interval, perlu ditentukan terlebih dahulu besarnya tingkat keyakinan yang diberi simbol $1 - \alpha$ di mana α adalah tingkat signifikansi. Besarnya tingkat keyakinan misal 90% maka $\alpha = 10\%$, 95% maka $\alpha = 5\%$, atau angka lainnya. Perhatikan Tingkat signifikansi pada Gambar 7.1.



Gambar 7.1 Tingkat Signifikansi

Terdapat 3 rumus pendugaan interval rata-rata μ :

Rumus pertama,

$$\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

Rumus ini berlaku untuk varians populasi (σ^2) diketahui dengan sampel besar ($n \geq 30$) dari populasi yang tak terbatas atau populasi terbatas dengan penarikan sampel dilakukan dengan pengembalian (*with replacement*).

- $\bar{X}$ = rata-rata sampel
- σ = standar deviasi populasi
- n = ukuran sampel
- $Z_{\frac{\alpha}{2}}$ = skor Z untuk tingkat kepercayaan tertentu

Rumus kedua,

$$\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} < \mu < \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

Rumus ini digunakan untuk populasi terbatas dengan sampel n dari populasi N dan σ diketahui.

- $\bar{X}$ = rata-rata sampel
- σ = standar deviasi populasi
- n = ukuran sampel
- N = ukuran populasi

Rumus ketiga,

$$\bar{X} - t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} < \mu < \bar{X} + t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

Rumus ini berlaku untuk varians populasi tidak diketahui dengan sampel kecil ($n < 30$).

- $\bar{X}$ = rata-rata sampel
- s = standar deviasi sampel (sebagai estimasi σ)
- n = ukuran sampel
- $t_{\frac{\alpha}{2}}$ = nilai t dari distribusi t-student dengan $df = n - 1$

Contoh 7.1

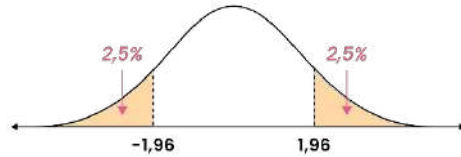
Terdapat seratus orang calon mahasiswa sebagai sampel acak yang sudah mengikuti tes IQ. Diketahui dari seratus orang tersebut memiliki rata-rata IQ 110 dan standar deviasi 20. Buatlah pendugaan interval dari rata-rata IQ dengan menggunakan tingkat keyakinan sebesar 95%

Jawab:

Diketahui $\bar{X} = 110$, $\sigma = 20$, $n = 100$ (karena $n \geq 30$ maka menggunakan distribusi normal)

Tingkat keyakinan 95% dengan $\alpha = 5\%$ maka $\frac{\alpha}{2} = 2,5\%$ sehingga nilai $Z_{\frac{\alpha}{2}} = \pm 1,96$

```
> qnorm(0.025)
[1] -1.959964
> qnorm(0.975)
[1] 1.959964
```



$$\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$
$$110 - 1,96 \left(\frac{20}{\sqrt{100}} \right) < \mu < 110 + 1,96 \left(\frac{20}{\sqrt{100}} \right)$$
$$106,08 < \mu < 113,92$$

$\therefore$ Rata-rata sebenarnya IQ mahasiswa diduga pada interval 106,08 dan 113,92 dengan tingkat keyakinan 95%.

Contoh 7.2

Lima orang mahasiswa FTI USAKTI dipilih secara acak untuk kemudian diukur tingginya. Berikut merupakan data kelima mahasiswa tersebut:

160, 170, 165, 175, 180

Buatlah pendugaan interval tentang rata-rata tinggi mahasiswa FTI USAKTI dengan tingkat kepercayaan sebesar 95%.

Jawab:

Diketahui $n = 5$ (karena $n < 30$ maka menggunakan distribusi t) jadi $df = n - 1 = 4$

Pertama, kita mencari nilai rata-rata dan standar deviasi. Selanjutnya kita mencari nilai $t_{\frac{\alpha}{2}}$.

Tingkat keyakinan 95% dengan $\alpha = 5\%$ maka $\frac{\alpha}{2} = 2,5\%$ sehingga nilai $t_{\frac{\alpha}{2}} = \pm 2,78$

```
> data <- c(160,170,165,175,180)
> mean(data)
[1] 170
> sd(data)
[1] 7.905694
> qt(0.025, df=4)
[1] -2.776445
```



$$\bar{X} - t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} < \mu < \bar{X} + t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$
$$170 - 2,78 \left(\frac{7,9057}{\sqrt{5}} \right) < \mu < 170 + 2,78 \left(\frac{7,9057}{\sqrt{5}} \right)$$
$$160,2 < \mu < 179,8$$

$\therefore$ Rata-rata tinggi mahasiswa FTI USAKTI diduga pada interval 160,2 dan 179,8 dengan tingkat keyakinan 95%.

B. Pendugaan Interval Proporsi (Persentase) $\hat{p}$

Pada pendugaan proporsi dilakukan untuk memperkirakan rentang nilai yang kemungkinan besar mencakup nilai proporsi populasi yang sebenarnya. Interval ini dihitung berdasarkan data sampel dan mempertimbangkan tingkat kepercayaan tertentu, seperti 90%, 95%, atau 99%. Misalnya, dalam penelitian pendapat umum untuk mengetahui berapa % barang produksi yang rusak, berapa % proyek yang tidak selesai tepat

pada waktunya, berapa % yang suka produk tertentu, dan lain sebagainya. Berikut ini merupakan rumus bagi pendugaan interval untuk proporsi $\hat{p}$.

$$\frac{X}{n} - Z_{\alpha/2} \sqrt{\frac{\frac{X}{n} \left(1 - \frac{X}{n}\right)}{n}} < \hat{p} < \frac{X}{n} + Z_{\alpha/2} \sqrt{\frac{\frac{X}{n} \left(1 - \frac{X}{n}\right)}{n}}$$

- X = banyaknya elemen dengan karakteristik tertentu
- n = ukuran sampel

Contoh 7.3

Seorang manajer perusahaan manufaktur akan memperkirakan berapa persen produksinya yang tidak sesuai standar. Dilakukan penelitian dan diambil sampel 250 produk dipilih secara acak. Ternyata ada 60 produk yang tidak sesuai standar. Dengan Tingkat keyakinan 95%, buatlah pendugaan interval persentase produk yang tidak sesuai standar.

Jawab:

Diketahui $n = 250$, $X = 60$, dan $Z_{\frac{\alpha}{2}} = \pm 1,96$

$$\frac{X}{n} - Z_{\alpha/2} \sqrt{\frac{\frac{X}{n} \left(1 - \frac{X}{n}\right)}{n}} < \hat{p} < \frac{X}{n} + Z_{\alpha/2} \sqrt{\frac{\frac{X}{n} \left(1 - \frac{X}{n}\right)}{n}}$$

$$\frac{60}{250} - 1,96 \sqrt{\frac{\frac{60}{250} \left(1 - \frac{60}{250}\right)}{250}} < \hat{p} < \frac{60}{250} + 1,96 \sqrt{\frac{\frac{60}{250} \left(1 - \frac{60}{250}\right)}{250}}$$

$$0,19 < \hat{p} < 0,29$$

$$19\% < \hat{p} < 29\%$$

$\therefore$ Interval antara 19% dan 29% akan memuat persentase produk yang tidak sesuai standar sebenarnya dengan probabilitas 95%. Jika $N = 1000$ maka banyaknya produk yang tidak sesuai standar antara 190 buah sampai 290 buah dengan tingkat keyakinan 95%.

C. Pendugaan Interval Selisih Dua Rata-Rata $\mu_1 - \mu_2$

Pendugaan interval selisih dua rata-rata digunakan untuk memperkirakan rentang nilai yang mungkin memuat selisih antara dua rata-rata populasi berdasarkan data sampel. Pendekatan ini berguna dalam membandingkan dua kelompok, misalnya perbedaan efektivitas dua metode produksi atau perbedaan rata-rata pendapatan antara dua wilayah. Terdapat dua kasus utama berdasarkan varians populasi (σ^2) yang diketahui atau tidak diketahui.

Rumus pertama,

$$(\bar{X}_1 - \bar{X}_2) - Z_{\alpha/2} \sigma_{(\bar{X}_1 - \bar{X}_2)} < (\mu_1 - \mu_2) < (\bar{X}_1 - \bar{X}_2) + Z_{\alpha/2} \sigma_{(\bar{X}_1 - \bar{X}_2)}$$

$$\text{di mana } n \geq 30 \text{ dan } \sigma_{(\bar{X}_1 - \bar{X}_2)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Rumus ini berlaku untuk varians populasi σ_1^2 dan σ_2^2 diketahui.

- $\bar{X}_1, \bar{X}_2$ = rata-rata sampel dari populasi 1 dan 2
- σ_1^2, σ_2^2 = varians populasi 1 dan 2
- n_1, n_2 = ukuran sampel dari populasi 1 dan 2

Rumus kedua,

$$(\bar{X}_1 - \bar{X}_2) - t_{\alpha/2} s_{(\bar{X}_1 - \bar{X}_2)} < (\mu_1 - \mu_2) < (\bar{X}_1 - \bar{X}_2) + t_{\alpha/2} s_{(\bar{X}_1 - \bar{X}_2)}$$

$$\text{di mana } n < 30 \text{ dan } s_{(\bar{X}_1 - \bar{X}_2)} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Rumus ini berlaku untuk varians populasi σ_1^2 dan σ_2^2 tidak diketahui (menggunakan estimasi dari sampel)

- $\bar{X}_1, \bar{X}_2$ = rata-rata sampel dari populasi 1 dan 2

- s_1^2, s_2^2 = varians sampel 1 dan 2
- $t_{\alpha/2}$ = nilai t dari distribusi t-student dengan $df = n - 1$

Contoh 7.4

Seorang analyst pada perusahaan sedang melakukan pengamatan terhadap suatu jenis bola lampu dengan merk yang berbeda yaitu merk A dan merk B. Analyst ingin mengetahui apakah ada selisih atau perbedaan rata-rata lamanya hidup (*expected life*) dari kedua lampu. Diselidiki sebanyak 100 buah lampu untuk masing-masing merk dan diperoleh data:

Merk A:

Menyala rata-rata = 3600 jam dan standar deviasi = 200 jam

Merk B:

Menyala rata-rata = 3500 jam dan standar deviasi = 200 jam

Dengan menggunakan tingkat kepercayaan 90%, hitunglah pendugaan interval selisih rata-rata lamanya hidup dari kedua bola lampu tersebut!

Jawab:

Diketahui: $n_1 = 100$, $\bar{X}_1 = 3600$, $\sigma_1 = 200$ maka $\sigma_1^2 = 40.000$

$n_2 = 100$, $\bar{X}_2 = 3500$, $\sigma_2 = 200$ maka $\sigma_2^2 = 40.000$

karena $\alpha = 10\%$ maka $\alpha/2 = 5\%$ sehingga $Z_{\alpha/2} = \pm 1,64$

$$\sigma_{(\bar{X}_1 - \bar{X}_2)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{40.000}{100} + \frac{40.000}{100}} \approx 28,28$$

Dengan menggunakan rumus pertama:

$$(\bar{X}_1 - \bar{X}_2) - Z_{\alpha/2} \sigma_{(\bar{X}_1 - \bar{X}_2)} < (\mu_1 - \mu_2) < (\bar{X}_1 - \bar{X}_2) + Z_{\alpha/2} \sigma_{(\bar{X}_1 - \bar{X}_2)}$$

$$(3600 - 3500) - 1,64(28,28) < (\mu_1 - \mu_2) < (3600 - 3500) + 1,64(28,28)$$

$$53,62 < (\mu_1 - \mu_2) < 146,38$$

$\therefore$ Interval antara 53,62 jam dan 146,38 jam akan memuat selisih rata-rata lamanya hidup dari bola lampu kedua merk dengan tingkat kepercayaan 90%.

D. Pendugaan Interval Selisih Dua Proporsi $P_1 - P_2$

Rumus dari selisih dua proporsi atau persentase adalah sebagai berikut.

$$(\hat{p}_1 - \hat{p}_2) - Z_{\alpha/2} S_{(\hat{p}_1 - \hat{p}_2)} < (P_1 - P_2) < (\hat{p}_1 - \hat{p}_2) + Z_{\alpha/2} S_{(\hat{p}_1 - \hat{p}_2)}$$

$$\text{di mana } n \geq 30 \text{ dan } S_{(\hat{p}_1 - \hat{p}_2)} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$\hat{p}_1$ dan $\hat{p}_2$ dibaca P_1 dan P_2 topi atau “cap”, artinya $\hat{p}_1$ dan $\hat{p}_2$ merupakan perkiraan P_1 dan P_2

Contoh 7.5

Pabrik keramik PT. XYZ melakukan penelitian pada dua buah mesin untuk mengetahui apakah ada perbedaan antara persentase keramik yang rusak pada mesin tersebut. Pada mesin A dan B masing-masing diambil sampel 120 produk. Ternyata dari mesin A terdapat 90 produk tidak rusak & dari mesin B terdapat 78 produk tidak rusak. Buatlah pendugaan interval dari selisih/perbedaan persentase tentang produk yang tidak rusak pada kedua mesin dengan tingkat keyakinan 90%.

Jawab:

Diketahui $\hat{p}_1 = \frac{X_1}{n_1} = \frac{90}{120} = 0,75$ dan $\hat{p}_2 = \frac{X_2}{n_2} = \frac{78}{120} = 0,65$

$$\hat{p}_1 - \hat{p}_2 = 0,75 - 0,65 = 0,1$$

$$S_{(\hat{p}_1 - \hat{p}_2)} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = \sqrt{\frac{(0,75)(0,25)}{120} + \frac{(0,65)(0,35)}{120}}$$

$$\approx 0,059$$

karena $\alpha = 10\%$ maka $\alpha/2 = 5\%$ sehingga $Z_{\alpha/2} = \pm 1,64$

Dengan menggunakan selisih dua proporsi:

$$(\hat{p}_1 - \hat{p}_2) - Z_{\alpha/2} s_{(\hat{p}_1 - \hat{p}_2)} < (P_1 - P_2) < (\hat{p}_1 - \hat{p}_2) + Z_{\alpha/2} s_{(\hat{p}_1 - \hat{p}_2)}$$

$$0,1 - 1,64(0,059) < (P_1 - P_2) < 0,1 + 1,64(0,059)$$

$$0,003 < (P_1 - P_2) < 0,197$$

$\therefore$ Interval antara 0,3% dan 19,7% akan memuat perbedaan persentase mengenai produk yang tidak rusak dari kedua mesin tersebut dengan probabilitas 90%.

LATIHAN 7

Pendugaan Interval

1. Rata-rata nilai pesanan (*order*) yang diterima oleh suatu Perusahaan sebanyak 36 kali bernilai Rp12,74 juta dengan simpangan baku Rp 2,45 juta. Buatlah pendugaan interval rata-rata pesanan dengan tingkat keyakinan 95%.
2. Suatu kegiatan pengawasan mutu barang (*quality control*) ingin mengestimasi proporsi mesin yang rusak. Untuk maksud tersebut dipilih 120 mesin secara acak, ternyata ada 18 mesin yang rusak. Buatlah pendugaan interval persentase mesin yang rusak dengan tingkat keyakinan 90% & 97%. Bandingkan hasilnya dan berikan komentar!
3. Sebuah pabrik mobil ingin memutuskan apakah akan menggunakan ban merk A atau B bagi mobil produksi terbarunya. Untuk membantu mencapai keputusan tersebut, sebuah percobaan dilakukan dengan menggunakan 200 ban untuk masing-masing merk A dan merk B. Ban-ban tersebut dipasang dan digunakan sampai aus sehingga harus diganti. Hasilnya adalah:
 - Merk A: rata-rata 41.000 km dan simpangan baku 5000 km
 - Merk B: rata-rata 42.500 km dan simpangan baku 5200 km

Dengan menggunakan tingkat keyakinan 95%

- a. Gambarkan kurva normalnya
 - b. Hitunglah pendugaan interval selisih rata-rata jarak tempuh yang bisa dicapai kedua merk ban tersebut & simpulkan hasil yang diperoleh.
4. Tim penelitian teknik mesin sedang menguji daya tahan sebuah jenis bantalan logam yang digunakan dalam mesin industri. Diketahui umur pakai (dalam jam) dari 10 sampel bantalan yang diambil secara acak. Hasil pengukuran menunjukkan bahwa rata-rata umur pakai bantalan adalah 1200 jam dengan simpangan baku sampel 100 jam. Dengan

tingkat kepercayaan 95%, tentukan interval kepercayaan untuk rata-rata umur pakai bantalan logam tersebut!

5. Suatu Perusahaan teknik menguji dua jenis bahan pelapis tahan panas untuk komponen mesin. Pengujian dilakukan dengan melihat proporsi komponen yang mengalami kerusakan setelah digunakan selama 1000 jam. Hasilnya adalah:

- Dari sampel 200 komponen dengan bahan pelapis A, ditemukan bahwa 30 komponen mengalami kerusakan.
- Dari sampel 250 komponen dengan bahan pelapis B, ditemukan bahwa 40 komponen mengalami kerusakan.

Dengan tingkat kepercayaan 95%, tentukan interval kepercayaan untuk selisih proporsi komponen yang mengalami kerusakan antara bahan pelapis A dan bahan pelapis B.

6. Tim insinyur dari PT. XYZ sedang menguji kualitas pengelasan pada sambungan pipa di proyek konstruksi. Dari 20 sambungan pipa yang diperiksa secara acak, ditemukan bahwa 3 sambungan memiliki cacat yang dapat mempengaruhi kekuatan struktur. Dengan tingkat kepercayaan 90%, tentukan interval kepercayaan untuk proporsi sebenarnya dari sambungan pipa yang memiliki cacat dalam proyek tersebut.
7. Tim Teknik Mesin USAKTI ingin membandingkan efisiensi dua jenis oli pelumas pada suhu tinggi dengan mengukur rata-rata keausan komponen mesin (dalam mikrometer) setelah 100 jam operasi.

- Dari sampel 12 mesin yang menggunakan Oli A, diperoleh rata-rata keausan 15,2 mikrometer dengan simpangan baku 1,8 mikrometer.
- Dari sampel 10 mesin yang menggunakan Oli B, diperoleh rata-rata keausan 17,5 mikrometer dengan simpangan baku 2,1 mikrometer.

Dengan tingkat kepercayaan 95%, tentukan interval kepercayaan untuk selisih rata-rata keausan antara mesin yang menggunakan Oli A dan Oli B.

BAB 8

Pengujian Hipotesis

Pendahuluan

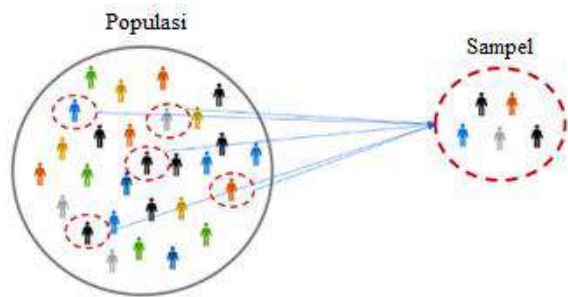
Kemampuan Akhir yang Diharapkan	Mahasiswa mampu menjelaskan sampel, populasi, hipotesis, jenis kesalahan, serta menyelesaikan masalah pengujian hipotesis rata-rata, proporsi dan varians.
Bahan Kajian	- Sampel, populasi, jenis kesalahan, dan hipotesis - Uji hipotesis rata-rata - Uji hipotesis proporsi - Uji hipotesis varians
Metode Pembelajaran	<i>Problem based learning</i> , demonstrasi <i>software R</i>
Pengalaman Belajar Mahasiswa	- Menganalisis kasus nyata untuk membedakan antara populasi dan sampel - Menyusun hipotesis nol dan alternatif dari pernyataan masalah statistik - Melakukan uji hipotesis untuk rata-rata, proporsi, dan varians menggunakan R - Menginterpretasikan hasil pengujian hipotesis secara statistik dan praktis
Kriteria Penilaian (Indikator)	- Mampu mengidentifikasi dengan jelas populasi dan sampel dalam konteks masalah. - Melakukan uji hipotesis (rata-rata, proporsi, varians) secara benar dan logis. - Menggunakan R untuk menguji hipotesis dan membaca hasil output secara akurat. - Menjelaskan proses uji statistik dari awal hingga kesimpulan dengan runtut dan tepat.

Tujuan Pembelajaran

- 1. Mahasiswa mampu memahami dan menjelaskan sampel, populasi, dan hipotesis
- 2. Mahasiswa memahami masalah yang berkaitan dengan pengujian hipotesis rata-rata
- 3. Mahasiswa memahami masalah yang berkaitan dengan pengujian hipotesis proporsi
- 4. Mahasiswa memahami masalah yang berkaitan dengan pengujian hipotesis varians

A. Sampel dan Populasi

Populasi adalah keseluruhan elemen (himpunan) dengan ciri-ciri tertentu yang sama. Artinya populasi adalah himpunan semua individu yang menggambarkan suatu informasi atau data untuk penelitian. Sampel adalah himpunan bagian dari populasi. Setiap sampel yang diambil haruslah representatif artinya segala karakteristik populasi tercerminkan pada sampel yang diambil tersebut. Sebagai contoh, populasi adalah seluruh mahasiswa Universitas Trisakti, sedangkan sampel yang diambil adalah sekelompok mahasiswa Teknik Mesin Universitas Trisakti.



Gambar 8.1. Populasi dan Sampel

Pengumpulan data bertujuan untuk menarik kesimpulan tentang populasi dari sampel itu diambil. Satuan-satuan ukuran kuantitatif dalam populasi disebut *parameter*, sedangkan satuan-satuan kuantitatif pada

sampel disebut *statistik*. Berikut ini merupakan ukuran dari parameter dan statistik.

Ukuran	Parameter (Populasi)	Statistik (Sampel)
Rata-Rata	μ	$\bar{x}$
Standar Deviasi	σ	s_n
Varians	σ^2	s_n^2

B. Definisi Hipotesis

Hipotesis adalah suatu anggapan yang mungkin benar dan sering digunakan sebagai dasar pembuat keputusan. Namun karena adanya kemungkinan kesalahan, maka perlu diuji terlebih dahulu berdasarkan data hasil observasi. Sebagai contoh:

- Seorang pejabat bank beranggapan bahwa penurunan suku bunga deposito tidak mempengaruhi jumlah tabungan deposito, maka diputuskan untuk menurunkan suku bunga deposito.
- Peningkatan rasio kompresi pada mesin pembakaran akan meningkatkan efisiensi termal hingga batas tertentu sebelum terjadi knocking yang mengurangi performa mesin, maka perlu diuji hipotesisnya.
- Penggunaan material komposit berbasis serat karbon pada komponen kendaraan akan meningkatkan rasio kekuatan terhadap berat dibandingkan dengan material baja konvensional, maka perlu diuji hipotesisnya.

Hipotesis statistik merupakan suatu pernyataan atau dugaan yang mungkin benar atau mungkin salah tentang parameter dari satu atau lebih populasi yang bisa diuji secara empiris berdasarkan data. Ingat, yang diuji dalam hipotesis adalah **parameter**, maka notasi dalam hipotesis statistika adalah parameter μ (untuk rata-rata), σ^2 (untuk varians), **p** (untuk proporsi),

B (untuk koefisien regresi), ρ (untuk koefisien korelasi), dan lain sebagainya.

Pengujian hipotesis statistik adalah prosedur yang memungkinkan keputusan untuk ditolak atau tidak menolak hipotesis yang sedang diuji. Dalam “menerima” atau “menolak” suatu hipotesis, ada hal yang perlu diperhatikan bahwa penolakan suatu hipotesis berarti menyimpulkan bahwa hipotesis itu salah, sedangkan menerima suatu hipotesis mengimplikasikan bahwa tidak memiliki bukti untuk mempercayai sebaliknya. Karena pengertian ini, peneliti cenderung mengambil pernyataan yang diharapkan akan ditolak sebagai hipotesisnya.

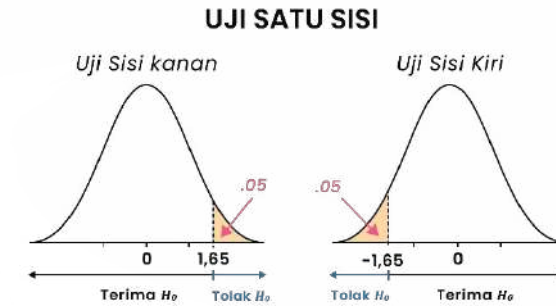
Jenis hipotesis statistik terdiri atas:

- Hipotesis nol H_0 : suatu pernyataan atau dugaan yang akan dibuktikan
- Hipotesis alternatif H_1 : hipotesis yang bertentangan dengan H_0

Terdapat 2 tipe hipotesis:

1. Hipotesis Satu Arah/ Hipotesis Satu Sisi

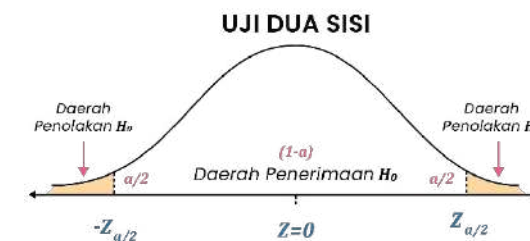
Hipotesis satu arah adalah hipotesis yang menyatakan adanya hubungan antara variabel dengan arah yang sudah ditentukan, baik itu **meningkat** atau **menurun**. Misalnya, metode pembelajaran PBL akan **meningkatkan** hasil belajar siswa dibandingkan metode ceramah; penggunaan pelumas sintetis akan **mengurangi** gesekan lebih efektif dibandingkan pelumas konvensional, dan lainnya. Ciri khas hipotesis satu arah adalah adanya kata-kata seperti *meningkatkan*, *menurunkan*, *mempercepat*, *memperlambat*, dan lain-lain. Hipotesis alternatif pada hipotesis satu arah ditunjukkan dengan tanda $>$ (*lebih dari*) atau $<$ (*kurang dari*).



Gambar 8.2. Ilustrasi Uji Satu Sisi

2. Hipotesis Dua Arah/ Hipotesis Dua Sisi

Hipotesis dua arah adalah hipotesis yang menyatakan adanya hubungan atau perbedaan antara variabel **tanpa menentukan arah** pengaruhnya bisa meningkat atau menurun. Misalnya, kenaikan harga bahan bakar **berpengaruh** terhadap jumlah penggunaan kendaraan pribadi; penggunaan pelumas sintetis **berpengaruh** terhadap tingkat gesekan pada mesin, dan lainnya. Ciri khas hipotesis dua arah adalah tidak menyebutkan apakah pengaruhnya positif atau negatif, hanya menyatakan bahwa ada hubungan atau pengaruh antara variabel. Hipotesis alternatifnya menunjukkan tanda $\neq$ yang artinya peneliti ingin melihat apakah ada suatu perbedaan.



Gambar 8.3. Ilustrasi Uji Dua Sisi

C. Jenis Kesalahan

Terdapat dua jenis kesalahan yang mungkin terjadi dalam pengujian hipotesis. Kesalahan yang disebabkan karena kita menolak hipotesis nol padahal hipotesis nol itu benar disebut kesalahan tipe I atau *Type I Error*. Kesalahan tipe I disebut juga sebagai *false positive* di mana probabilitas

terjadinya kesalahan ini disebut taraf signifikansi (α). Sebagai contoh, seorang pasien didiagnosis mengidap kanker (menolak H_0), padahal sebenarnya pasien tersebut sehat.

Sebaliknya, kesalahan yang diakibatkan karena kita menerima hipotesis nol padahal hipotesis itu salah disebut kesalahan tipe II atau *Type II Error*. Kesalahan tipe II disebut sebagai *false negative* dengan probabilitas terjadinya kesalahan ini disebut β . Misalnya, seorang pasien sebenarnya memiliki kanker, tetapi hasil tes menyatakan pasien tersebut sehat (gagal menolak H_0). Perhatikan tabel berikut.

	H_0 benar	H_1 benar
Menerima H_0	Keputusan yang benar $(1 - \alpha)$	Keputusan yang salah Kesalahan Tipe II atau β
Menolak H_0	Keputusan yang salah/ Kesalahan Tipe I atau α	Keputusan yang benar $(1 - \beta)$

- Beberapa hal yang berkaitan dengan uji hipotesis diantaranya:
- Tingkat signifikansi (α): peluang melakukan kesalahan tipe I
 - Tingkat kepercayaan/taraf nyata $(1 - \alpha)$: peluang kepercayaan untuk dapat menolak H_0
 - Tingkat kekuatan uji $(1 - \beta)$: seberapa besar peluang menolak H_0 jika H_0 salah.
 - Menurunkan α mengurangi kemungkinan kesalahan tipe I tetapi meningkatkan kesalahan tipe II
 - Menambah ukuran sampel n dapat mengurangi kesalahan tipe II.
 - Semakin kecil α maka penelitian semakin meyakinkan atau semakin beresiko penelitian maka α semakin kecil

D. Pengujian Hipotesis Tentang Rata-Rata μ

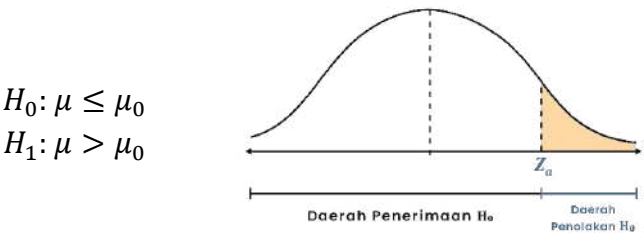
Pengujian hipotesis rata-rata bertujuan untuk mengetahui apakah rata-rata suatu populasi berbeda dari nilai tertentu. Urutan yang perlu diperhatikan dalam pengujian hipotesis adalah sebagai berikut:

1. Menentukan hipotesis nol H_0 dan hipotesis alternatif H_1

Tentukan hipotesis berdasarkan pernyataan penelitian:

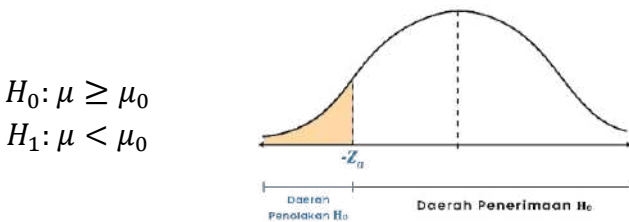
- Uji satu sisi (kanan)

Menguji apakah rata-rata lebih besar dari nilai tertentu



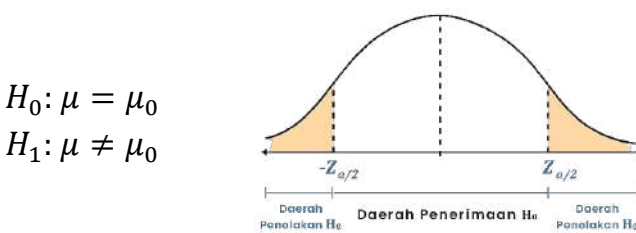
- Uji satu sisi (kiri)

Menguji apakah rata-rata lebih kecil dari nilai tertentu



- Uji dua sisi

Menguji apakah rata-rata berbeda dari nilai tertentu



2. Tentukan tingkat signifikansi α atau taraf nyata

Biasanya $\alpha = 0,05$ (5%) atau $\alpha = 0,01$ (1%) tergantung tingkat kepercayaan yang diinginkan.

3. Tentukan statistik uji yang digunakan

Pemilihan uji bergantung pada ukuran sampel dan apakah varians populasi diketahui atau tidak

- Jika varians populasi diketahui (σ^2) atau $n > 30$ maka gunakan uji-Z dengan statistik uji:

$$Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

- Jika varians populasi tidak diketahui atau $n \leq 30$ maka gunakan uji-t dengan statistik uji:

$$t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

di mana:

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

μ_0 = rata-rata populasi yang diuji

σ = standar deviasi populasi (jika diketahui)

s = standar deviasi sampel (jika σ tidak diketahui)

n = ukuran sampel

4. Tentukan wilayah kritis, nilai kritis, dan kesimpulan

- Cari nilai kritis z atau t berdasarkan α dan jenis uji (satu sisi atau dua sisi) dengan menggunakan Tabel distribusi-Z untuk uji-Z dan Tabel distribusi -t untuk uji -t
- Wilayah kritis adalah daerah di mana H_0 akan ditolak.

- Bandingkan dengan nilai statistik uji dan tarik kesimpulan dengan ketentuan yaitu

Tolak H_0 : jika nilai statistik uji berada di wilayah kritis

Terima H_0 : jika nilai statistik uji berada di luar wilayah kritis

Tolak H_0 : jika nilai p-value $< \alpha$

Terima H_0 : jika nilai p-value $> \alpha$

Sebagian besar penelitian menggunakan p-value sebagai acuan. **P-value** adalah peluang yang digunakan untuk mengukur apakah kita memiliki cukup bukti untuk menolak hipotesis nol.

Contoh 8.1

Sebuah perusahaan manufaktur suku cadang otomotif ingin menguji kualitas baut yang dihasilkan oleh mesin baru. Menurut spesifikasi, panjang rata-rata baut yang ideal adalah 50 mm. Selanjutnya dilakukan penyelidikan dan diambil sampel 100 baut secara acak dan diperoleh rata-rata panjang sampel 49,6 mm dan standar deviasi 1,2 mm. Dengan menggunakan tingkat kepercayaan 95%.

- Ujilah apakah mesin baru menghasilkan baut dengan panjang rata-rata kurang dari 50 mm?
- Ujilah apakah mesin baru menghasilkan baut dengan panjang rata-rata lebih dari 50 mm?
- Ujilah apakah mesin baru menghasilkan baut dengan panjang rata-rata berbeda dari 50 mm?

Jawab:

Langkah awal (sama untuk semua bagian a, b dan c):

Karena $n > 30$ maka akan dicari nilai statistik uji (nilai Z-hitung) dengan menggunakan uji Z:

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{49,6 - 50}{1,2/\sqrt{100}} \approx -3,33$$

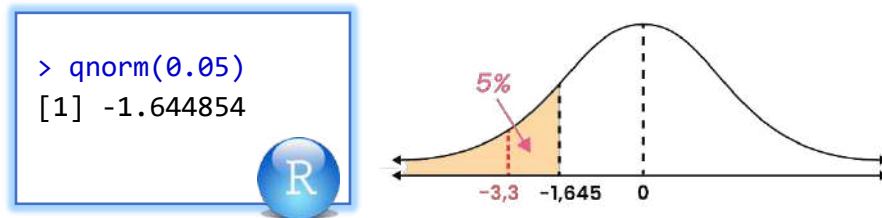
a. Uji sisi kiri:

$H_0: \mu \geq 50$ (panjang sesuai atau lebih besar dari standar)

$H_1: \mu < 50$ (panjang kurang dari standar)

Nilai kritis:

karena tingkat keyakinan 95% dan $\alpha = 0,05$ (5%) maka nilai kritis $Z_\alpha = -1,645$ (nilai Z-tabel)



Berikut adalah penjelasan kurva distribusi normal untuk pengujian sisi kiri:

- Area biru di kiri adalah daerah penolakan dengan $\alpha = 5\%$ yang merupakan wilayah kritis
- Garis merah menunjukkan statistik uji yaitu z-hitung -3,33
- Nilai uji statistik berada pada wilayah kritis $\rightarrow$ Tolak H_0

Kesimpulan:

Diperoleh hasil z-hitung -3,33 dan z-tabel -1,645 dengan uji sisi kiri. Karena nilai z-hitung yaitu -3,33 berada pada wilayah kritis maka **Tolak H_0** . Dengan demikian dapat disimpulkan terdapat bukti yang cukup untuk menyatakan bahwa **mesin baru menghasilkan baut dengan panjang rata-rata kurang dari 50 mm.**

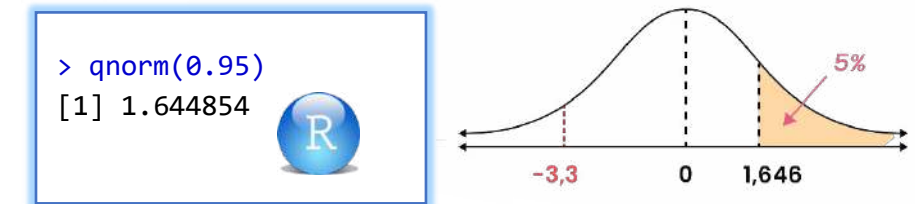
b. Uji sisi kanan:

$H_0: \mu \leq 50$ (panjang sesuai atau lebih kecil dari standar)

$H_1: \mu > 50$ (panjang lebih dari standar)

Nilai kritis:

karena tingkat keyakinan 95% dan $\alpha = 0,05$ (5%) maka nilai kritis $Z_\alpha = 1,645$ (nilai Z-tabel)



Kesimpulan:

Diperoleh hasil z-hitung -3,33 dan z-tabel 1,645 dengan uji sisi kanan. Karena nilai z-hitung yaitu -3,33 tidak berada pada wilayah kritis maka **gagal tolak H_0** . Dengan demikian, tidak terdapat bukti yang cukup untuk menyatakan bahwa **mesin baru menghasilkan baut dengan panjang lebih dari 50 mm.**

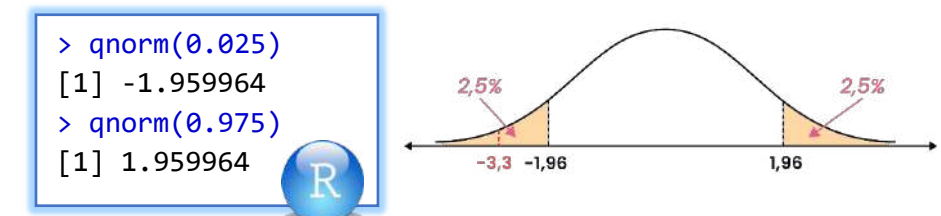
c. Uji dua sisi:

$H_0: \mu = 50$ (panjang sesuai standar)

$H_1: \mu \neq 50$ (panjang berbeda dari standar)

Nilai kritis:

karena tingkat keyakinan 95% dan $\frac{\alpha}{2} = 0,025$ (2,5%) maka nilai kritis $Z_{\frac{\alpha}{2}} = \pm 1,96$ (nilai Z-tabel)



Berikut adalah penjelasan kurva distribusi normal untuk pengujian dua sisi:

- Area biru di kiri dan kanan adalah daerah penolakan dengan $\alpha = 5\%$ dan masing-masing 2,5% per sisi yang merupakan wilayah kritis
- Garis merah menunjukkan statistik uji yaitu z-hitung -3,33
- Nilai uji statistik berada pada wilayah kritis $\rightarrow$ Tolak H_0

Kesimpulan:

Diperoleh hasil z-hitung -3,33 dan z-tabel $\pm 1,96$ dengan uji dua sisi. Karena nilai z-hitung yaitu -3,33 berada pada wilayah kritis maka **Tolak H_0** . Dengan demikian, terdapat bukti yang cukup untuk menyatakan bahwa **mesin baru menghasilkan baut dengan panjang rata-rata berbeda dari standar yaitu 50 mm.**

Contoh 8.2

Sebuah perusahaan ingin memastikan bahwa mesin bubut CNC generasi baru mampu menghasilkan batang logam dengan rata-rata diameter tidak kurang dari 20 mm. Diketahui bahwa standar diameter minimum yang dapat diterima adalah 20 mm. Dalam uji kualitas, perusahaan mengambil sampel acak sebanyak 15 batang logam dan diketahui bahwa rata-rata diameter sampel 19,7 mm dan standar deviasi sampel 1 mm. Dengan menggunakan tingkat kepercayaan 95%, apakah mesin bubut CNC generasi baru menghasilkan batang logam dengan **diameter kurang dari 20 mm?**

Jawab:

Uji sisi kiri:

$H_0: \mu \geq 20$ (diameter sesuai atau lebih besar dari standar)

$H_1: \mu < 20$ (diameter kurang dari standar)

Statistik Uji:

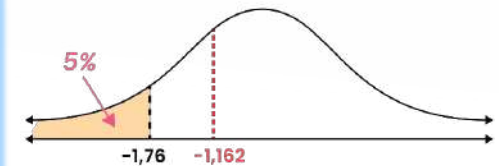
Karena $n < 30$ maka akan dicari nilai statistik uji (nilai t-hitung) dengan menggunakan uji-t:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{19,7 - 20}{1/\sqrt{15}} \approx -1,162$$

Nilai kritis:

karena tingkat keyakinan 95% dan $\alpha = 0,05$ (5%) maka nilai kritis $t_{\alpha} = -1,761$ (nilai t-tabel)

```
> qt(0.05,df=14)
[1] -1.76131
```



Kesimpulan:

Diperoleh hasil t-hitung -1,162 dan t-tabel -1,761 dengan uji sisi kiri. Karena nilai t-hitung yaitu -1,162 tidak berada pada wilayah kritis maka **gagal tolak H_0** . Dengan demikian, **tidak ada cukup bukti** untuk menyatakan bahwa mesin bubut CNC generasi baru menghasilkan batang logam dengan diameter **kurang dari 20 mm.**

E. Pengujian Hipotesis Tentang Proporsi

Pengujian hipotesis proporsi digunakan untuk menguji klaim tentang persentase (proporsi) suatu karakteristik dalam populasi. Misalnya, ingin diketahui:

- Apakah lebih dari 60% produk mesin A layak pakai
- Apakah kurang dari 5% komponen dari Mesin Bubut X mengalami keretakan setelah pengujian beban?
- Apakah proporsi produk cacat berbeda antara mesin produksi A dan mesin produksi B?

Berikut ini merupakan urutan yang perlu diperhatikan dalam pengujian hipotesis proporsi:

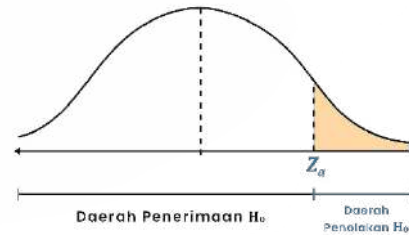
1. Menentukan hipotesis nol H_0 dan hipotesis alternatif H_1

Tentukan hipotesis berdasarkan pernyataan penelitian:

- Uji satu sisi (kanan)

$$H_0: p \leq p_0$$

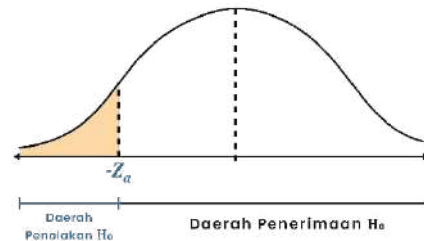
$$H_1: p > p_0$$



- Uji satu sisi (kiri)

$$H_0: p \geq p_0$$

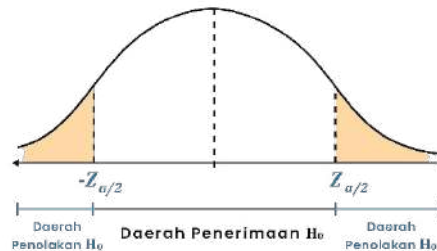
$$H_1: p < p_0$$



- Uji dua sisi

$$H_0: p = p_0$$

$$H_1: p \neq p_0$$



2. Tentukan tingkat signifikansi α atau taraf nyata

Biasanya taraf signifikansi $\alpha = 0,05$ (5%) atau $\alpha = 0,01$ (1%) tergantung tingkat kepercayaan yang diinginkan/diberikan.

3. Tentukan statistik uji yang digunakan

Statistik uji pada proporsi menggunakan uji-z:

$$Z = \frac{\frac{X}{n} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$

di mana:

X = banyaknya elemen sampel dengan karakteristik tertentu

p_0 = proporsi populasi berdasarkan hipotesis

n = ukuran sampel

4. Tentukan nilai kritis dan tarik kesimpulan

- Bandingkan Z-hitung dengan Z kritis dari tabel z normal
- Bandingkan dengan nilai statistik uji dan tarik kesimpulan dengan ketentuan yaitu

Tolak H_0 : jika nilai statistik uji berada di wilayah kritis

Terima H_0 : jika nilai statistik uji berada di luar wilayah kritis

Tolak H_0 : jika nilai p-value $< \alpha$

Terima H_0 : jika nilai p-value $> \alpha$

Contoh 8.3

Sebuah pabrik mesin mengklaim bahwa 90% dari suku cadangnya bebas cacat. Untuk memverifikasi klaim tersebut, seorang inspektur secara acak memeriksa 100 suku cadang, dan menemukan bahwa 85 di antaranya bebas cacat. Dengan tingkat signifikansi 5%, uji apakah klaim pabrik benar?

Jawab:

$$H_0: p = 0,9$$

$$H_1: p \neq 0,9$$

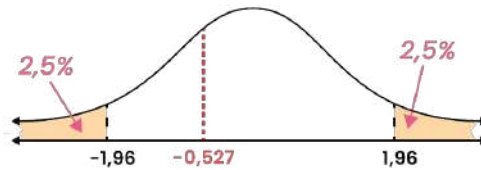
Statistik uji proporsi:

$$Z = \frac{\frac{X}{n} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\frac{85}{100} - 0,9}{\sqrt{\frac{0,9(1-0,9)}{100}}} \approx -0,527$$

Nilai kritis:

karena tingkat signifikansi 5% dan $\frac{\alpha}{2} = 0,025$ (2,5%) maka nilai kritis $Z_{\frac{\alpha}{2}} = \pm 1,96$ (nilai Z-tabel)

```
> qnorm(0.025)
[1] -1.959964
> qnorm(0.975)
[1] 1.959964
```



Kesimpulan:

Diperoleh hasil z-hitung -0,527 dan z-tabel $\pm 1,96$ dengan uji dua sisi. Karena nilai z-hitung yaitu -0,527 tidak berada pada wilayah kritis maka **gagal tolak H_0** . Dengan demikian, tidak terdapat bukti yang cukup untuk menyatakan bahwa proporsi suku cadang bebas cacat **berbeda** dari 90%.

F. Pengujian Hipotesis Tentang Varians

Pengujian hipotesis varians pada praktiknya dipergunakan sebagai variasi dari suatu kumpulan nilai hasil observasi. Ingat, bahwa akar dari varians merupakan simpangan baku/standar deviasi. Sebagai ilustrasi, suatu perusahaan ingin mengetahui variasi komponen dari mesin produksi; pejabat bank ingin mengetahui variasi pengeluaran kartu kredit; pemilik pabrik ingin mengetahui variasi ukuran dari hasil produksi; dan lain sebagainya.

Pengujian hipotesis varians pada dasarnya sama seperti pengujian hipotesis rata-rata dan proporsi. Seperti kita ketahui, jika suatu sampel acak ditarik dari suatu populasi dengan distribusi normal, maka rasio mengikuti fungsi Kai-Kuadrat χ^2 (*Chi-Square*) dengan derajat kebebasan $(n - 1)$ dan rumus statistik uji yaitu:

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

di mana:

n = ukuran sampel

s^2 = varians sampel

σ^2 = varians populasi berdasarkan hipotesis

Perumusan hipotesis varians seperti halnya dengan perumusan hipotesis rata-rata dan proporsi yaitu sebagai berikut:

- Uji sisi kanan

$$H_0: \sigma^2 \leq \sigma_0^2$$

$$H_1: \sigma^2 > \sigma_0^2$$

- Uji sisi kiri

$$H_0: \sigma^2 \leq \sigma_0^2$$

$$H_1: \sigma^2 > \sigma_0^2$$

- Uji dua sisi

$$H_0: \sigma^2 = \sigma_0^2$$

$$H_1: \sigma^2 \neq \sigma_0^2$$

Contoh 8.4

Sebuah pabrik mobil ingin apakah penggunaan jenis bahan baku tertentu mempengaruhi **variasi umur pakai ban** (dinyatakan dalam varians jarak tempuh sebelum aus). Menurut standar produksi, varians umur pakai ban yang ideal adalah 5000 km. Setelah menerapkan bahan baku baru, diambil **sampel acak sebanyak 25 ban**, dan diperoleh varians sampel sebesar 5300 km. Dengan menggunakan tingkat signifikansi $\alpha = 10\%$.

- Ujilah apakah variasi umur pakai ban lebih besar dari varians ideal.
- Ujilah apakah variasi umur pakai ban berbeda dari varians ideal.

Jawab:

Hitung **statistik uji** dengan menggunakan uji *Chi-Squared* untuk bagian a dan b:

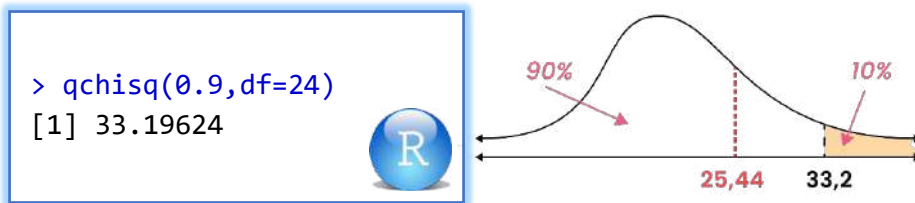
$$\chi^2_{hitung} = \frac{(n-1)s^2}{\sigma^2} = \frac{(25-1)5300}{5000} = 25,44$$

a. Uji sisi kanan

$$H_0: \sigma^2 \leq 5000$$

$$H_1: \sigma^2 > 5000$$

Selanjutnya menentukan nilai kritis dengan $df = 24$ dan $\alpha = 10\%$ (uji sisi kanan), sehingga diperoleh $\chi^2_{0,10(24)} = 33,20$.



Kesimpulan:

Diperoleh hasil $\chi^2_{hitung} = 25,44$ dan $\chi^2_{tabel} = 33,20$ dengan uji sisi kanan. Karena nilai χ^2_{hitung} yaitu 25,44 tidak berada pada wilayah kritis maka **gagal tolak H_0** . Dengan demikian, Tidak terdapat cukup bukti untuk menyatakan bahwa **variasi umur pakai ban lebih besar dari varians ideal 5000 km** pada tingkat signifikansi 10%.

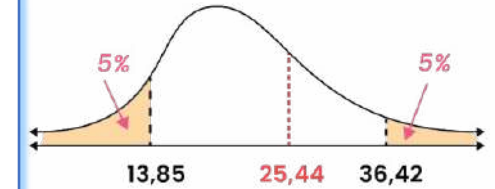
b. Uji dua sisi

$$H_0: \sigma^2 = 5000$$

$$H_1: \sigma^2 \neq 5000$$

Selanjutnya menentukan nilai kritis dengan $df = 24$ dan $\frac{\alpha}{2} = 5\%$ (uji dua sisi), sehingga diperoleh batas bawah $\chi^2_{0,05(24)} \approx 13,85$ dan batas atas $\chi^2_{0,95(24)} \approx 36,42$.

```
> qchisq(0.05, df=24)
[1] 13.84843
> qchisq(0.95, df=24)
[1] 36.41503
```



Kesimpulan:

Diperoleh hasil $\chi^2_{hitung} = 25,44$, batas bawah $\chi^2_{tabel} = 13,85$, dan batas atas $\chi^2_{tabel} \approx 36,42$ dengan uji dua sisi. Karena nilai χ^2_{hitung} yaitu 25,44 tidak berada pada wilayah kritis maka **gagal tolak H_0** . Dengan demikian, tidak terdapat cukup bukti untuk menyatakan bahwa **variasi umur pakai ban berbeda dari varians ideal 5000 km** pada tingkat signifikansi 10%.

G. Analisis Data Tentang Pengujian Hipotesis

Pada topik ini akan dibahas uji beda rata-rata satu sampel (*one sample t-test*) dan uji beda rata-rata dua sampel (*two sample t-test*) dengan menggunakan data Excel yang diolah datanya dengan software R Studio. Dalam suatu penelitian, peneliti sering dihadapkan dengan analisis data dalam melihat hubungan antar beberapa populasi berdasarkan parameter yang dimiliki

- Jika sampel berasal dari satu populasi & ingin dibandingkan apakah nilai rata-rata dari sampel tsb sama dengan satu nilai yang akan dihipotesiskan → menggunakan uji beda rata-rata satu sampel (*one sample t test*)
- Jika sampel berasal dari dua populasi yang berbeda → menggunakan uji beda rata-rata dua sampel (*two sample t test*)
- Jika sampel berasal dari lebih dua populasi yang berbeda → menggunakan analisis keragaman (*Analysis of Variance – ANOVA*)

Mengapa menggunakan Uji-t?

Uji beda biasanya dilakukan untuk membandingkan nilai tengah (rata-rata/modus/median) kelompok dengan suatu nilai. Misalnya, suatu penelitian ingin mengetahui apakah terdapat perbedaan tingkat kesembuhan pasien dengan diberikan vaksin baru. Uji yang bisa diterapkan yaitu uji Z, namun uji Z memiliki beberapa syarat yaitu:

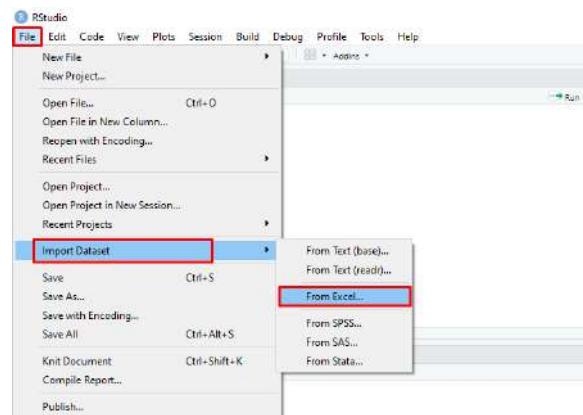
- Data memiliki distribusi normal
- Ragam populasi σ^2 harus diketahui

Namun biasanya data yang kita peroleh adalah data sampel bukan populasi dan data sampel biasanya berukuran kecil. Jika data berukuran kecil maka ragam sampel s_n^2 tidak cukup akurat untuk mendekati atau menduga ragam populasi σ^2 sehingga hasil uji Z menjadi kurang valid. Oleh karena itu, kita menggunakan uji t dalam uji beda. Uji t dapat digunakan untuk sampel berukuran kecil dan memiliki distribusi normal. Namun sebelum membahas contoh studi kasus, berikut akan dijelaskan untuk *import data* Excel dengan menggunakan R-Studio.

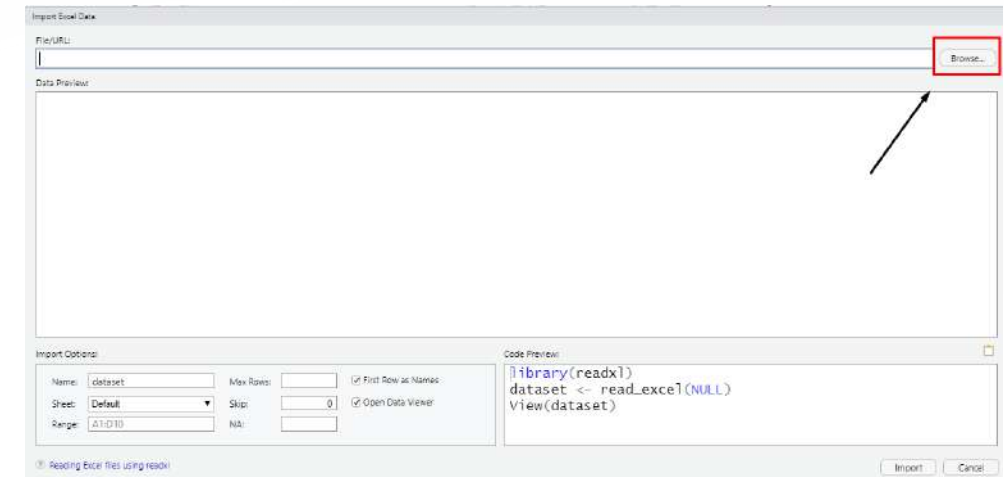
G.1. Import Data

Manajemen data diperlukan untuk mengolah data di R. Jika data yang kita miliki tersimpan dalam bentuk format Excel, CSV dan lain-lain, kita tidak perlu untuk mengetik ulang satu persatu data tsb ke R Studio namun cukup dengan menggunakan fitur *Import Data* untuk mengambil data dari *software* lain. Berikut ini merupakan langkah-langkah dalam *Import Data*:

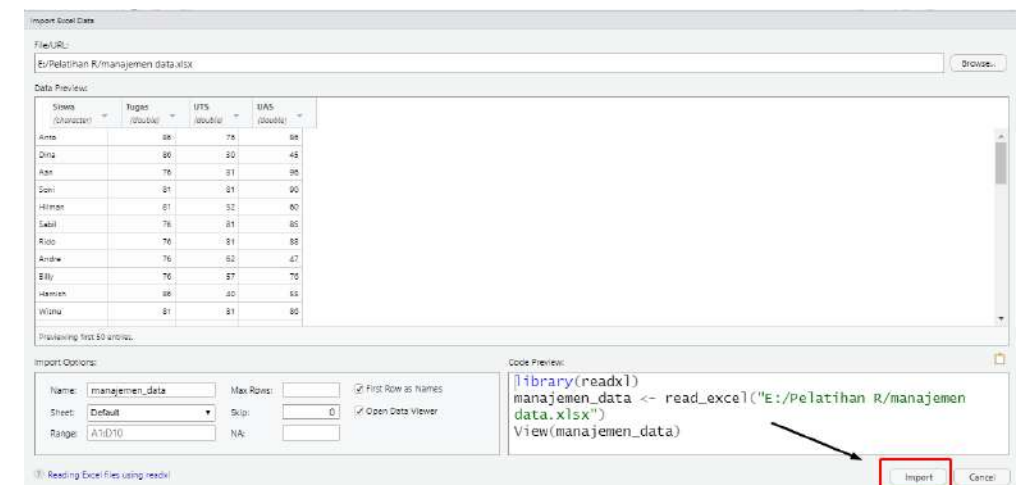
1. Klik File → Import Dataset → From Excel



2. Setelah klik “**From Excel**” maka muncul tampilan seperti di bawah ini, kemudian pilih “**Browse**”



3. Pilih file Excel atau file lain yang akan kita olah di R Studio.
4. Setelah memilih file yang diinginkan, akan tampil layar seperti di bawah ini. Selain itu, *Code Preview* juga akan berubah menyesuaikan lokasi file berada (perhatikan kotak merah pada gambar), kemudian klik “**Import**”.



5. Selanjutnya pada tampilan R *Console* akan tampil seperti gambar di bawah ini:

Siswa	Tugas	UTS	UAS
1 Anto	86	76	86
2 Dina	86	30	45
3 Aan	76	81	96
4 Soni	81	81	90
5 Hilman	81	52	60
6 Sabil	76	81	85
7 Rido	76	81	88
8 Andre	76	62	47
9 Billy	76	57	76
10 Hamish	86	40	55
11 Wisnu	81	81	86
12 Irfan	76	81	89

```

> library(readxl)
> manajemen_data <- read_excel("E:/Pelatihan R/manajemen data.xlsx")
> View(manajemen_data)
  
```

Jika import data telah berhasil maka data sudah bisa diolah sesuai kebutuhan.

G.2. Uji Beda Rata-Rata Satu Sampel (One Sample T-Test)

Sintaks uji beda rata-rata satu sampel adalah sebagai berikut:

```
t.test(x, alternative , mu, conf.level )
```

keterangan:

- x : data
- alternative: “less” untuk uji sisi kiri; “greater” untuk uji sisi kanan; “two.sided” untuk uji dua sisi
- mu: nilai tertentu yang ingin dibandingkan dengan nilai uji statistik
- conf.level : tingkat kepercayaan

Contoh 8.5

Pada suatu pengawasan masa studi akan dicari tahu lamanya mahasiswa menyelesaikan proposal tugas akhir. Kemudian diambil sampel secara acak sebanyak 20 mahasiswa dalam menyusun proposal tugas akhir. Kaprodi menyatakan bahwa penyusunan proposal tugas akhir idealnya adalah 80 hari. Dengan menggunakan tingkat kepercayaan 95%.

- Ujilah apakah waktu penyusunan proposal tugas akhir mahasiswa **kurang** dari 80 hari
- Ujilah apakah waktu penyusunan proposal tugas akhir mahasiswa **lebih** dari 80 hari
- Ujilah apakah waktu penyelesaian suatu kasus di pengadilan **berbeda** dari 80 hari

Data lengkap dalam format Excel, silakan scan QR code berikut atau kunjungi tautan: <https://bit.ly/OneSampletTest>



Jawab:

- Uji Sisi Kiri:

$H_0: \mu \geq 80$ Waktu penyelesaian proposal tugas akhir ideal atau lebih dari 80 hari

$H_1: \mu < 80$ Waktu penyelesaian proposal tugas akhir kurang dari ideal

Lakukan *import data* file excel ke R. Setelah import data, selanjutnya mencari nilai uji statistik rata-rata satu sampel dengan uji t pada R:

```
> t.test(one_sample_t_test$hari, alternative = "less", mu= 80, conf.level = 0.95) #nilai uji statistic dengan uji t
> qt(0.05,df=19) #nilai t tabel
[1] -1.729133
```



Diperoleh output ditunjukkan pada gambar berikut.

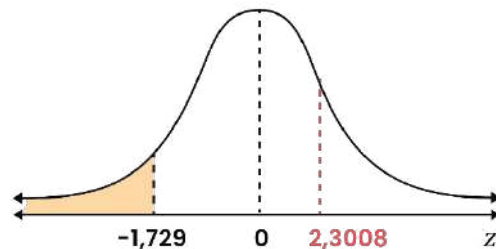
```
> t.test(one_sample_t_test$hari,alternative = "less",mu= 80, conf.level = 0.95)

One Sample t-test

data:  one_sample_t_test$hari
t = 2.3008, df = 19, p-value = 0.9836
alternative hypothesis: true mean is less than 80
95 percent confidence interval:
 -Inf 97.25258
sample estimates:
mean of x
 89.85
```

Kesimpulan:

Diperoleh hasil t-hitung 2,3008 dan t-tabel 1,729, karena t-hitung tidak berada pada wilayah kritis maka gagal tolak H_0 . Dengan hasil yang sama menggunakan *p-value*, diperoleh *p-value* 0,9836 dan $\alpha = 0,05$ sehingga *p-value* $> \alpha$ maka gagal tolak H_0 . Dengan demikian, dapat disimpulkan bahwa tidak terdapat cukup bukti waktu penyelesaian proposal tugas akhir kurang dari ideal (80 hari) dengan tingkat kepercayaan 95%.



b. Uji Sisi Kanan:

$H_0: \mu \leq 80$ Waktu penyelesaian proposal tugas akhir ideal atau kurang dari 80 hari

$H_1: \mu > 80$ Waktu penyelesaian proposal tugas akhir lebih dari ideal

Lakukan *import data* file excel ke R. Setelah import data, selanjutnya mencari nilai uji statistik rata-rata satu sampel dengan uji t pada R:

```
> t.test(one_sample_t_test$hari,alternative="greater",mu=
80, conf.level = 0.95) #nilai uji statistic dengan uji t
> qt(0.95,df=19) #nilai t tabel
[1] 1.729133
```



Diperoleh output ditunjukkan pada gambar berikut.

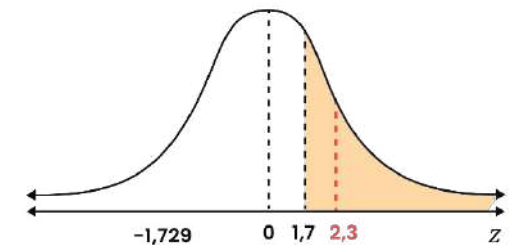
```
> t.test(one_sample_t_test$hari,alternative = "greater",mu= 80, conf.level = 0.95)
#nilai t hitung

One Sample t-test

data:  one_sample_t_test$hari
t = 2.3008, df = 19, p-value = 0.01645
alternative hypothesis: true mean is greater than 80
95 percent confidence interval:
 82.44742      Inf
sample estimates:
mean of x
 89.85
```

Kesimpulan:

Diperoleh hasil t-hitung 2,3008 dan t-tabel 1,729, karena t-hitung berada pada wilayah kritis maka tolak H_0 . Dengan hasil yang sama menggunakan *p-value*, diperoleh *p-value* 0,01645 dan $\alpha = 0,05$ sehingga *p-value* $< \alpha$ maka tolak H_0 . Dengan demikian, dapat disimpulkan bahwa terdapat cukup bukti waktu penyelesaian proposal tugas akhir lebih dari ideal (80 hari) dengan tingkat kepercayaan 95%.



c. Uji Dua Sisi:

$H_0: \mu = 80$ Waktu penyelesaian proposal tugas akhir ideal 80 hari

$H_1: \mu \neq 80$ Waktu penyelesaian proposal tugas akhir berbeda dari ideal

Lakukan *import data* file excel ke R. Setelah import data, selanjutnya mencari nilai uji statistik rata-rata satu sampel dengan uji t pada R:

```
> t.test(one_sample_t_test$hari,alternative="two.sided",mu=
80, conf.level = 0.95) #nilai uji statistic dengan uji t
> qt(0.025,df=19) #nilai t tabel
[1] -2.093024
> qt(0.975,df=19) #nilai t tabel
[1] 2.093024
```



Diperoleh output ditunjukkan pada gambar berikut.

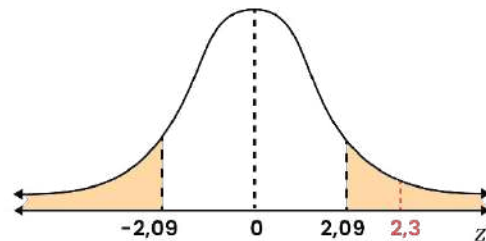
```
> t.test(one_sample_t_test$hari, alternative = "two.sided", mu = 80, conf.level = 0.95)
#nilai t hitung

One Sample t-test

data: one_sample_t_test$hari
t = 2.3008, df = 19, p-value = 0.0329
alternative hypothesis: true mean is not equal to 80
95 percent confidence interval:
 80.88957 98.81043
sample estimates:
mean of x
 89.85
```

Kesimpulan:

Diperoleh hasil t-hitung 2,3008 dan t-tabel $\pm 2,09$, karena t-hitung berada pada wilayah kritis maka tolak H_0 . Dengan hasil yang sama menggunakan *p-value*, diperoleh *p-value* 0,0329 dan $\alpha = 0,05$ sehingga *p-value* $< \alpha$ maka tolak H_0 . Dengan demikian, dapat disimpulkan bahwa terdapat cukup bukti waktu penyelesaian proposal tugas akhir berbeda dari ideal (80 hari) dengan tingkat kepercayaan 95%.



G.3.1. Uji Beda Rata-Rata Dua Sampel Independen (*Unpaired Two Sample t-Test*)

Uji beda rata-rata dua sampel independent dilakukan untuk mengetahui rata-rata antar kedua populasi sama atau tidak. Kedua populasi ini merupakan sampel yang saling independent artinya kedua populasi tersebut tidak saling berhubungan. Syarat-syarat penggunaan uji dua sampel saling bebas:

- Jumlah data kurang dari 30, jika sama atau lebih dari 30 gunakan uji Z
- Data berskala interval atau rasio
- Data dari kedua sampel berdistribusi normal: bisa menggunakan uji Saphiro-Wilks
- Perlu diketahui apakah kedua kelompok berasal dari populasi yang ragamnya sama atau tidak

Berikut ini sintaks untuk uji normalitas dan uji ragam/varians:

```
shapiro.test(x) #uji normalitas
Var.test(a,b) #uji ragam
```

Selanjutnya menguji *Unpaired Two Sample t-Test* dapat dilakukan dengan sintaks

```
t.test(a,b, var.equal, alternative, paired)
```

keterangan:

- x : data
- a,b : data a dan data b adalah data yang akan dibandingkan
- var.equal : TRUE jika varians sama, FALSE jika varians kedua kelompok beda

- paired : FALSE untuk uji bukan berpasangan

Contoh 8.6

Seorang dosen statistik ingin menguji apakah terdapat perbedaan kemampuan pemahaman konsep antar kelas reguler dan ekstensi. Peneliti tersebut mengambil sampel secara acak sebanyak 15 orang mahasiswa dari masing-masing kelas. Apakah ada perbedaan kemampuan antar mahasiswa kelas reguler dan ekstensi pada mata kuliah statistik?

Data lengkap dalam format Excel, silakan scan QR code berikut atau kunjungi tautan:

<https://bit.ly/UnpairedSampleTest1>



Jawab:

Langkah pertama, uji dua sisi:

$H_0: \mu_a = \mu_b$ Kemampuan mahasiswa kelas reguler dan ekstensi sama

$H_1: \mu_a \neq \mu_b$ Terdapat perbedaan kemampuan antar mahasiswa kelas reguler dan ekstensi

Langkah kedua, lakukan *import data* file excel ke R. Setelah import data, selanjutnya uji normalitas pada data dengan hipotesis:

H_0 = Data berdistribusi normal

H_1 = Data tidak berdistribusi normal

```
> a <- unpaired_two_sample_1$ekstensi
> b <- unpaired_two_sample_1$reguler
> shapiro.test(a)
      Shapiro-Wilk normality test
data:  a
W = 0.92384, p-value = 0.2204
> shapiro.test(b)
      Shapiro-Wilk normality test
data:  b
W = 0.96701, p-value = 0.8115
```



Berdasarkan output di atas diperoleh dua nilai *p-value* 0,2204 dan 0,8115 dimana keduanya lebih besar dari taraf signifikansi yang digunakan yaitu 0,05. Karena *p-value* > α maka gagal tolak H_0 maka dengan keyakinan 95% maka cukup bukti menyatakan distribusi data kedua grup kelas tersebut **berdistribusi normal**. Selanjutnya dilakukan uji varians dengan hipotesis awal dalam uji kesamaan ragam adalah varian antara dua kelompok yang diuji adalah sama atau homogen.

```
> var.test(a,b)
      F test to compare two variances
data:  a and b
F = 0.82426, num df = 14, denom df = 14, p-value = 0.7227
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.2767275 2.4551214
sample estimates:
ratio of variances
      0.8242569
```



Diperoleh *p-value* yaitu 0,7227 dimana lebih besar dari 0,05 dan dapat dikatakan bahwa dengan tingkat keyakinan 95% cukup bukti menyatakan bahwa ragam/varian antar dua kelompok adalah sama atau homogen. Langkah selanjutnya mencari nilai uji statistik dengan uji t pada R:

```
> t.test(a,b, var.equal = TRUE, alternative = "two.sided", conf.level = 0.95 )
```



Output sebagai berikut:

```
> t.test(a,b, var.equal = TRUE, alternative = "two.sided", conf.level = 0.95 )

Two Sample t-test

data: a and b
t = 12.671, df = 28, p-value = 4.081e-13
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 25.59742 35.46925
sample estimates:
mean of x mean of y
 84.73333  54.20000
```

Dari hasil *output* diperoleh *p-value* $4,08 \times 10^{-13}$ lebih kecil dari taraf signifikansi (α) yaitu 0,05 sehingga tolak H_0 maka dapat disimpulkan dengan tingkat kepercayaan 95% cukup bukti terdapat perbedaan kemampuan antar kelas ekstensi dan regular pada mata kuliah statistik.

G.3.2. Uji Beda Rata-Rata Dua Sampel Berpasangan (*Paired Two Sample t-Test*)

Uji beda rata-rata dua sampel berpasangan dilakukan ketika dua kelompok berasal dari populasi yang sama, di mana kelompok pertama merupakan kelompok dengan kondisi awal, sedangkan kelompok kedua merupakan kelompok yang telah menerima perlakuan tertentu. Misalnya, seorang guru ingin mengetahui apakah terdapat perbedaan signifikan pada hasil belajar suatu kelas yang diajar menggunakan metode ceramah (kelompok 1) dibandingkan dengan metode *contextual learning* (kelompok 2). Sintaks uji *t* berpasangan di R adalah sebagai berikut:

```
t.test(a,b, alternative , paired, conf.level )
```

Keterangan:

- a : data kelompok 1
- b : data kelompok 2
- alternative : “two.sided” untuk uji dua sisi
- paired : TRUE untuk uji berpasangan
- conf.level : tingkat kepercayaan

Contoh 8.7

Suatu perusahaan yang memproduksi motor ingin mengetahui apakah terdapat perbedaan yang signifikan pada penjualan motor antara sebelum dan sesudah adanya ketetapan kenaikan pajak kendaraan. Selanjutnya diambil data secara acak dari 15 dealer. Ujilah apakah terdapat perbedaan yang signifikan pada penjualan motor antara sebelum dan sesudah adanya ketetapan kenaikan pajak kendaraan dengan tingkat kepercayaan 95%?

Jawab:

Langkah pertama adalah menentukan hipotesisnya:

$H_0: \mu_a = \mu_b$ Penjualan motor antara sebelum dan sesudah kenaikan pajak adalah sama

$H_1: \mu_a \neq \mu_b$ Terdapat perbedaan penjualan motor antara sebelum dan sesudah kenaikan pajak

Langkah kedua, lakukan *import data* file excel ke R. Setelah import data, selanjutnya uji normalitas pada data dengan menggunakan uji normalitas (*Shapiro Wilk Normality Test*). Berikut ini merupakan hipotesis uji normalitas:

H_0 = Data berdistribusi normal

H_1 = Data tidak berdistribusi normal

```
> before <- paired_t_test$sebelum
> after <- paired_t_test$sesudah
> shapiro.test(before- after)
Shapiro-Wilk normality test
data: before - after
W = 0.93176, p-value = 0.2898
```



Data lengkap dalam format Excel, silakan scan QR code berikut atau kunjungi tautan: <https://bit.ly/PairedtTest>



Dari hasil di atas diperoleh $p\text{-value}$ $0,2898 > 0,05 (\alpha)$ sehingga gagal tolak H_0 maka cukup bukti data berdistribusi normal. Selanjutnya dilakukan analisis uji t berpasangan dengan sintaks sebagai berikut:

```
> t.test(before,after, alternative = "two.sided", paired  
= TRUE, conf.level = 0.95)
```



Output sebagai berikut:

```
> t.test(before,after, alternative = "two.sided", paired = TRUE, conf.level = 0.95)

Paired t-test

data: before and after
t = -1.0312, df = 14, p-value = 0.32
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.695969  1.295969
sample estimates:
mean of the differences
-1.2
```

Dari hasil *output* diperoleh $p\text{-value}$ 0,32 lebih besar dari taraf signifikansi (α) yaitu 0,05 sehingga gagal tolak H_0 maka dapat disimpulkan tidak cukup bukti terdapat perbedaan penjualan motor antara sebelum dan sesudah kenaikan pajak dengan tingkat kepercayaan 95%.

LATIHAN 8

Pengujian Hipotesis

1. Sebuah perusahaan manufaktur sedang menguji kualitas produk pelat baja yang diproduksi oleh mesin baru. Mesin tersebut diklaim menghasilkan pelat dengan ketebalan rata-rata 10 mm. Untuk memverifikasi klaim tersebut, dilakukan pengukuran terhadap 50 pelat baja secara acak, dan hasilnya menunjukkan bahwa rata-rata ketebalan pelat adalah 12,6 mm, dengan simpangan baku 1,5 mm. Dengan tingkat kepercayaan 95%:
 - a. Ujilah apakah ketebalan rata-rata pelat yang dihasilkan oleh mesin tersebut lebih dari 10 mm.
 - b. Ujilah apakah ketebalan rata-rata pelat yang dihasilkan oleh mesin tersebut berbeda dari 10 mm.
2. Suatu bengkel teknik mesin sedang menguji performa kompresor udara baru yang dirancang untuk menghasilkan tekanan rata-rata minimal 8 bar. Untuk menguji klaim tersebut, dilakukan pengujian terhadap 25 unit kompresor secara acak, dan diperoleh rata-rata tekanan 7,6 bar dengan simpangan baku 0,9 bar. Dengan tingkat kepercayaan 95%:
 - Ujilah apakah kompresor tersebut menghasilkan tekanan rata-rata kurang dari 8 bar.
 - Ujilah apakah kompresor tersebut menghasilkan tekanan rata-rata berbeda dari 8 bar.
3. Pabrik PT.XYZ memproduksi komponen mesin dan mengklaim bahwa setidaknya 90% dari produknya lolos uji mutu. Seorang inspektur memeriksa 180 komponen secara acak dan menemukan bahwa 154 komponen lolos uji. Dengan tingkat signifikansi 5%:
 - a. Apakah proporsi komponen yang lolos uji kurang dari 90%.
 - b. Apakah proporsi komponen yang lolos uji lebih dari 90%.
4. Perusahaan energi terbarukan mengembangkan panel surya baru yang diklaim memiliki tingkat keberhasilan konversi energi sebesar 92%

dalam kondisi tertentu. Untuk menguji klaim tersebut, tim peneliti mengamati 200 panel surya yang dipasang di berbagai lokasi, dan menemukan bahwa 162 panel berhasil mencapai standar efisiensi yang ditentukan. Dengan tingkat signifikansi 10%, ujilah apakah tingkat keberhasilan panel surya berbeda dari 92%.

5. Laboratorium X ingin mengevaluasi stabilitas output daya turbin angin dari hasil riset. Berdasarkan spesifikasi pabrik, varian daya output maksimum yang diperbolehkan adalah 25 (dalam satuan kW^2) untuk menjaga performa optimal. Untuk menguji klaim tersebut, dilakukan pengukuran terhadap sampel 16 hari pengoperasian, dan diperoleh simpangan baku sampel sebesar 6 kW . Dengan tingkat signifikansi 5%. Apakah varians daya output turbin lebih besar dari yang diperbolehkan (25 kW^2)
6. Divisi HRD pada PT. XYZ berpendapat bahwa rata-rata pengeluaran per hari karyawannya lebih dari Rp70 ribu. Untuk menguji pendapat tersebut ada 10 karyawan yang diwawancara dan diperoleh data pengeluaran per hari (dalam ribu rupiah) adalah sebagai berikut: 35, 50, 63, 65, 80, 105, 120, 130, 145, 150. Dengan menggunakan $\alpha = 10\%$, ujilah pendapat tersebut.
7. Suatu obat ketegangan syaraf diduga hanya 75% efektif. Hasil percobaan dengan obat baru terhadap 50 orang dewasa penderita ketegangan syaraf yang diambil secara acak, ternyata dari hasil penelitian terdapat obat tersebut bekerja secara efektif pada 85 orang. Apakah ini merupakan bukti yang cukup untuk menyimpulkan bahwa obat baru itu lebih baik dari yang beredar sekarang? Gunakan $\alpha = 1\%$.
8. Seorang pejabat perusahaan batu baterai mengatakan bahwa penggantian suatu bahan baku akan mempengaruhi variasi lama hidup batu baterai yang dinyatakan dengan varians 120 jam. Untuk maksud tersebut dipilih sampel acak sebanyak 40 buah baterai dari jenis yang sama. Setelah dilakukan penggantian bahan baku tersebut ternyata **variens** batu baterai adalah 132 jam. Dengan menggunakan tingkat kepercayaan 95%. Ujilah pendapat tersebut dengan uji sisi kiri.

BAB 9

Anlisis Korelasi & Regresi

Pendahuluan

Kemampuan Akhir yang Diharapkan	Mahasiswa mampu memahami masalah yang berkaitan dengan analisis regresi sederhana, berganda, dan logistik
Bahan Kajian	- Analisis Regresi Sederhana - Analisis Regresi Berganda - Analisis Regresi Logistik
Metode Pembelajaran	Ceramah interaktif, studi kasus, <i>collaborative learning</i>
Pengalaman Belajar Mahasiswa	- Mengidentifikasi permasalahan nyata dalam bidang teknik atau industri yang dapat dianalisis menggunakan model regresi. - Menganalisis regresi sederhana, berganda, dan logistik untuk mengetahui hubungan antara satu variabel terikat dengan satu atau lebih variabel bebas. - Menghitung koefisien regresi dan konstanta menggunakan perangkat lunak statistik. - Menafsirkan hasil analisis, termasuk koefisien determinasi dan uji signifikansi.
Kriteria Penilaian (Indikator)	- Mampu merumuskan masalah dengan tepat dan menentukan jenis regresi yang sesuai berdasarkan karakteristik data. - Mampu menghitung koefisien regresi, nilai R^2, serta menginterpretasikan arah dan kekuatan hubungan antar variabel. - Mampu menjalankan analisis menggunakan perangkat lunak dan menafsirkan output dengan benar.

Tujuan Pembelajaran

1. Memahami konsep dasar regresi sederhana, berganda, dan logistik serta perbedaannya dalam konteks analisis hubungan antar variabel.
2. Menafsirkan koefisien regresi, nilai p, dan koefisien determinasi secara statistik dan substantif sesuai konteks masalah.
3. Menggunakan perangkat lunak statistik (seperti *R Studio*) dalam melakukan analisis regresi, menguji signifikansi model, serta menafsirkan hasil analisis secara tepat.

ANALISIS REGRESI

Analisis regresi dalam Statistika adalah suatu metode untuk menentukan hubungan sebab akibat antara satu peubah atau variabel dengan satu atau lebih variabel lain. Variabel-variabel tersebut dikenal sebagai variabel *prediktor* atau variabel bebas dan variabel *respon* atau variabel tak bebas. Hubungan antar variabel tersebut dapat dibuat dalam bentuk persamaan matematik atau model matematik. Dalam rentang tertentu analisis regresi dapat digunakan untuk memprediksi nilai variabel tak bebas bila variabel bebas diketahui berdasarkan persamaan regresi yang sudah didapatkan sebelumnya. Dengan analisis regresi pula seorang peneliti dapat mengetahui suatu variabel bebas mempengaruhi atau tidak variabel tak bebas dan mengukur kekuatan hubungan antar variabel bebas dan tak bebas.

A. Regresi Linear Sederhana

Regresi linear sederhana merupakan suatu model regresi yang menggambarkan hubungan linear antara satu peubah/variabel bebas (x) dengan peubah atau variabel tidak bebas (y) secara linear. Yang dimaksud linier disini adalah linier dalam parameter. Model regresi linear adalah sebagai berikut:

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (9.1)$$

y disebut peubah/variabel tidak bebas (*dependent variabel*) atau variabel respon.

β_0 dan β_1 adalah parameter-parameter model dan x disebut peubah atau variabel bebas (*independent variable*) atau variabel prediktor. Sedangkan ε disebut error/galat atau residual model.

Persamaan (9.1) diestimasi atau ditaksir atau diduga oleh persamaan $\hat{y} = b_0 + b_1 x$ dengan b_0 adalah taksiran/dugaan/estimasi untuk β_0 merupakan intersep/titik potong dengan sumbu y , b_1 adalah taksiran dari β_1 merupakan slope/kemiringan dan $\hat{y}$ adalah nilai prediksi dari y . Dalam model regresi ε atau residual diasumsikan berdistribusi normal dengan mean (rata-rata) mendekati 0 dan simpangan baku σ . Jika residual mendekati 0 berarti selisih nilai taksiran dan nilai sebenarnya sangat kecil atau nilai taksiran mendekati nilai yang sebenarnya.

A.1 Diagram pencar atau *scatter plot*

Scatter plot menunjukkan plot titik dalam diagram kartesius. Setiap titik merupakan koordinat dari 2 peubah yaitu 1 peubah pada sumbu x dan 1 peubah lainnya pada sumbu y . Terdapat banyak cara untuk membuat *Scatter plot*, di R Studio salah satunya dengan menggunakan fungsi **plot(x,y)**.

```
plot(x,y, xlab, ylab, main, col)
```

Keterangan:

- x : data berupa vektor numerik yang digunakan pada sumbu x
- y : data berupa vektor numerik yang digunakan pada sumbu y
- $xlab$: nama label pada sumbu x
- $ylab$: nama label pada sumbu y
- $main$: judul dari plot
- col : warna *scatterplot*

Contoh 9.1

Sebuah pabrik mobil ingin mengetahui apakah penambahan sejenis bahan baku (x) dapat mempengaruhi kenaikan lamanya hidup ban (y) yang diproduksinya. Berikut ini merupakan percobaan yang dihasilkan dari

penambahan bahan baku. Data hasil percobaan ditunjukkan pada tabel berikut. Gambarlah *scatter plot* nya.

x	1	2	4	6	7	9	10	12	15
y	3	5	7	8	10	13	12	14	17

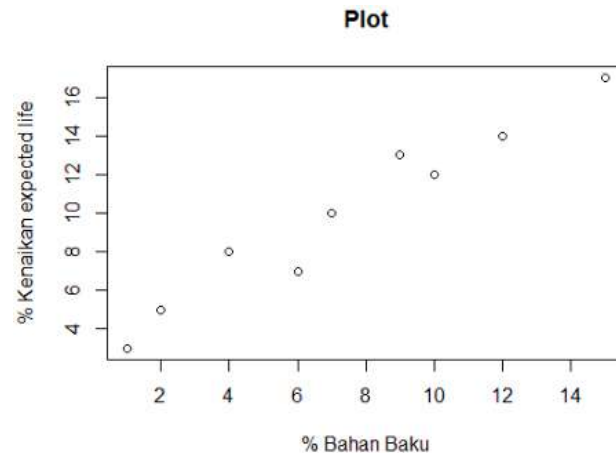
Jawab:

Untuk mempermudah pembuatan diagram pencar untuk contoh soal ini menggunakan R Studio . Dengan menggunakan fungsi **plot(x,y)** maka sintaksnya adalah:

```
> x <- c(1,2,4,6,7,9,10,12,15)
> y <- c(3,5,8,7,10,13,12,14,17)
> plot(x,y, xlab = "% Bahan Baku", ylab = "% Kenaikan expected life", main = "Plot" )
```



Output:



A.2 Metode Kuadrat Terkecil

Metode Kuadrat Terkecil atau *Method of Least Squares* adalah prosedur secara statistik untuk menemukan persamaan regresi atau garis regresi

terbaik (*best-fitting line*) yaitu dengan prinsip meminimalkan jumlah kuadrat dari deviasi/simpangan/selisih dari nilai pengamatan y dengan nilai prediksi $\hat{y}$ yang disebut dengan jumlah kuadrat error atau *sum of square of error* (SSE).

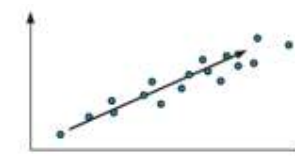
$$SSE = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - b_0 - b_1 x)^2$$

Penduga kuadrat terkecil untuk parameter β_0 dan β_1 diestimasi oleh b_0 dan b_1 dengan $b_1 = \frac{S_{xy}}{S_{xx}}$ dan $b_0 = \bar{y} - b_1 \bar{x}$. Bila jumlah pengamatan adalah n maka $\bar{y} = \frac{\sum y_i}{n}$ dan $\bar{x} = \frac{\sum x_i}{n}$. Besaran S_{xy} dan S_{xx} didefinisikan dengan $S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n}$ dan $S_{xx} = \sum (x_i - \bar{x})^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$.

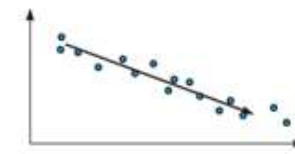
A.3 Koefisien Korelasi

Koefisien korelasi merupakan ukuran yang menunjukkan tingkat keeratan hubungan antara dua variabel atau peubah. Ukuran ini paling sering digunakan untuk mengetahui seberapa kuat dan searah apa hubungan antara dua variabel. Dalam pembahasan berikut, akan dijelaskan mengenai arah, linearitas, dan kekuatan dari suatu korelasi.

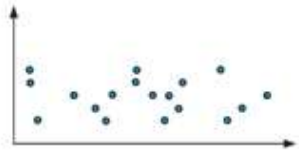
Interpretasi Arah Korelasi



Untuk tren yang umumnya meningkat, korelasinya dikatakan positif. Peningkatan variabel independen berarti variabel dependen secara umum meningkat.



Untuk tren yang umumnya menurun, korelasinya dikatakan negatif. Peningkatan variabel independen berarti variabel dependen umumnya menurun.

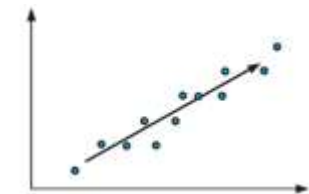


Untuk titik-titik yang tersebar secara acak, tanpa tren naik atau turun, kita katakan tidak ada korelasi.

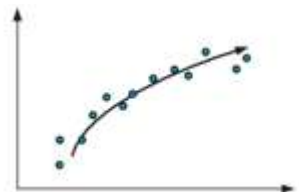
Gambar 9.1 Arah Korelasi

Interpretasi Linearitas Korelasi

Linearitas menunjukkan sejauh mana titik-titik data mengikuti pola garis lurus. Dengan kata lain, kita menilai apakah hubungan antara dua variabel dapat digambarkan secara kira-kira dengan garis lurus.



Titik-titik data tampak mengikuti pola yang hampir membentuk garis lurus, sehingga dapat dikatakan kira-kira linier.



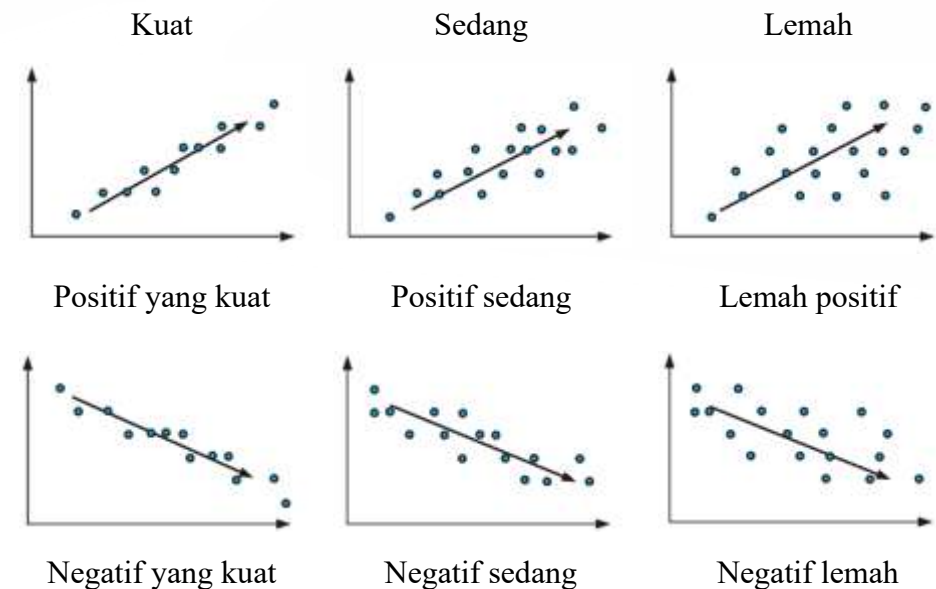
Titik-titik data membentuk pola yang melengkung, sehingga tidak mengikuti tren linier.

Gambar 9.2 Linearitas

Interpretasi Kekuatan Korelasi

Gambar di bawah ini menunjukkan berbagai tingkat kekuatan dan arah hubungan antara dua variabel. Jika titik-titik data membentuk pola yang menaik dan rapat di sekitar garis lurus, maka hubungan disebut korelasi positif yang kuat, sedangkan jika pola menaik masih terlihat tetapi lebih menyebar disebut korelasi positif sedang atau lemah. Sebaliknya, jika titik-titik membentuk pola menurun, hubungan tersebut bersifat korelasi negatif, yang dapat dikategorikan sebagai kuat, sedang, atau lemah tergantung pada seberapa rapat titik-titik mengikuti garis tren. Semakin dekat titik-titik data

terhadap garis lurus, semakin kuat hubungan antarvariabel; dan semakin menyebar titik-titik tersebut, semakin lemah korelasinya.



Gambar 9.3 Kekuatan Korelasi

Koefisien korelasi dilambangkan dengan huruf **r** (*Pearson's correlation coefficient*). Rumus koefisien korelasi pearson pada persamaan (9.2) dibawah ini dan pada tabel berikut ditunjukkan pula interpretasi dari nilai korelasi (**r**):

$$r_{xy} = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{\{n \sum x^2 - (\sum x)^2\} \{n \sum y^2 - (\sum y)^2\}}} \text{ atau } r_{xy} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} \quad (9.2)$$

Nilai **r** selalu berada di antara **-1 dan +1**, yaitu menunjukkan arah dan kekuatan hubungan antara dua variabel. Berikut ini Tabel interval nilai korelasi:

Interval Korelasi	Interpretasi
$0 < r < 0.49$	Hubungan lemah
$0.5 < r < 0.79$	Hubungan sedang
$0.79 < r < 1$	Hubungan kuat

Keeratan hubungan antara variabel respons dan variabel prediktor dapat pula dilihat dari nilai **koefisien determinasi** r^2 . Koefisien determinasi dapat diinterpretasikan sebagai persen pengurangan variasi total dalam percobaan yang diperoleh dengan menggunakan garis regresi $\hat{y} = b_0 + b_1x$, dari pada mengabaikan x dan menggunakan rata-rata sampel $\bar{y}$ untuk memprediksi variabel respons y atau koefisien determinasi menjelaskan besarnya variasi respons yang dapat dijelaskan prediktor.

Penggunaan Regresi Untuk Tujuan Prediksi

Setelah Anda menemukan persamaan garis regresi, Anda dapat menggunakannya untuk memprediksi nilai y untuk nilai x tertentu. Jika nilai x yang Anda gunakan dalam persamaan berada dalam rentang data (interpolasi), maka prediksi tersebut dapat dianggap reliabel. Namun, jika nilai y berada di luar rentang data (ekstrapolasi), prediksi tersebut harus diperlakukan dengan hati-hati karena tidak ada jaminan bahwa hubungan tersebut berlanjut melampaui nilai yang diamati.

Contoh 9.2

Berdasarkan data dengan nilai x antara 10 dan 20 dengan koefisien korelasi 0,93 garis regresi terbentuk: $y = 10,2x - 5,3$.

- a. Gunakan persamaan ini untuk memprediksi nilai y ketika:
 - i. $x = 3$
 - ii. $x = 13$
- b. Berikan komentar tentang keandalan (*reliabel*) jawaban pada bagian a.

Jawab:

- a. i. $y = 10,2(3) - 5,3 = 25$
 ii. $y = 10,2(13) - 5,3 = 127$
- b. Prediksi ketika $x = 13$ dapat dianggap andal karena 13 berada dalam rentang nilai x yang diketahui. Prediksi saat $x = 3$, bagaimanapun tidak

dapat dianggap andal, karena hubungan tersebut harus diekstrapolasi secara signifikan melampaui rentang data yang diberikan untuk membuat prediksi ini.

Contoh 9.3

Tentukan persamaan regresi untuk data berpasangan pada tabel berikut ini.

No	x_i	y_i	No	x_i	y_i
1	39	65	6	47	89
2	43	78	7	28	73
3	21	52	8	75	98
4	64	82	9	34	56
5	57	92	10	52	75

Jawab:

Cara menghitung manual untuk menentukan b_0 dan b_1 . Untuk mempermudah dibuat tabel sebagai berikut.

No	x_i	y_i	x_i^2	x_iy_i	y_i^2
1	39	65	1521	2535	4225
2	43	78	1849	3354	6084
3	21	52	441	1092	2704
4	64	82	4096	5248	6724
5	57	92	3249	5244	8464
6	47	89	2209	4183	7921
7	28	73	784	2044	5329
8	75	98	5625	7350	9604
9	34	56	1156	1904	3136

10	52	75	2704	3900	5625
Jumlah	460	760	23.634	36.854	59.816

$$b_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} = \frac{36.854 - (460)(760)/10}{23.634 - (460)^2/10} = \frac{1894}{2474} = 0,76556$$

$$b_0 = \bar{y} - b_1 \bar{x} = \frac{\sum y_i}{n} - b_1 \frac{\sum x_i}{n} = \frac{760}{10} - 0,76556 \frac{460}{10} = 76 - (0,76556)(46) = 40,78424$$

$$S_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n} = 59.816 - \frac{(760)^2}{10} = 2056$$

$$r^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}} = \frac{(1894)^2}{(0,76556)(2056)} = \frac{3587236}{5086544} = 0,70524$$

$$r = 0,89786$$

Jadi diperoleh persamaan regresi untuk data di atas adalah :

$$\hat{y} = b_0 + b_1 x = 40,78424 + 0,76556x$$

Dari persamaan regresi tersebut dapat dikatakan jika variable x adalah 0 maka variable y bernilai 40,78424. Koefisien regresi 0,76556 menunjukkan setiap kenaikan 1 satuan pada variable x maka variable y rata-rata akan naik sebesar 0,76556. Hubungan antara variabel x dan variable y cukup erat yang ditunjukkan oleh nilai koefisien korelasi $r = 0,89786$. Nilai koefisien determinasi $r^2 = 0,70524$ menyatakan bahwa 70,5% variasi respon dapat dijelaskan oleh prediktor.

Contoh 9.4

Sebuah pabrik mobil ingin mengetahui apakah penambahan sejenis bahan baku dapat mempengaruhi kenaikan lamanya hidup ban sampai aus yang diproduksinya. Berikut ini merupakan percobaan yang dihasilkan dari penambahan bahan baku:

X (%)	1	2	4	6	7
Y (%)	3	5	7	8	10

X = % penambahan bahan baku

Y = % penambahan lamanya hidup ban (hingga aus)

- Buat persamaan regresi liniernya
- Apakah % kenaikan bahan baku berpengaruh terhadap % kenaikan lamanya hidup ban (hingga aus) dengan tingkat keyakinan 95%?
- Hitung koefisien korelasi pearson dan interpretasikan hasilnya.
- Presiksilah kenaikan % lamanya hidup ban (hingga aus) jika bahan baku ditambah 5%

Jawab:

- Persamaan regresi linier: $Y = a + bX$

X	Y	XY	X^2	Y^2
1	3	3	1	9
2	5	10	4	25
4	7	28	16	49
6	8	48	36	64
7	10	70	49	100
$\Sigma X = 20$	$\Sigma Y = 33$	$\Sigma XY = 159$	$\Sigma X^2 = 106$	$\Sigma Y^2 = 247$

Cara perhitungan manual

Hitung b :

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

$$b = \frac{33 - (1,0385)(20)}{5} = \frac{795 - 660}{530 - 400} = \frac{135}{130} = 1,0385$$

Hitung a:

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

$$a = \frac{33 - (1,0385)(20)}{5} = \frac{33 - 20,77}{5} = \frac{12,23}{5} = 2,446$$

Maka, persamaan regresinya:

$$Y = 2,446 + 1,0385X$$

Cara menggunakan R

```
> x <- c(1,2,4,6,7)
> y <- c(3,5,7,8,10)
> persamaan_regresi <- lm(y~x)
> summary(persamaan_regresi)
```

Output:

```
Call:
lm(formula = y ~ x)

Residuals:
    1      2      3      4      5 
-0.4846  0.4769  0.4000 -0.6769  0.2846 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.4462    0.5619   4.354  0.02239 *
x            1.0385    0.1220   8.510  0.00341 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6222 on 3 degrees of freedom
Multiple R-squared:  0.9602,    Adjusted R-squared:  0.947
F-statistic: 72.42 on 1 and 3 DF, p-value: 0.003408
```

Maka, persamaan regresinya $Y = 2,4462 + 1,0385X$.

Diperoleh koefisien determinasi (R-Squared) yaitu 0.9602 sehingga dapat dikatakan variasi % kenaikan lamanya hidup ban (hingga aus) dapat dijelaskan oleh variasi % kenaikan bahan baku sebesar 96.02%. Sedangkan sisanya 3.98% dijelaskan oleh variabel lain yang tidak dibahas oleh penelitian tersebut.

- b. Gunakan uji signifikan dengan hipotesis:

$H_0: b = 0$ (tidak ada pengaruh)

$H_1: b \neq 0$ (ada pengaruh)

Berdasarkan hasil *output* dengan menggunakan *software* R, diperoleh *p-value* sebesar 0.0034 dan lebih kecil dari alpha (0.05) sehingga tolak H_0 . Maka dapat dikatakan bahwa terdapat hubungan yang signifikan antara % kenaikan bahan baku terhadap % kenaikan lamanya hidup ban (hingga aus)

- c. Hitung koefisien korelasi Pearson (r)

$$r = \frac{n \sum XY - \sum X \sum Y}{\sqrt{[n \sum X^2 - (\sum X)^2][n \sum Y^2 - (\sum Y)^2]}}$$

$$r = \frac{5(159) - 20(33)}{\sqrt{(5.106 - 400)(5.247 - 1089)}}$$

$$= 0,9798$$

Nilai $r = 0,9798 \rightarrow$ mendekati 1 $\rightarrow$ hubungan sangat kuat & positif antara penambahan bahan baku dan peningkatan umur ban.

- d. Gunakan persamaan regresi:

$$Y = 2,446 + 1,0385(5) = 2,446 + 5,1925 = 7,6385$$

Jadi, prediksi kenaikan umur ban adalah sekitar 7,64% jika bahan baku ditambah 5%.

Contoh 9.5

Suatu penelitian ingin mengetahui pengaruh tinggi badan (x) terhadap berat badan (y) di suatu daerah. Pada penelitian ini diambil sampel secara acak sebanyak 20 orang di daerah tersebut.

Data lengkap dalam format Excel, silakan scan QR code berikut atau kunjungi tautan: <https://bit.ly/RegresiLinear-R>

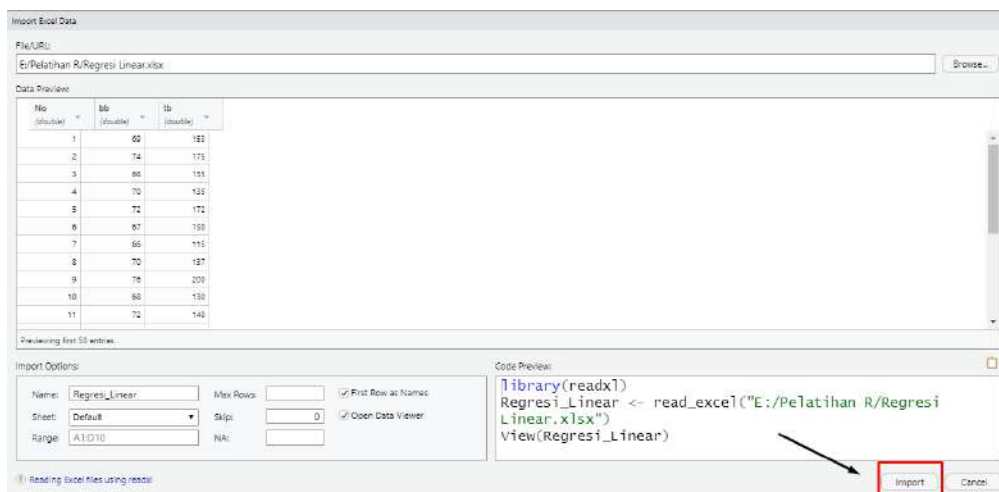


Berdasarkan data tersebut tentukan:

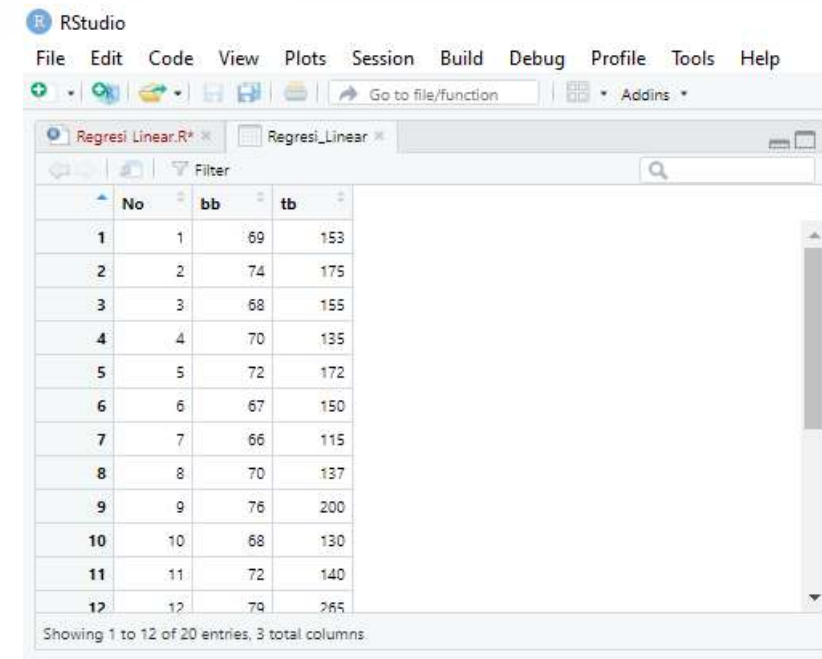
- Model regresi linear.
- Koefisien determinasi
- Dengan menggunakan uji koefisien regresi, apakah tinggi badan berpengaruh terhadap berat badan?
- Tentukan korelasi antara tinggi badan dan berat badan. Berikan interpretasi dari hasil tersebut.
- Prediksi berat badan jika tinggi badan 150 cm

Jawab:

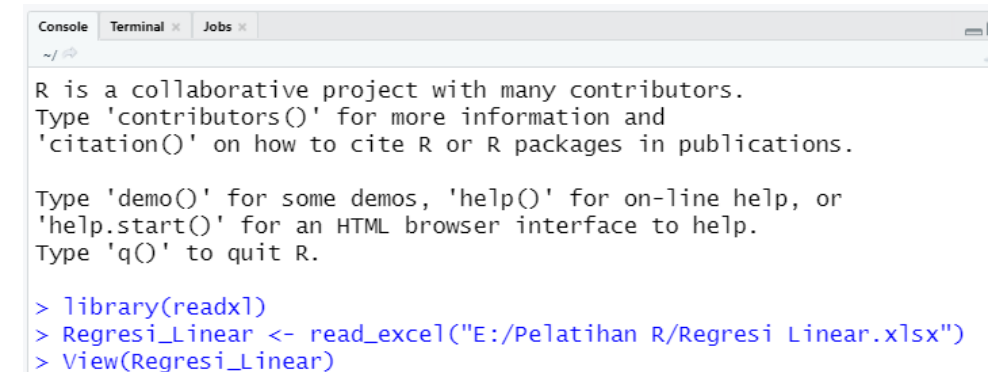
Langkah pertama adalah melakukan import data dari Excel ke R.



Setelah klik “**import**” maka akan muncul tampilan seperti gambar di bawah ini:



Kemudian pada tampilan R-Console akan tampil seperti ini:



Analisis regresi dengan menggunakan fungsi **lm()** dan untuk hasil analisisnya menggunakan fungsi **summary()**. Sintaksnya adalah sebagai berikut:

```

> library(readxl)
> Regresi_Linear <- read_excel("E:/Pelatihan
R/Regresi Linear.xlsx")
> View(Regresi_Linear)
> Tinggi_badan <- Regresi_Linear$tb #variabel bebas
(x)
> Berat_badan <- Regresi_Linear$bb #variabel tidak
bebas (y)
> model <- lm(Berat_badan~Tinggi_badan)#model
regresi linear sederhana
> summary(model)

```

Output:

```
> summary(model)
```

Call:

```
lm(formula = Berat_badan ~ Tinggi_badan)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.2413	-1.1053	0.0798	1.2527	2.7587

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	56.72250	1.83090	30.981	< 2e-16 ***
Tinggi_badan	0.08942	0.01085	8.244	1.59e-07 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.795 on 18 degrees of freedom

Multiple R-squared: 0.7906, Adjusted R-squared: 0.779

F-statistic: 67.97 on 1 and 18 DF, p-value: 1.592e-07

- a. Hasil model regresi yang ditunjukkan pada gambar output dapat ditulis:

$$\hat{y} = 0.0894X + 56.722$$

Dari persamaan model regresi tersebut dapat dikatakan jika variabel tinggi badan (x) bernilai 0 maka nilai variabel y (berat badan) yaitu 56.722 kg sebagai nilai konstanta. Koefisien regresi 0.0894 menunjukkan setiap kenaikan 1 cm pada variabel tinggi badan maka rata-rata berat badan akan naik sebesar 0.0894 kg. Jika dilihat dari kasus ini, nilai intersep tidak memiliki makna karena populasi yang dimaksud pasti memiliki tinggi badan dan tidak mungkin 0 cm, sehingga kita tidak perlu memaksakan untuk menginterpretasikan nilai intersep jika tak bermakna.

- b. Koefisien determinasi juga disebut R-Squared yang dilambangkan sebagai R^2 . Pada hasil output menunjukkan nilai R-Squared yang diperoleh yaitu 0.779 atau 77.9%. Sehingga dapat dikatakan variasi berat badan dijelaskan oleh variasi tinggi badan sebesar 77.9%, sisa 22.1% dijelaskan oleh variabel lain yang tidak dibahas dalam penelitian.
- c. Uji koefisien dilakukan untuk mengetahui apakah terdapat pengaruh signifikan variabel x terhadap variabel y.

Hipotesis nol dan alternatif adalah sebagai berikut:

$H_0: \rho = 0$ artinya tidak ada pengaruh tinggi badan terhadap berat badan

$H_1: \rho \neq 0$ artinya ada pengaruh antara tinggi badan terhadap berat badan

Hasil analisis dari uji koefisien regresi:

Dari hasil analisis diperoleh nilai p-value koefisien x sebesar 1.592e-07 lebih kecil dari alfa (α) yaitu 0.05. Karena nilai p-value < α maka tolak H_0 . Sehingga dapat disimpulkan bahwa tinggi badan berpengaruh terhadap berat badan dengan tingkat keyakinan 95%.

- d. Korelasi yang dibahas adalah korelasi *Pearson Product-Moment*. Korelasi ini adalah korelasi yang paling banyak digunakan untuk melihat ukuran korelasi antar dua variabel. Sintaks untuk menentukan nilai korelasi adalah:

```
> cor.test(Tinggi_badan,Berat_badan) #menentukan nilai korelasi
```

Output:

```
> cor.test(Tinggi_badan,Berat_badan)

Pearson's product-moment correlation

data: Tinggi_badan and Berat_badan
t = 8.2444, df = 18, p-value = 1.592e-07
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.7364087 0.9556610
sample estimates:
cor
0.8891701
```

Pada output di atas diperoleh nilai korelasi yaitu 0.889117 yang artinya variabel tinggi badan memiliki korelasi/hubungan yang kuat dengan variabel berat badan (lihat Tabel Interpretasi Korelasi)

- e. Prediksi dalam analisis regresi untuk menentukan nilai y dapat dilakukan dengan syarat nilai variabel x yang akan diprediksi masih dalam interval nilai x pada data sebelumnya. Pada kasus ini variabel tinggi badan (X) pada data yang kita miliki yaitu pada interval 112-265, oleh karena itu jika kita ingin memprediksi berat badan (Y) maka harus menggunakan nilai X antara 112-265. Misal kita ingin memprediksi nilai berat badan jika tinggi badannya sebesar 150 maka sintaksnya yaitu:

```
> predict(model, data.frame(Tinggi_badan=c(150)),
interval = "pred")
Output:
      fit      lwr      upr
1 70.13552 66.25707 74.01398
```

Hasil output menunjukkan bahwa jika tinggi badan sebesar 150 cm maka berat badan yang diperoleh yaitu 70.135 kg dengan dugaan prediksi terendah yaitu 66.257 kg dan dugaan prediksi tertinggi yaitu 74.014 kg.

B. Regresi Linear Berganda

Regresi Linear Berganda merupakan model regresi yang membentuk hubungan linear antara satu variabel tak bebas (y) dengan dua atau lebih variabel bebas ($x_1, x_2, \dots, x_n$). Model regresi linear berganda yaitu:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon \quad (9.2)$$

Dengan y merupakan variable/peubah tidak bebas (dependent variable), parameter $\beta_n, n = 0, 1, 2, \dots, n$ disebut koefisien regresi. Sedangkan x_n merupakan variabel bebas (independent variable) ke n dan ε = error pada mode atau residual. Dalam regresi linier berganda ε diasumsikan berdistribusi normal dengan rata-rata (mean) 0 dan simpangan baku σ . Asumsi lainnya adalah tidak ada korelasi antar variabel prediktor. Jika antar variable predictor terjadi korelasi maka dugaan /taksiran parameter model tidak tepat.

B.1. Penduga atau estimasi dari Parameter Model

Metode kuadrat terkecil digunakan dalam mengestimasi koefisien-koefisien regresi. Misal ada $n > k$ data pengamatan yang ada, y_i adalah pengamatan ke- i dari variable respon dan x_{ij} adalah pengamatan ke- i dari variable predictor ke- j . Dengan asumsi residual ε tidak berkorelasi dan berdistribusi normal dengan rata-rata 0 dan simpangan baku σ maka model sampel dapat ditulis sebagai berikut :

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_n x_{in} + \varepsilon_i \quad \text{atau} \quad y_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \varepsilon_i, \quad i = 1, 2, 3, \dots, n.$$

Fungsi kudrat terkecil Adalah :

$$S(\beta_0, \beta_1, \dots, \beta_k) = \sum_{i=1}^n \varepsilon_i^2$$
$$= \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij} \right)^2 \quad (9.3)$$

Selanjutnya agar diperoleh penduga kuadrat terkecil maka fungsi diturunkan terhadap $\beta_0, \beta_1, \dots, \beta_k$ dan disama dengan nol.

$$\frac{\partial S}{\partial \beta_0} \Big|_{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k} = -2 \sum_{i=1}^n \left(y_i - \hat{\beta}_0 - \sum_{j=1}^k \hat{\beta}_j x_{ij} \right) = 0 \quad (9.4)$$

Dan

$$\frac{\partial S}{\partial \beta_j} \Big|_{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k} = -2 \sum_{i=1}^n \left(y_i - \hat{\beta}_0 - \sum_{j=1}^k \hat{\beta}_j x_{ij} \right) x_{ij} = 0, j = 1, 2, \dots, k$$

Dengan menyederhanakan kedua persamaan diatas diperoleh :

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n x_{i1} + \hat{\beta}_2 \sum_{i=1}^n x_{i2} + \dots + \hat{\beta}_k \sum_{i=1}^n x_{ik} = \sum_{i=1}^n y_i$$

$$\hat{\beta}_0 \sum_{i=1}^n x_{i1} + \hat{\beta}_1 \sum_{i=1}^n x_{i1}^2 + \hat{\beta}_2 \sum_{i=1}^n x_{i1}x_{i2} + \dots + \hat{\beta}_k \sum_{i=1}^n x_{i1}x_{ik} = \sum_{i=1}^n x_{i1}y_i$$

$$\begin{matrix} \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{matrix}$$

$$\hat{\beta}_0 \sum_{i=1}^n x_{ik} + \hat{\beta}_1 \sum_{i=1}^n x_{ik}x_{i1} + \hat{\beta}_2 \sum_{i=1}^n x_{ik}x_{i2} + \dots + \hat{\beta}_k \sum_{i=1}^n x_{ik}^2 = \sum_{i=1}^n x_{ik}y_i$$

Ada $p = k + 1$ persamaan dengan masing-masing memuat koefisien regresi yang tidak dikenal. Penyelesaian/Solusi persamaan tersebut akan menjadi penduga kuadrat terkecil $\beta_0, \beta_1, \dots, \beta_k$.

Sistem persamaan diatas dapat dituliskan dalam bentuk matriks sebagai berikut :

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\text{Dengan, } \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \text{ dan } \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$\mathbf{y}$ adalah matriks $n \times 1$ atau vector kolom dari pengamatan, $\mathbf{X}$ Adalah matriks $n \times p$ dari variable predictor, $\boldsymbol{\beta}$ Adalah matriks $p \times 1$ atau vector kolom dari koefisien regresi dan $\boldsymbol{\varepsilon}$ adalah matriks $n \times 1$ atau vector kolom dari error atau residual.

$$S(\boldsymbol{\beta}) = \sum_{i=1}^n \varepsilon_i^2 = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (9.5)$$

$S(\boldsymbol{\beta})$ dapat dinyatakan pula dalam bentuk sebagai

$$\begin{aligned} S(\boldsymbol{\beta}) &= \mathbf{y}'\mathbf{y} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{y}'\mathbf{y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \end{aligned}$$

Karena $\boldsymbol{\beta}'\mathbf{X}'\mathbf{y}$ Adalah matriks 1×1 atau scalar maka berlaku $(\boldsymbol{\beta}'\mathbf{X}'\mathbf{y})' = \mathbf{y}'\mathbf{X}\boldsymbol{\beta}$ Adalah scalar yang sama.

$$\text{Penduga kudrat terkecil, } \frac{\partial S}{\partial \boldsymbol{\beta}} \Big|_{\hat{\boldsymbol{\beta}}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = 0$$

$$\text{atau } \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$$

atau $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, dengan syarat $(\mathbf{X}'\mathbf{X})^{-1}$ ada. $(\mathbf{X}'\mathbf{X})^{-1}$ ada bila predictor bebas linier atau kolom dari matriks $\mathbf{X}$ yang merupakan kombinasi linier dari kolomyang lain.

Dari persamaan $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$ dapat dituliskan dalam bentuk perkalian matriks berikut ,

$$\begin{bmatrix} n & \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i2} & \dots & \sum_{i=1}^n x_{ik} \\ \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i1}^2 & \sum_{i=1}^n x_{i1}x_{i2} & \dots & \sum_{i=1}^n x_{i1}x_{ik} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \sum_{i=1}^n x_{ik} & \sum_{i=1}^n x_{ik}x_{i1} & \sum_{i=1}^n x_{ik}x_{i2} & \dots & \sum_{i=1}^n x_{ik}^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{i1}y_i \\ \vdots \\ \sum_{i=1}^n x_{ik}y_i \end{bmatrix} \quad (9.6)$$

Contoh 9.6

Seseorang sedang menganalisis rute layanan mesin penjual otomatis (*vending machine*) dalam sistem distribusi minuman ringan. Ia tertarik untuk memprediksi jumlah waktu yang dibutuhkan oleh pengemudi untuk melayani *vending machine* disebuah gerai. Data pengamatan ditunjukkan pata table berikut.

No Pengamatan	Waktu Pengiriman y	Jumlah Kasus x_1	Jarak (kaki) x_2
1	18.68	7	560
2	11.50	3	220
3	12.03	3	340
4	14.88	4	80
5	13.75	6	150
6	18.11	7	330
7	8.00	2	110
8	17.83	7	210
9	79.24	30	1460
10	21.50	5	605
11	40.33	16	688
12	21.00	10	215
13	13.50	4	255
14	19.75	6	462
15	24.00	9	448

16	29.00	10	776
17	15.35	6	200
18	19.00	7	132
19	9.50	3	36
20	35.10	17	770
21	17.90	10	140
22	52.32	26	810
23	18.75	9	450
24	19.83	8	635
25	10.75	4	150

Jawab:

Matriks $\mathbf{X}$ dan vektor $\mathbf{y}$ adalah :

$$\mathbf{X} = \begin{bmatrix} 1 & 7 & 560 \\ 1 & 3 & 220 \\ 1 & 3 & 340 \\ 1 & 4 & 80 \\ 1 & 6 & 150 \\ 1 & 7 & 330 \\ 1 & 2 & 110 \\ 1 & 7 & 210 \\ 1 & 30 & 1460 \\ 1 & 5 & 605 \\ 1 & 16 & 688 \\ 1 & 10 & 215 \\ 1 & 4 & 255 \\ 1 & 6 & 462 \\ 1 & 9 & 448 \\ 1 & 10 & 776 \\ 1 & 6 & 200 \\ 1 & 7 & 132 \\ 1 & 3 & 36 \\ 1 & 17 & 770 \\ 1 & 10 & 140 \\ 1 & 26 & 810 \\ 1 & 9 & 450 \\ 1 & 8 & 635 \\ 1 & 4 & 150 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 16.68 \\ 11.50 \\ 12.03 \\ 14.88 \\ 13.75 \\ 18.11 \\ 8.00 \\ 17.83 \\ 79.24 \\ 21.50 \\ 40.33 \\ 21.00 \\ 13.50 \\ 19.75 \\ 24.00 \\ 29.00 \\ 15.35 \\ 19.00 \\ 9.50 \\ 35.10 \\ 17.90 \\ 52.32 \\ 18.75 \\ 19.83 \\ 10.75 \end{bmatrix}$$

$$\text{Matriks } X'X = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 7 & 3 & \dots & 4 \\ 560 & 220 & \dots & 150 \end{bmatrix} \begin{bmatrix} 1 & 7 & 560 \\ 1 & 3 & 220 \\ \vdots & \vdots & \vdots \\ 1 & 4 & 150 \end{bmatrix}$$

$$= \begin{bmatrix} 25 & 219 & 10232 \\ 219 & 3055 & 133899 \\ 10232 & 133899 & 6725688 \end{bmatrix}$$

$$\text{Dan vector } X'y = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 7 & 3 & \dots & 4 \\ 560 & 220 & \dots & 150 \end{bmatrix} \begin{bmatrix} 16.68 \\ 11.50 \\ \vdots \\ 10.75 \end{bmatrix} = \begin{bmatrix} 559.60 \\ 7375.44 \\ 337072.00 \end{bmatrix}$$

Penduga kuadrat terkecil dari β adalah : $\beta = (X'X)^{-1}X'y$ atau

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 25 & 219 & 10232 \\ 219 & 3055 & 133899 \\ 10232 & 133899 & 6725688 \end{bmatrix}^{-1} \begin{bmatrix} 559.60 \\ 7375.44 \\ 337072.00 \end{bmatrix}$$

$$= \begin{bmatrix} 0.11321518 & -0.00444859 & -0.00008367 \\ -0.00444859 & 0.00274378 & -0.00004786 \\ -0.00008367 & -0.00004786 & 0.00000123 \end{bmatrix} \begin{bmatrix} 559.60 \\ 7375.44 \\ 337072.00 \end{bmatrix}$$

$$= \begin{bmatrix} 2.34123115 \\ 1.61590712 \\ 0.01438483 \end{bmatrix}$$

Penduga kuadrat terkecil yang sesuai dengan data pelayanan minuman botol pada mesin penjual otomatis atau persamaan regresinya adalah

$$\hat{y} = 2.34123115 + 1.61590712 x_1 + 0.01438483 x_2$$

Untuk mengetahui apakah persamaan regresi tersebut merupakan model yang baik maka perlu analisis lebih lanjut apakah uji asumsi klasik terpenuhi.

Contoh 9.7

Suatu penelitian ingin mengetahui pengaruh variabel motivasi, peran dosen, fasilitas dan lingkungan keluarga terhadap pemahaman konsep siswa. Pada kasus ini diambil 261 siswa untuk dijadikan sampel. Dimisalkan variabel motivasi dengan x_1 , peran dosen dengan x_2 , fasilitas dengan x_3 , lingkungan keluarga dengan x_4 dan pemahaman konsep dengan y .

Data lengkap dalam format Excel, silakan scan QR code berikut atau kunjungi tautan: <https://bit.ly/RegresiBerganda-R>

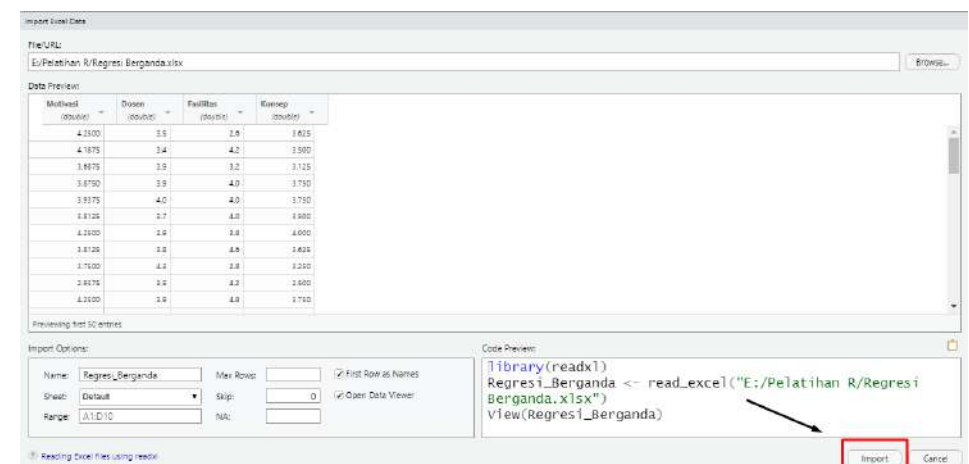


Berdasarkan data tersebut tentukan:

- Model regresi berganda.
- Koefisien determinasi
- Dengan menggunakan uji F, tentukan apakah variabel motivasi (X_1), peran dosen (X_2) dan fasilitas (X_3) secara simultan berpengaruh terhadap pemahaman konsep (Y)?
- Dengan menggunakan uji parsial, tentukan apakah variabel motivasi (X_1), peran dosen (X_2) dan fasilitas (X_3) secara parsial berpengaruh terhadap pemahaman konsep (Y)?

Jawab:

Langkah pertama adalah melakukan import data dari Excel ke R. Lakukan cara ini sesuai dengan panduan import data sebelumnya.



Setelah klik “**Import**” pada gambar di atas maka akan muncul tampilan berikut

	Motivasi	Dosen	Fasilitas	Konsep
1	4.2500	3.5	2.6	3.625
2	4.1875	3.4	4.2	3.500
3	3.6875	3.9	3.2	3.125
4	3.8750	3.9	4.0	3.750
5	3.9375	4.0	4.0	3.750

Kemudian pada tampilan R-Console akan tampil seperti gambar di bawah ini

```

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> library(readxl)
> Regresi_Berganda <- read_excel("E:/Pelatihan R/Regresi Berganda.xlsx")
> View(Regresi_Berganda)
  
```

Analisis regresi dengan menggunakan fungsi **lm()** dan untuk hasil analisisnya menggunakan fungsi **summary()**. Sintaksnya adalah sebagai berikut:

```

> library(readxl)
> Regresi_Berganda <- read_excel("E:/Pelatihan R/Regresi Berganda.xlsx")
> View(Regresi_Berganda)
> modelreg <- lm(Konsep~Motivasi+ Dosen +
Fasilitas,Regresi_Berganda)
> summary(modelreg)
  
```

Output:

```

> summary(modelreg)

Call:
lm(formula = Konsep ~ Motivasi + Dosen + Fasilitas, data = Regresi_Berganda)

Residuals:
    Min       1Q   Median       3Q      Max
-0.59366 -0.18863 -0.00878  0.18302  0.55201

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.09117   0.14191   0.642  0.52115
Motivasi     0.73240   0.03999  18.315 < 2e-16 ***
Dosen        0.13719   0.04611   2.975  0.00320 **
Fasilitas     0.08002   0.02612   3.063  0.00242 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2634 on 257 degrees of freedom
Multiple R-squared:  0.732    Adjusted R-squared:  0.7289 
F-statistic: 234 on 3 and 257 DF, p-value: < 2.2e-16
  
```

- a. Hasil model regresi berganda yang ditunjukkan pada gambar output dapat ditulis:

$$\hat{Y} = 0.09117 + 0.7324X_1 + 0.13719X_2 + 0.08002X_3$$

Dari model di atas dapat kita lihat bahwa nilai slope atau koefisien X_1 yaitu sebesar 0.7324 artinya setiap kenaikan 1 satuan skor motivasi maka pemahaman konsep akan mengalami kenaikan sebesar 0.7324 satuan jika variabel lainnya tetap. Pada model tersebut juga diperoleh koefisien X_2 sebesar 0.13719 artinya setiap kenaikan 1 satuan skor peran dosen maka nilai pemahaman konsep siswa akan naik sebesar 0.13719 jika variabel lainnya tetap. Begitu juga dengan nilai koefisien X_3 yaitu 0.08002. Selanjutnya diperoleh nilai intersep sebesar 0.09117 artinya jika variabel motivasi (X_1), peran dosen (X_2), dan fasilitas (X_3) bernilai 0 maka nilai pemahaman konsep yaitu 0.09117.

- b. Koefisien determinasi untuk regresi linear berganda sedikit berbeda dengan regresi linear sederhana. Untuk regresi linear sederhana, dilihat dari nilai *Rsquare* nya cukup, tetapi untuk regresi linear berganda yang dilihat adalah *Adjusted R Square* nya. Pada output di atas terlihat bahwa nilai dari *Adjusted R Square* adalah 0.7289 sehingga dapat dikatakan

bahwa variabel motivasi, peran dosen dan fasilitas berpengaruh terhadap pemahaman konsep sebesar 72.89% dan sisanya 27.11% dipengaruhi variabel lain yang tidak dibahas pada penelitian.

- c. Uji koefisien regresi simultan digunakan untuk mengetahui seberapa berpengaruh variabel independen terhadap variabel dependen secara bersama-sama dalam sebuah model regresi berganda. Uji statistic yang digunakan adalah uji statistic F. Pada persamaan di bawah diberikan rumus F_0 dalam menguji model

$$F_0 = \frac{R^2/k}{(1-R^2)/(n-k-1)}$$

Hipotesis yang akan diuji adalah:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_j = \beta_k = 0$$

(tidak ada pengaruh dari $X_1, X_2, \dots, X_j, \dots, X_k$ terhadap Y)

$$H_a : \beta_j \neq 0$$

(paling sedikit ada satu variabel X yang mempengaruhi Y , misalnya $X_j \rightarrow \beta_j \neq 0$)

Kriteria pengambilan keputusan adalah sebagai berikut:

- Jika nilai signifikansi (Sig.) < 0.05 maka H_0 ditolak artinya variabel motivasi (X_1), peran dosen (X_2) dan fasilitas (X_3) secara simultan berpengaruh terhadap pemahaman konsep (Y)
- Jika nilai signifikansi (Sig.) > 0.05 maka H_0 diterima artinya variabel motivasi (X_1), peran dosen (X_2) dan fasilitas (X_3) secara simultan tidak berpengaruh terhadap pemahaman konsep (Y)

Hasil Analisis Uji Koefisien Regresi Simultan:

Berdasarkan Uji-F pada output di atas diperoleh F-Statistic yaitu 234 dan p-value kurang dari $2.2e-16$. Oleh karena nilai p-value lebih kecil dari $\alpha = 5\%$ sehingga tolak H_0 dan dapat dikatakan bahwa variabel motivasi (X_1), peran dosen (X_2) dan fasilitas (X_3) secara simultan

berpengaruh terhadap pemahaman konsep (Y) atau model regresi yang diperoleh telah sesuai.

- d. Uji parsial digunakan untuk mengetahui apakah masing-masing variabel bebas yaitu motivasi (X_1), peran dosen (X_2) dan fasilitas (X_3) secara parsial berpengaruh terhadap pemahaman konsep (Y) atau tidak. Hipotesis nol dan alternatif adalah sebagai berikut:

$$H_0 : \beta_j = 0 \text{ artinya tidak ada pengaruh variabel } X_j \text{ terhadap } Y$$

$$H_1 : \beta_j \neq 0 \text{ artinya ada pengaruh antara variabel } X_j \text{ terhadap } Y$$

Kriteria pengujian uji-t parsial dapat menggunakan dasar pengambilan keputusan seperti di bawah ini:

- Jika nilai signifikansi (Sig.) < 0.05 maka variabel independen secara parsial ada hubungan berpengaruh terhadap variabel dependen
- Jika nilai signifikansi (Sig.) > 0.05 maka variabel independen secara parsial tidak ada hubungan berpengaruh terhadap variabel dependen

Hasil Analisis Uji Koefisien Regresi Parsial:

Terlihat dari output bahwa nilai p-value dari uji t nya untuk motivasi yaitu kurang dari $2e-16$, peran dosen yaitu 0.00320, dan fasilitas yaitu 0.00242. Oleh karena, seluruh nilai $\Pr(> |t|)$ lebih kecil dari $\alpha = 5\%$ sehingga tolak H_0 dan dapat dikatakan bahwa variabel motivasi (X_1), peran dosen (X_2) dan fasilitas (X_3) berpengaruh signifikan secara parsial terhadap kemampuan pemahaman konsep siswa.

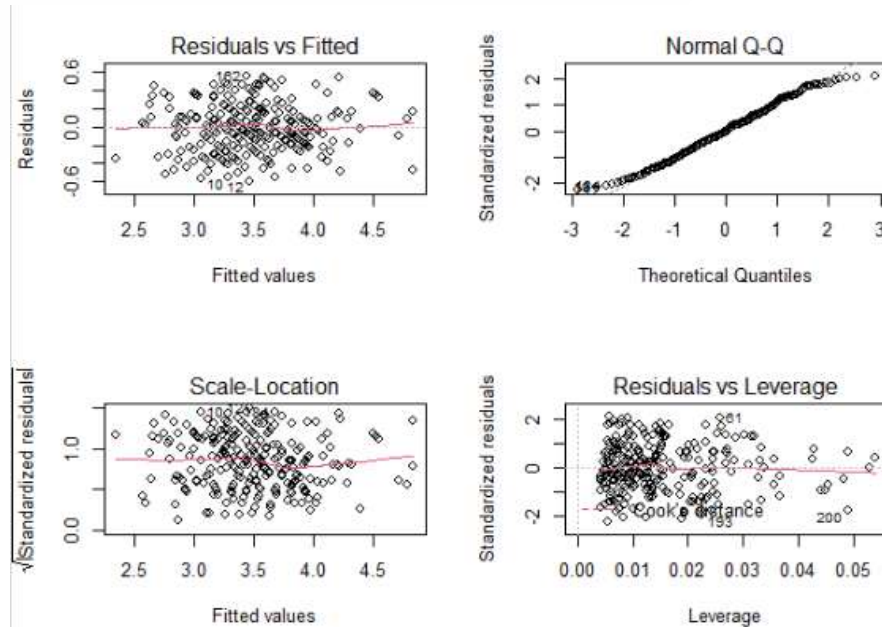
Uji Asumsi Klasik

1. Uji Normalitas

Uji Normalitas adalah untuk menguji apakah nilai residual yang dihasilkan dari regresi terdistribusi normal. Pengujian normalitas salah satunya menggunakan *normal probability plot* atau dengan melihat histogram yang membandingkan antara data observasi dengan distribusi yang mendekati distribusi normal. Sintaksnya adalah sebagai berikut:

```
> par(mfrow=c(2,2))
> plot(modelreg)
```

Output:



Interpretasi Hasil Uji Normalitas:

Pada gambar Normal QQ di atas menunjukkan bahwa titik-titik data berada disekitar garis lurus maka nilai residual yang dihasilkan regresi terdistribusi normal, sehingga dapat dikatakan model regresi memenuhi asumsi normalitas.

2. Uji Multikolinearitas

Uji multikolinearitas adalah untuk menguji apakah di dalam model regresi ditemukan adanya korelasi antar variabel bebas (independen). Model regresi yang baik seharusnya tidak memiliki korelasi antar variabel independen. Untuk mendeteksi adanya multikolinearitas adalah dengan menggunakan nilai *Tolerance* dan *Variance Inflation*

Factor (VIF). Berikut merupakan pedoman eputusan berdasarkan nilai *variance Inflation factor* (VIF).

- Jika nilai VIF < 10,00 maka tidak terjadi multikolinearitas dalam model regresi
- Jika nilai VIF > 10,00 maka terjadi multikolinearitas dalam model regresi

Sintaks untuk menguji multikolinearitas:

```
> library(car)
> vif(modelreg)
Motivasi      Dosen      Fasilitas
1.541923     1.674491     1.198591
```

Interpretasi Hasil Uji Multikolinearitas:

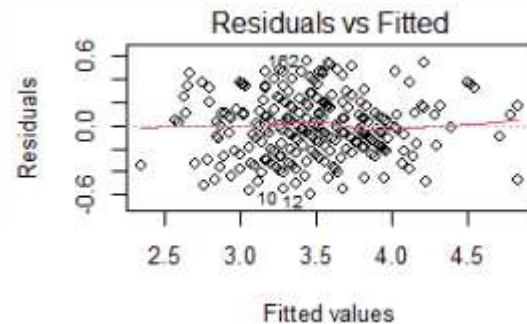
Berdasarkan Gambar di atas diperoleh nilai VIF untuk variabel Motivasi, Peran Dosen dan Fasilitas berada di bawah 10 maka tidak terjadi multikolinearitas, sehingga dapat dikatakan model regresi memenuhi asumsi multikolinearitas.

3. Uji Heterokedastisitas

Uji heteroskedastisitas adalah untuk mengetahui apakah di dalam model regresi terjadi ketidaksamaan variance dari residual satu pengamatan ke pengamatan lain. Model regresi yang bersifat homoskedastisitas dimana residual antar suatu pengamatan ke pengamatan lainnya tetap merupakan model regresi yang baik. Atau dengan kata lain regresi yang baik seharusnya tidak terjadi heteroskedastisitas. Berikut ini merupakan kriterianya:

- Jika terjadi pola tertentu, seperti titik-titik yang membentuk pola teratur (bergelombang, melebar kemudian menyempit), maka terjadi heteroskedastisitas.
- Jika tidak ada pola yang jelas, serta titik-titik tersebut menyebar di atas dan di bawah angka 0 pada sumbu Y, maka terjadi homoskedastisitas.

Untuk menguji heteroskedastisitas adalah dengan melihat gambar plot titik-titik data. Sintaks yang digunakan sama dengan sintaks pada uji normalitas.



Interpretasi Hasil Uji Heteroskedastisitas:

Berdasarkan Gambar di atas bahwa titik-titik data tersebar dan tidak membentuk pola tertentu maka tidak terjadi perbedaan varians residual, sehingga dapat dikatakan model regresi memenuhi uji asumsi heteroskedastisitas.

4. Uji Autokorelasi

Uji autokorelasi adalah untuk menguji apakah dalam model regresi linear terdapat korelasi antara kesalahan pengganggu pada periode t dengan kesalahan pengganggu pada periode $t - 1$ (sebelumnya). Pengujian adanya autokorelasi dalam suatu model regresi dilakukan dengan menggunakan Statistik d Durbin-Watson (*The Durbin-Watson d Statistics*) yang ditunjukkan pada persamaan di bawah ini:

$$d = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} \quad (9.7)$$

dimana:

d = Durbin Watson

e_t = nilai residu dari persamaan regresi periode t

e_{t-1} = nilai residu dari persamaan regresi periode $t - 1$

Berikut ini sintaks untuk menguji autokorelasi:

```
> library(lmtest)
> dwtest(modelreg)
```

Output:

```
> dwtest(modelreg)
```

Durbin-Watson test

data: modelreg

DW = 1.8796, p-value = 0.1656

alternative hypothesis: true autocorrelation is greater than 0

Interpretasi Hasil Uji Autokorelasi:

Berdasarkan Gambar di atas diperoleh p-value yaitu 0.1656, nilai p-value lebih besar dari 0.05 maka tidak terjadi pengaruh antar data terdahulu dengan data yang baru, sehingga dapat dikatakan asumsi tidak ada autokorelasi antar data pada model regresi terpenuhi.

C. Regresi Logistik

Regresi logistik adalah model matematik yang bertujuan untuk melihat hubungan beberapa variabel independent (X) dengan variabel dependen (Y) yang bersifat binary atau dikotomis (Kleinbaum & Klein, 2002). Contohnya adalah sebagai berikut:

1. Seorang dokter akan menganalisis probabilitas pasiennya terjangkit TBC dengan memprediksinya menggunakan faktor berat badan, suhu tubuh, jenis kelamin, nafsu makan, demam dan riwayat keluarga. (Misalkan pasien terjangkit TBC diberi angka 1, sedangkan yang tidak terjangkit diberi angka 0)
2. Seorang rektor akan menganalisis faktor-faktor yang mempengaruhi penentuan beasiswa PPA (mendapatkan atau tidak) dengan menggunakan faktor IPK, tingkat penghasilan orang tua, jenis kelamin,

keaktifan berorganisasi dan rata-rata SKS. (Misalkan mahasiswa yang mendapat beasiswa diberi angka 1, sedangkan yang tidak mendapatkan beasiswa diberi angka 0)

Regresi logistik tidak perlu mensyaratkan asumsi berdistribusi normal pada variabel bebasnya & dapat digunakan apabila asumsi multivariat normal distribusi tidak terpenuhi. Variabel respon pada regresi linear diasumsikan berdistribusi normal, sedangkan pada regresi logistic diasumsikan berdistribusi bernoulli. Regresi logistik dapat dikelompokkan menjadi dua jenis, yaitu:

1. Regresi Logistik Biner

Digunakan pada kasus yang memiliki dua kategori pada variabel dependent (Y), contohnya mendapatkan beasiswa atau tidak mendapatkan, untung atau rugi, terjangkit covid-19 atau tidak.

2. Regresi Logistik Multinomial

Digunakan pada kasus yang memiliki lebih dari dua kategori pada variabel dependent (Y). Model regresi logistik yang digunakan yaitu:

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 X)}{1 + \exp(\beta_0 + \beta_1 X)} \quad (9.8)$$

Transformasi dari $\pi(x)$ yang menjadi fokus pada regresi logistic adalah transformasi logit. Transformasi tersebut didefinisikan dalam bentuk $\pi(x)$, sebagai:

$$\ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_1 X + \dots + \beta_p X_p \quad (9.9)$$

Catatan:

Exp atau ditulis “e” merupakan fungsi eksponen yang merupakan invers dari logaritma natural. Nilai e yaitu nilai konstanta 2,71828182845904 atau dibulatkan menjadi 2,72.

Contoh 9.8

Suatu penelitian ingin mengetahui faktor-faktor yang mempengaruhi diterimanya suatu beasiswa. Terdapat 3 variabel bebas yang dikaji yaitu

variabel skor toefl (X_1), IPK (X_2) dan ranking asal Universitas (X_3). Sedangkan variabel tidak bebas yaitu variabel beasiswa (Y) yang terdiri atas dua kategori/biner, yaitu diterima (1) dan tidak diterima (0). Pada kasus ini diambil sampel secara acak sejumlah 400 pendaftar & menggunakan tingkat signifikansi 5%.

Berdasarkan data tersebut tentukan:

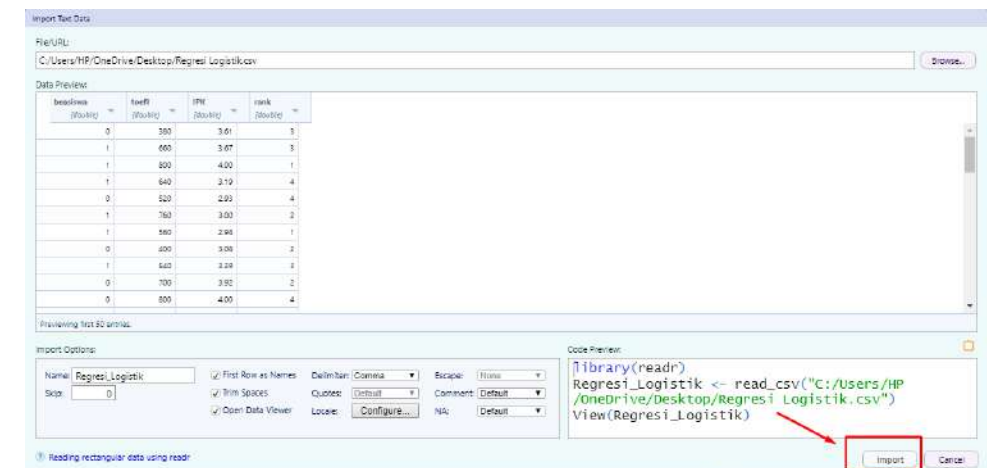
Data lengkap dalam format Excel, silakan scan QR code berikut atau kunjungi tautan: <https://bit.ly/RegresiLogistik-R>



- Model regresi logistik
- Dengan menggunakan uji hipotesis, apakah terdapat pengaruh variabel bebas terhadap variabel tidak bebas?
- Dengan menggunakan uji *Goodness of Fit*, tentukan kecocokan model regresi logistik yang diperoleh.
- Tentukan nilai Rasio Kecenderungan (*Odds Ratio*) dari model regresi logistik yang dihasilkan. Berikan interpretasi dari nilai tersebut.

Jawab:

Langkah pertama adalah melakukan import data dari Excel ke R. Lakukan cara ini sesuai dengan panduan import data.



Setelah klik “**Import**” pada gambar di atas maka akan muncul tampilan berikut

	beasiswa	toefl	IPK	rank
1	0	380	3.61	3
2	1	660	3.67	3
3	1	800	4.00	1
4	1	640	3.19	4
5	0	520	2.93	4
6	1	760	3.00	2
7	1	560	2.98	1

Kemudian pada tampilan R-Console akan tampil seperti gambar di bawah ini

```
> library(readr)
> Regresi_Logistik <- read_csv("C:/Users/HP/OneDrive/Desktop/Regresi Logistik.csv")

-- Column specification -----
cols(
  beasiswa = col_double(),
  toefl = col_double(),
  IPK = col_double(),
  rank = col_double()
)
> View(Regresi_Logistik)
```

Analisis regresi logistic dengan menggunakan fungsi **glm()** dan untuk hasil analisisnya menggunakan fungsi **summary()**. Sintaksnya adalah sebagai berikut:

```
> library(readr)
> Regresi_Logistik <- read_csv("C:/Users/HP/OneDrive/Desktop/Regresi Logistik.csv")
> View(Regresi_Logistik)
> str(Regresi_Logistik)
> Beasiswa <- Regresi_Logistik$beasiswa
> Toefl <- Regresi_Logistik$toefl
> IPK <- Regresi_Logistik$IPK
> rank <- factor(Regresi_Logistik$rank) #mengubah ke tipe faktor
> model_regresi_logistik <- glm(Beasiswa~ Toefl+IPK+rank, family = "binomial")
> print(model_regresi_logistik)
> summary(model_regresi_logistik)
```

Output **model_regresi_logistik**:

```
> model_regresi_logistik

Call: glm(formula = Beasiswa ~ Toefl + IPK + rank, family = "binomial")

Coefficients:
(Intercept)  Toefl      IPK      rank2      rank3      rank4
-3.989979    0.002264    0.804038   -0.675443   -1.340204   -1.551464

Degrees of Freedom: 399 Total (i.e. Null); 394 Residual
Null Deviance: 500
Residual Deviance: 458.5 AIC: 470.5
```

Output **summary()**:

```
> summary(model_regresi_logistik)

Call:
glm(formula = Beasiswa ~ Toefl + IPK + rank, family = "binomial")

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6268  -0.8662  -0.6388   1.1490   2.0790

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.989979   1.139951  -3.500  0.000465 ***
Toefl        0.002264   0.001094   2.070  0.038465 *
IPK          0.804038   0.331819   2.423  0.015388 *
rank2       -0.675443   0.316490  -2.134  0.032829 *
rank3       -1.340204   0.345306  -3.881  0.000104 ***
rank4       -1.551464   0.417832  -3.713  0.000205 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 499.98  on 399  degrees of freedom
Residual deviance: 458.52  on 394  degrees of freedom
AIC: 470.52
```

a. Hasil model regresi logistik setelah ditransformasi logit, persamaannya menjadi:

$$\hat{g}(D) = -3.89 + 0.0022 \text{ Toefl} + 0.8040 \text{ IPK} - 0.67 \text{ Rank2} - 1.34 \text{ Rank3} - 1.55 \text{ Rank4}$$

b. Perintah **summary** merupakan ringkasan terkait model yang digunakan & berisi koefisien logistik, Wald Z-Statistiknya dan nilai p-value. Dengan menggunakan uji hipotesis sebagai berikut:

H0: Tidak terdapat pengaruh variabel bebas terhadap variabel tidak bebas

H1: Terdapat pengaruh variabel bebas terhadap variabel tidak bebas

Pada output `summary()` diperoleh semua p-value kurang dari tingkat signifikansi 5% maka tolak H0, sehingga dapat dikatakan bahwa terdapat pengaruh variabel bebas terhadap variabel tidak bebas.

- c. Dalam menentukan kecocokan model dengan Hosmer Lemeshow Test (*Goodness of Fit Test*). Jika nilai p-value lebih besar dari nilai taraf signifikansi misal 5% menunjukkan kecocokan model yang baik & sebaliknya. Berikut merupakan sintaksnya:

```
> library(ResourceSelection)
> hosmer_test <- hoslem.test(Beasiswa,
fitted(model_regresi_logistik), g=10)
> print(hosmer_test)
```

Output:

```
> print(hosmer_test)
```

Hosmer and Lemeshow goodness of fit (GOF) test

```
data: Beasiswa, fitted(model_regresi_logistik)
X-squared = 11.085, df = 8, p-value = 0.1969
```

Dari hasil diatas dapat dilihat dari nilai p-value sebesar 0.1969 yang menunjukkan nilai p-value lebih dari 5% sehingga dapat disimpulkan bahwa model regresi logistik yang terbentuk sudah sesuai.

- d. Berikut sintaks untuk menentukan nilai *Odds Ratio*:

```
> exp(coef(model_regresi_logistik))

(Intercept)  Toefl      IPK      rank2      rank3      rank4
0.01850     1.00226    2.23454    0.50893    0.26179    0.21193
```

Dari hasil di atas diperoleh bahwa siswa dengan nilai Toefl dan IPK lebih tinggi dari variabel lainnya maka memiliki kecenderungan lebih besar untuk diterima beasiswa. Siswa dengan IPK lebih tinggi 1 unit maka akan 2.2 kali lebih tinggi peluang diterima beasiswa tersebut.

Selanjutnya jika kita lihat rank, yang menjadi dasar (baseline) adalah peringkat (ranking) 1. Sehingga pada *Odds Ratio* pada Rank 2 adalah 0.5. Hal tersebut menunjukkan bahwa siswa dengan rank 2 memiliki peluang 0.5 kali untuk diterima beasiswa dibandingkan siswa dengan rank 1. Dengan kata lain siswa dengan rank 1 pada asal universitasnya memiliki peluang 2 kali lipat lebih besar dibandingkan dengan siswa dengan rank 2 dan seterusnya.

LATIHAN 9

Analisis Korelasi dan Regresi

1. Diketahui data berpasangan seperti berikut :

x	1	2	3	4	5	6
y	5.6	4.6	4.5	3.7	3.2	2.7

Tentukan :

- Persamaan regresi untuk data tersebut
 - Berapakah nilai y bila $x = 3.5$
2. Nilai dua variabel yang diukur untuk sembilan barang berbeda ditunjukkan dalam tabel.

x	6	3	5	8	7	5	3	6	9
y	31	35	52	22	26	26	41	42	31

- Buatlah diagram pencar (*scatter plot*) dari data di atas.
 - Hitung koefisien korelasi *Pearson Product Momen*.
 - Tentukan persamaan regresi linear dari data yang diberikan.
3. Berikut ini ditampilkan data dari *Motor Trend US* yang terdiri dari data **konsumsi bahan bakar mobil** dalam satuan miles/gallon (mpg) dan **berat mobil** (1000 lbs).

Data lengkap dalam format Excel, silakan scan QR code berikut atau kunjungi tautan: <https://bit.ly/DataRegresiMobil>



Dengan menggunakan data tersebut:

- Apakah berat mobil berpengaruh secara signifikan terhadap konsumsi bahan bakar? (**Catatan:** rumuskan H_0 dan H_1 dan berikan kesimpulan anda dengan jelas)
- Buat persamaan regresi linearnya
- Hitung koefisien korelasi pearson (r) dan **interpretasikan** hasilnya

- Prediksilah konsumsi bahan bakar jika berat mobil 2150 (dalam 1000 lbs).

4. Seorang peneliti ingin memprediksi kandungan karbon monoksida yang dihasilkan oleh rokok jika diketahui kandungan nikotinnya. Sampel diambil dari beberapa merk rokok. Hasil pengamatan dinyatakan dalam tabel berikut.

Rokok	Karbon Monoksida (mg)	Nikotin (mg)	Rokok	Karbon Monoksida (mg)	Nikotin (mg)
1	13.6	0.86	14	10.2	0.78
2	16.6	1.06	15	9.5	0.74
3	23.5	2.03	16	1.5	0.13
4	10.2	0.67	17	18.5	1.26
5	5.4	0.40	18	12.6	1.08
6	15.0	1.04	19	17.5	0.96
7	9.0	0.76	20	4.9	0.42
8	12.3	0.95	21	15.9	1.01
9	16.3	1.12	22	8.5	0.61
10	15.4	1.02	23	10.6	0.69
11	13.0	1.01	24	13.9	1.02
12	14.4	0.90	25	14.9	0.82
13	10.0	0.57			

- Tentukan persamaan regresi untuk data tersebut
 - Berapa mg karbon monoksida yang dihasilkan bila kandungan nikotin dalam rokok adalah 14,5 mg?
5. Untuk mempelajari keuntungan investasi beberapa perusahaan, diperoleh data seperti pada tabel dibawah ini. Data dicatat dalam satuan M (milyar) rupiah dan dikumpulkan dari 10 perusahaan pada tahun yang sama. Variabel y mewakili laba tahun berjalan, x_1 mewakili modal investasi, dan x_2 mewakili pengeluaran dalam bentuk iklan.

y	x_1	x_2
15	25	4
16	1	5
2	6	3

3	30	1
12	29	2
1	20	0
16	12	4
18	15	5
13	6	4
2	16	2

- Dengan metode kuadrat terkecil tentukan persamaan regresi untuk data tersebut.
 - Berapakah laba yang diperoleh bila modal investasi 10 M rupiah dan pengeluaran untuk iklan 2,5 M rupiah.
6. Data dalam tabel berikut merupakan skor rasa dan tekstur (antara 0 dan 100) untuk 12 merek hamburger beserta harga, jumlah kalori, jumlah lemak, dan jumlah natrium per burger.

Merk	Skor	Harga (x_1)	Kalori (x_2)	Lemak (x_3)	Natrium (x_4)
1	70	91	110	4	310
2	45	68	90	0	420
3	43	92	80	1	280
4	41	75	120	5	370
5	39	88	90	0	410
6	30	67	140	4	440
7	68	73	120	4	430
8	56	92	170	6	520
9	40	71	130	4	180
10	34	67	110	2	180
11	30	92	100	1	330
12	26	95	130	2	340

7. Sebuah penelitian ingin mengetahui faktor-faktor yang mempengaruhi mahasiswa lulus ujian sertifikasi profesional (Y). Variabel respon Y bersifat biner: **lulus = 1 dan tidak lulus = 0**. Terdapat tiga variabel prediktor yang dianalisis:

X_1 = Jumlah jam belajar persiapan (jam),

X_2 = Persentase kehadiran kelas (%),

X_3 = Skor praktek akhir (0–100).

Diambil sampel acak sebanyak $n = 500$ mahasiswa. Analisis regresi logistik biner menghasilkan ringkasan parameter sebagai berikut (nilai koefisien sudah diestimasi):

Parameter	Koefisien (B)	Std. Error	z	p-value
Intercept	-6.200	0.920	-6.74	< 0.001
X_1 (Jam belajar)	0.085	0.015	5.67	< 0.001
X_2 (Kehadiran %)	0.030	0.008	3.75	< 0.001
X_3 (Skor praktek)	0.045	0.012	3.75	< 0.001

Gunakan tingkat signifikansi $\alpha = 0.05$. Berdasarkan output tersebut, tentukan:

- Model regresi logistik (tuliskan persamaan logit dan persamaan probabilitas).
- Uji hipotesis apakah masing-masing variabel bebas berpengaruh signifikan terhadap probabilitas lulus. Nyatakan hipotesis nol & alternatif, statistik uji yang digunakan, dan keputusan untuk tiap variabel.
- Uji *Goodness-of-Fit* sederhana: jelaskan pendekatan yang dapat digunakan dan bagaimana kamu menilai kecocokan model berdasarkan output tambahan berikut (misalkan nilai Hosmer–Lemeshow $\chi^2 = 8.2$ dengan $df = 8$, $p = 0.413$). Apa kesimpulannya?
- Hitung *Odds Ratio* (OR) untuk tiap variabel (dari koefisien B) dan berikan interpretasi praktis untuk masing-masing OR.

BAB 10

Peta Kendali Untuk Variabel

Pendahuluan

Kemampuan Akhir yang Diharapkan	Mahasiswa memahami masalah yang berkaitan dengan Peta Kendali $\bar{x}$ dan R , Peta Kendali $\bar{x}$ dan s , serta Peta Kendali Individual (<i>I-Chart</i>) dan Moving Range (<i>MR-Chart</i>).
Bahan Kajian	- Peta Kendali $\bar{x}$ dan R - Peta Kendali $\bar{x}$ dan s - Peta Kendali Individual (<i>I-Chart</i>) dan Moving Range (<i>MR-Chart</i>)
Metode Pembelajaran	Ceramah interaktif, studi kasus, <i>collaborative learning</i>
Pengalaman Belajar Mahasiswa	- Mengolah data hasil proses produksi atau simulasi untuk membuat berbagai jenis peta kendali ($\bar{x}-R$, $\bar{x}-s$, dan $I-MR$) - Menganalisis dan menginterpretasikan hasil peta kendali untuk menentukan kestabilan dan kemampuan proses.
Kriteria Penilaian (Indikator)	- Ketepatan perhitungan dan penyajian peta kendali (nilai rata-rata, simpangan baku, rentang, serta batas kendali atas dan bawah). - Ketepatan interpretasi kondisi proses (dalam kendali atau tidak) dan kemampuan menjelaskan penyebab variasi proses.

Tujuan Pembelajaran

1. Mahasiswa memahami masalah yang berkaitan dengan Peta Kendali $\bar{x}$ dan R
2. Mahasiswa memahami masalah yang berkaitan dengan Peta Kendali $\bar{x}$ dan s
3. Mahasiswa memahami masalah yang berkaitan dengan Peta Kendali Individual (*I-Chart*) dan Moving Range (*MR-Chart*).

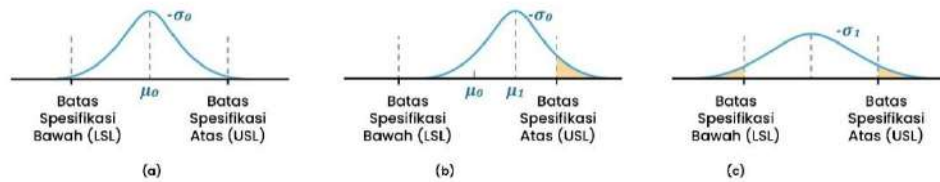
A. Pendahuluan

Banyak karakteristik kualitas dapat dinyatakan dalam bentuk ukuran numerik. Misalnya, diameter suatu bantalan dapat diukur dengan menggunakan mikrometer dan dinyatakan dalam milimeter. Karakteristik kualitas yang dapat diukur, seperti dimensi, berat, atau volume, disebut **variabel**. Peta kendali untuk variabel banyak digunakan dalam pengendalian kualitas, dan merupakan salah satu alat utama dalam tahap analisis dan pengendalian. Ketika menangani karakteristik kualitas berupa variabel, penting untuk memantau baik rata-rata nilai karakteristik kualitas maupun variabilitasnya. Pengendalian rata-rata proses atau tingkat kualitas rata-rata biasanya dilakukan menggunakan **peta kendali rata-rata (peta kendali $\bar{x}$)**.

Variabilitas suatu proses dapat dipantau menggunakan **peta kendali simpangan baku (peta kendali s)** atau **peta kendali rentang (peta kendali R)**. Biasanya, peta $\bar{x}$ dan R disusun secara terpisah untuk setiap karakteristik kualitas yang diamanati. Peta $\bar{x}$ dan R (atau s) termasuk di antara teknik pemantauan dan pengendalian proses statistik yang paling penting dan bermanfaat.

Menjaga kestabilan baik rata-rata proses maupun variabilitas proses sangatlah penting. Gambar 10.1 menggambarkan output dari suatu proses produksi. Pada Gambar 10.1a, baik rata-rata proses (μ) maupun simpangan baku (σ) berada dalam kondisi terkendali pada nilai nominalnya (misal μ_0 dan σ_0); sehingga sebagian besar output proses tetap berada dalam batas spesifikasi. Namun, pada Gambar 10.1b, rata-rata proses bergeser dari target ke nilai $\mu_1 > \mu_0$ yang menyebabkan meningkatnya proporsi produk cacat. Sedangkan pada Gambar 10.1c, simpangan baku meningkat menjadi $\sigma_1 >$

σ_0 atau variabilitas proses meningkat. Hal ini mengakibatkan peningkatan jumlah produk gagal, meskipun nilai rata-rata proses masih berada pada nilai nominalnya.



Gambar 10.1. Perbandingan Perubahan Rata-Rata dan Variabilitas Proses terhadap Batas Spesifikasi

Keterangan:

- Batas Spesifikasi Bawah atau *Lower Specification Limit* (LSL)
- Batas Spesifikasi Atas atau *Upper Specification Limit* (USL)
- μ_0 = rata-rata proses awal
- μ_1 = rata-rata proses baru
- σ_0 = simpangan baku awal
- σ_1 = simpangan baku baru

B. Peta Kendali $\bar{x}$ dan R

Misalkan suatu karakteristik kualitas terdistribusi normal dengan rata-rata μ dan simpangan baku σ , di mana baik μ maupun σ diketahui. Jika $x_1, x_2, \dots, x_n$ adalah sampel berukuran n , maka rata-rata dari sampel ini adalah

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} \quad (10.1)$$

dan kita tahu bahwa $\bar{x}$ terdistribusi normal dengan rata-rata μ dan simpangan baku $\sigma_{\bar{x}} = \sigma/\sqrt{n}$. Lebih lanjut, probabilitas bahwa nilai rata-rata sampel berada di antara:

$$\mu + z_{\alpha/2} \sigma_{\bar{x}} = \mu + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \text{ dan } \mu - z_{\alpha/2} \sigma_{\bar{x}} = \mu - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad (10.2)$$

adalah sebesar $1 - \alpha$.

Oleh karena itu, jika μ dan σ telah diketahui, persamaan (10.2) dapat digunakan untuk menentukan batas kendali atas dan bawah pada grafik pengendalian rata-rata sampel. Jika rata-rata suatu sampel berada di luar batas kendali tersebut, hal ini menunjukkan bahwa nilai rata-rata proses kemungkinan sudah tidak sama dengan μ . Dalam pembahasan ini diasumsikan bahwa distribusi karakteristik kualitas bersifat normal. Namun demikian, hasil perhitungan tetap cukup akurat meskipun distribusi sebenarnya tidak normal, karena adanya Teorema Limit Tengah. Dalam praktiknya, nilai μ dan σ biasanya tidak diketahui, sehingga perlu diestimasi dari data awal atau subkelompok yang dikumpulkan ketika proses dianggap berada dalam kondisi terkendali. Estimasi ini idealnya didasarkan pada 20 hingga 25 sampel. Misalkan tersedia m sampel, masing-masing terdiri dari n pengamatan terhadap karakteristik kualitas. Umumnya, ukuran sampel n relatif kecil sekitar 4, 5, atau 6. Ukuran sampel kecil ini biasanya disebabkan oleh konstruksi subkelompok rasional serta fakta bahwa biaya sampling dan inspeksi yang terkait dengan pengukuran variabel biasanya relatif besar. Misalkan $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m$ adalah rata-rata dari masing-masing sampel. Maka, estimasi terbaik dari μ , rata-rata proses, adalah rata-rata gabungan yaitu:

$$\bar{\bar{x}} = \frac{(\bar{x}_1 + \bar{x}_2 + \dots + \bar{x}_m)}{m} \quad (10.3)$$

Dengan demikian, $\bar{\bar{x}}$ akan digunakan sebagai garis tengah pada diagram $\bar{x}$. Untuk membangun batas kendali, kita membutuhkan estimasi dari simpangan baku σ . Jika $x_1, x_2, \dots, x_n$ adalah sampel berukuran n , maka rentang sampel adalah selisih antara nilai terbesar dan terkecil dari pengamatan yaitu:

$$R = x_{\max} - x_{\min}$$

Misalkan $R_1, R_2, \dots, R_m$ adalah rentang dari m sampel. Maka rentang rata-ratanya adalah:

$$\bar{R} = \frac{R_1 + R_2 + \dots + R_m}{m} \quad (10.4)$$

Batas kendali atas (*Upper Control Limit/UCL*) dan batas kendali bawah (*Lower Control Limit/LCL*) untuk diagram $\bar{X}$ dan $\bar{R}$ dapat dihitung dengan rumus:

Batas Kendali untuk Diagram $\bar{X}$

$$\begin{aligned} \text{UCL} &= \bar{\bar{X}} + A_2 \bar{R} \\ \text{Garis tengah} &= \bar{\bar{X}} \\ \text{LCL} &= \bar{\bar{X}} - A_2 \bar{R} \end{aligned} \quad (10.5)$$

Batas Kendali untuk Diagram R

$$\begin{aligned} \text{UCL} &= D_4 \bar{R} \\ \text{Garis tengah} &= \bar{R} \\ \text{LCL} &= D_3 \bar{R} \end{aligned} \quad (10.6)$$

Variabel acak $W = R/\sigma$ disebut sebagai **rentang relatif**. Parameter dari distribusi W merupakan fungsi dari ukuran sampel n . Rata-rata dari W adalah d_2 . Sehingga penduga untuk σ adalah $\hat{\sigma} = R/d_2$. Nilai dari d_2 untuk berbagai ukuran sampel tersedia di Tabel Lampiran VI. $\bar{R}$ adalah rentang rata-rata dari m sampel awal, maka σ dapat diestimasi menggunakan rumus:

$$\hat{\sigma} = \frac{\bar{R}}{d_2} \quad (10.7)$$

Penduga ini merupakan estimasi tak bias (*unbiased estimator*) dari σ , sehingga memberikan perkiraan yang baik terhadap simpangan baku proses sebenarnya.

Jika kita menggunakan $\bar{\bar{X}}$ sebagai estimator dari μ dan $\bar{R}/d_2$ sebagai estimator dari σ , maka parameter dari diagram $\bar{X}$ adalah

$$\text{UCL} = \bar{\bar{X}} + \frac{3}{d_2 \sqrt{n}} \bar{R}$$

$$\text{Garis Tengah} = \bar{\bar{X}} \quad (10.8)$$

$$\text{LCL} = \bar{\bar{X}} - \frac{3}{d_2 \sqrt{n}} \bar{R}$$

Jika kita mendefinisikan

$$A_2 = \frac{3}{d_2 \sqrt{n}} \quad (10.9)$$

maka persamaan (10.8) menjadi sama dengan persamaan (10.5).

Sekarang perhatikan diagram R dengan garis tengah $\bar{R}$. Untuk menentukan batas kendali, diperlukan estimasi simpangan baku dari R yaitu σ_R . Dengan asumsi bahwa karakteristik kualitas berdistribusi normal, σ_R dapat diperoleh dari distribusi rentang relatif $W = R/\sigma$. Simpangan baku dari W , misalnya d_3 , merupakan fungsi yang bergantung pada ukuran sampel n . Maka diperoleh:

$$R = W\sigma$$

Simpangan baku dari R adalah:

$$\sigma_R = d_3 \sigma$$

Karena σ tidak diketahui, kita dapat memperkirakan σ_R dengan

$$\hat{\sigma}_R = d_3 \frac{\bar{R}}{d_2} \quad (10.10)$$

Oleh karena itu, parameter dari diagram R dengan batas kendali tiga-sigma yang biasa digunakan adalah

$$\text{UCL} = \bar{R} + 3\hat{\sigma}_R = \bar{R} + 3d_3 \frac{\bar{R}}{d_2}$$

$$\text{Garis Tengah} = \bar{R} \quad (10.11)$$

$$\text{LCL} = \bar{R} - 3\hat{\sigma}_R = \bar{R} - 3d_3 \frac{\bar{R}}{d_2}$$

Jika kita menetapkan

$$D_3 = 1 - 3 \frac{d_3}{d_2} \quad \text{dan} \quad D_4 = 1 + 3 \frac{d_3}{d_2}$$

maka persamaan (10.11) dapat disederhanakan menjadi persamaan (10.6).

Contoh 10.1

Sebuah proses *hard-bake* digunakan bersamaan dengan teknik fotolitografi dalam proses pembuatan semikonduktor. Kita ingin menetapkan pengendalian statistik terhadap lebar aliran (*flow width*) dari *resist* pada proses ini menggunakan diagram $\bar{X}$ dan R . Sebanyak dua puluh lima sampel, masing-masing terdiri dari lima *wafer*, diambil pada saat proses dianggap berada dalam kondisi terkendali. Jarak waktu antar sampel atau subkelompok adalah satu jam. Data hasil pengukuran lebar aliran (dalam mikron) dari sampel-sampel tersebut ditampilkan pada tabel berikut. Buatlah peta kendali $\bar{X}$ dan R .

Sampel	Wafer 1	Wafer 2	Wafer 3	Wafer 4	Wafer 5
1	13,235	14,128	16,744	14,573	16,914
2	14,314	13,592	16,075	14,666	16,109
3	14,284	14,871	14,932	14,324	15,674
4	15,028	16,352	13,841	12,831	15,507
5	15,604	12,735	15,265	14,363	16,441
6	15,955	15,451	13,574	13,281	14,198
7	16,274	15,064	18,366	14,177	15,144
8	14,190	14,303	16,637	16,067	15,519
9	13,884	17,277	15,355	15,176	13,688
10	14,039	16,697	15,089	14,627	15,220
11	14,158	17,667	14,278	15,928	14,181
12	15,821	13,355	15,777	13,908	17,559
13	12,856	14,106	14,447	16,398	11,928
14	14,951	14,036	15,893	16,458	14,969
15	13,589	12,863	15,996	12,497	15,471
16	15,747	15,301	15,171	11,839	18,662
17	13,680	17,269	13,957	15,014	14,449

18	14,163	13,864	13,057	16,210	15,573
19	15,796	14,185	16,541	15,116	17,247
20	17,106	14,421	12,361	13,820	17,601
21	14,371	15,051	13,485	15,670	14,880
22	14,738	15,936	16,583	14,973	14,720
23	15,917	14,333	15,551	15,295	16,866
24	16,399	15,243	15,705	15,563	15,530
25	15,797	13,663	16,240	13,732	16,887

Jawab:

Saat membuat peta kendali $\bar{X}$ dan R , sebaiknya dimulai dengan peta R . Hal ini karena batas kendali pada peta $\bar{X}$ bergantung pada variasi proses. Jika variasi proses belum terkendali, maka batas kendali tersebut tidak akan memiliki makna yang signifikan. Dengan menggunakan data di atas kita bisa menghitung rata-rata $\bar{X}$ dan range R :

Sampel	Wafer 1	Wafer 2	Wafer 3	Wafer 4	Wafer 5	$\bar{x}_i$	R_i
1	13,235	14,128	16,744	14,573	16,914	15,119	0,3679
2	14,314	13,592	16,075	14,666	16,109	14,951	0,2517
3	14,284	14,871	14,932	14,324	15,674	14,817	0,1390
4	15,028	16,352	13,841	12,831	15,507	14,712	0,3521
5	15,604	12,735	15,265	14,363	16,441	14,882	0,3706
6	15,955	15,451	13,574	13,281	14,198	14,492	0,2674
7	16,274	15,064	18,366	14,177	15,144	15,805	0,4189
8	14,190	14,303	16,637	16,067	15,519	15,343	0,2447
9	13,884	17,277	15,355	15,176	13,688	15,076	0,3589
10	14,039	16,697	15,089	14,627	15,220	15,134	0,2658
11	14,158	17,667	14,278	15,928	14,181	15,242	0,3509
12	15,821	13,355	15,777	13,908	17,559	15,284	0,4204
13	12,856	14,106	14,447	16,398	11,928	13,947	0,4470
14	14,951	14,036	15,893	16,458	14,969	15,261	0,2422
15	13,589	12,863	15,996	12,497	15,471	14,083	0,3499
16	15,747	15,301	15,171	11,839	18,662	15,344	0,6823
17	13,680	17,269	13,957	15,014	14,449	14,874	0,3589

18	14,163	13,864	13,057	16,210	15,573	14,573	0,3153
19	15,796	14,185	16,541	15,116	17,247	15,777	0,3062
20	17,106	14,421	12,361	13,820	17,601	15,060	0,5240
21	14,371	15,051	13,485	15,670	14,880	14,691	0,2185
22	14,738	15,936	16,583	14,973	14,720	15,390	0,1863
23	15,917	14,333	15,551	15,295	16,866	15,592	0,2533
24	16,399	15,243	15,705	15,563	15,530	15,688	0,1156
25	15,797	13,663	16,240	13,732	16,887	15,264	0,3224
							$\sum \bar{x}_i$ = 37,64
							$\sum R_i$ = 8,13

Hitung garis tengah untuk peta R yaitu:

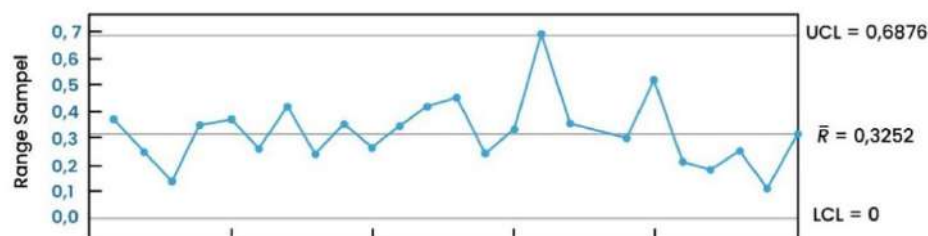
$$\bar{R} = \frac{\sum_{i=1}^{25} R_i}{25} = \frac{8,13}{25} = 0,32521$$

Ukuran pengamatan yaitu 5 subgrup, berdasarkan Tabel Lampiran VI diketahui bahwa $D_3 = 0$ dan $D_4 = 2,114$. Selanjutnya kita hitung UCL dan LCL:

$$UCL = \bar{R}D_4 = 0,32521(2,114) = 0,68749$$

$$LCL = \bar{R}D_3 = 0,32521(0) = 0$$

Peta kendali R :



Pada peta kendali R ditunjukkan tidak ada indikasi bahwa proses berada dalam kondisi tidak terkendali karena variabilitas proses masih dalam batas

kendali. Selanjutnya untuk membuat diagram $\bar{X}$. Garis tengah (*center line*) dihitung sebagai:

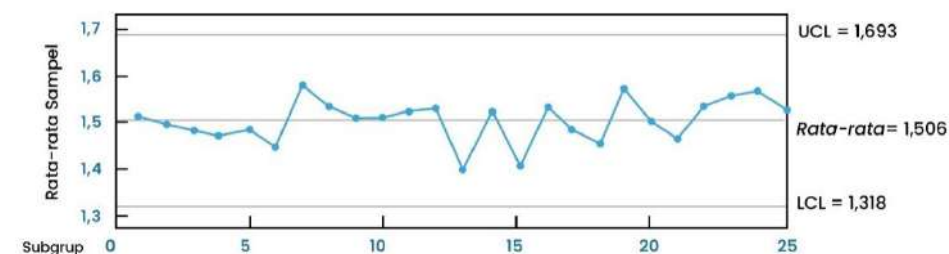
$$\bar{\bar{X}} = \frac{\sum_{i=1}^{25} \bar{X}_i}{25} = \frac{37,64}{25} = 1,5056$$

Berdasarkan Tabel Lampiran VI diketahui $A_2 = 0.577$ sehingga:

$$UCL = \bar{\bar{X}} + A_2\bar{R} = 1,5056 + (0,577)(0,325) = 1,693$$

$$LCL = \bar{\bar{X}} - A_2\bar{R} = 1,5056 - (0,577)(0,325) = 1,318$$

Peta kendali $\bar{X}$:



Pada gambar peta kendali $\bar{X}$ tidak ditemukan tanda bahwa proses berada di luar batas kendali. Maka dapat disimpulkan berdasarkan diagram $\bar{X}$ maupun R , proses berada dalam kendali pada tingkat yang ditetapkan.

C. Estimasi Kapabilitas Proses (C_p)

Grafik $\bar{X}$ dan R memberikan gambaran mengenai kinerja atau kemampuan (kapabilitas) suatu proses. Dari diagram $\bar{X}$, kita dapat memperkirakan lebar aliran rata-rata *resist* pada proses *hard-bake*, yaitu sebesar **1,5056 mikron**. Sementara itu, simpangan baku proses dapat diperkirakan menggunakan persamaan (10.6), yaitu:

$$\hat{\sigma} = \frac{\bar{R}}{d_2} = \frac{0,32521}{2,326} = 0,1398$$

di mana nilai d_2 untuk sampel berukuran lima terdapat pada Lampiran VI. Batas spesifikasi untuk lebar aliran ditetapkan sebesar **1,50 ± 0,50 mikron**. Data dari diagram kendali dapat digunakan untuk menilai sejauh mana proses mampu menghasilkan wafer yang sesuai dengan spesifikasi

tersebut. Dengan asumsi bahwa lebar aliran berdistribusi normal, memiliki rata-rata **1,5056** dan simpangan baku **0,1398**, kita dapat memperkirakan proporsi wafer yang tidak memenuhi spesifikasi sebagai berikut:

$$\begin{aligned} p &= P\{x < 1,00\} + P\{x > 2,00\} \\ &= \Phi\left(\frac{1,00 - 1,5056}{0,1398}\right) + 1 - \Phi\left(\frac{2,00 - 1,5056}{0,1398}\right) \\ &= (-3,61660) + 1 - \Phi(3,53648) \\ &\cong 0,00015 + 1 - 0,99980 \\ &\cong 0,00035 \end{aligned}$$

Artinya, sekitar 0,035 persen [350 bagian per juta (ppm)] wafer yang diproduksi akan berada di luar spesifikasi.

Cara lain untuk mengekspresikan kapabilitas proses adalah dalam hal **indeks kapabilitas proses** (C_p). C_p digunakan dalam menghitung kepresisian dari proses produksi, di mana C_p memperhatikan sebaran data apakah berada pada interval LSL & USL. C_p didefinisikan sebagai perbandingan antara rentang spesifikasi dan penyebaran alami proses (penyebaran 6-sigma) yaitu:

$$C_p = \frac{USL - LSL}{6\sigma} \tag{10.12}$$

Perhatikan bahwa sebaran proses 6σ merupakan definisi dasar kapabilitas proses. Karena σ biasanya tidak diketahui, kita harus menggantinya dengan estimasi. Kita sering menggunakan $\sigma = R/d_2$ sebagai estimasi σ , yang menghasilkan estimasi $\hat{C}_p$ sebesar C_p . Untuk proses *hard-bake*, $\bar{R}/d_2 = \hat{\sigma} = 0,1398$ dan diketahui $USL = 2$ dan $LSL = 1$, maka:

$$\hat{C}_p = \frac{2,00 - 1,00}{6 (0,1398)} = 1,192$$

Hal ini menunjukkan bahwa batas toleransi “alami” proses (tiga sigma di atas dan di bawah rata-rata) masih berada dalam batas spesifikasi atas dan bawah. Dengan demikian, hanya sebagian kecil wafer yang diperkirakan

akan berada di luar spesifikasi atau tidak sesuai. Adapun interpretasi nilai C_p ditampilkan pada Tabel berikut

No	Nilai C_p	Interpretasi
1	$C_p < 1$	Tidak Memadai (Not Capable). Proses akan menghasilkan produk cacat secara signifikan (melebihi batas spesifikasi), bahkan jika rata-rata proses berada di tengah spesifikasi.
2	$C_p = 1$	Minimum Acceptable. Proses hanya memenuhi spesifikasi secara teoritis.
3	$1 < C_p < 1,33$	Adekuat (Adequate). Memadai untuk proses lama atau proses non-kritis. Memberikan sedikit ruang toleransi (margin).
4	$C_p \geq 1,33$	Baik (Good). Tingkat yang disarankan untuk sebagian besar proses.
5	$C_p \geq 1,67$	Sangat Baik (Excellent). Umumnya diperlukan untuk proses yang mahal atau kritis, terutama di industri otomotif.
6	$C_p \geq 2$	Kelas Dunia (Six Sigma Level). Proses dengan kapabilitas sangat tinggi (hampir sempurna)

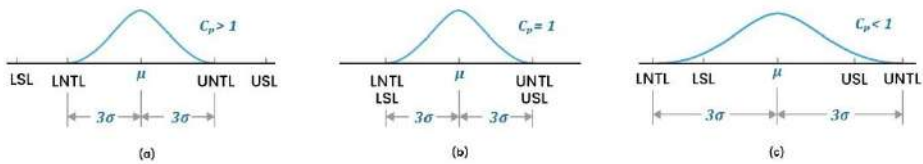
Indeks kapabilitas proses (C_p) juga dapat diinterpretasikan dengan cara lain. Besaran P menunjukkan persentase dari rentang spesifikasi yang digunakan oleh proses tersebut.

$$P = \left(\frac{1}{C_p}\right) 100\% \tag{10.13}$$

Untuk proses *hard-bake*, perkiraan P adalah

$$\hat{P} = \left(\frac{1}{C_p}\right) 100\% = \left(\frac{1}{1,192}\right)100\% =83,89$$

Artinya, proses tersebut menggunakan sekitar **83,89% dari total rentang** spesifikasi yang tersedia, sehingga masih terdapat ruang toleransi yang cukup baik untuk menjaga kualitas produk.



Gambar 10.3 . Kegagalan Proses dan Indeks Kapabilitas Proses (C_p)

Keterangan:

- LSL: *Lower Specification Limit* atau Batas Spesifikasi Bawah
- USL: *Upper Specification Limit* atau Batas Spesifikasi Atas
- LNTL: *Lower Natural Tolerance Limit* atau Batas Toleransi Alami Bawah
- UNTL: *Upper Natural Tolerance Limit* atau Batas Toleransi Alami Atas
- μ = rata-rata proses
- σ = standar deviasi proses

Gambar 10.3 menggambarkan tiga situasi yang berkaitan dengan C_p dan batas spesifikasi proses. Pada **Gambar 10.3a**, nilai C_p lebih besar dari satu. Hal ini berarti bahwa proses hanya menggunakan sebagian kecil dari rentang toleransi yang tersedia (kurang dari 100%). Akibatnya, jumlah produk yang tidak sesuai spesifikasi relatif sedikit akan diproduksi dalam proses ini. **Gambar 10.3b** menunjukkan proses dengan nilai $C_p = 1$, yang berarti proses tersebut menggunakan seluruh rentang toleransi yang tersedia. Untuk distribusi normal, kondisi ini menunjukkan bahwa sekitar 0,27% (atau 2700 ppm) produk yang dihasilkan tidak sesuai dengan spesifikasi. Sementara itu,

Gambar 10.3c menggambarkan proses dengan nilai $C_p < 1$, yang berarti proses tersebut melebihi batas toleransi yang ditetapkan. Dalam situasi ini, proses menjadi sangat sensitif terhadap variasi, sehingga menghasilkan jumlah produk cacat yang cukup besar.

D. Indeks Kapabilitas Proses Kritis (C_{pk})

C_{pk} mengukur kapabilitas proses aktual dengan mempertimbangkan pergeseran rata-rata proses $\bar{\bar{x}}$ dari target spesifikasi. Dengan kata lain, C_{pk} adalah nilai terendah (minimum) antara kapabilitas sisi atas C_{pu} dan kapabilitas sisi bawah C_{pl} .

$$C_{pk} = \min (C_{pu}, C_{pl}) \tag{10.14}$$

dimana

$$C_{pu} = \frac{USL - \bar{\bar{x}}}{3\sigma} \qquad C_{pl} = \frac{\bar{\bar{x}} - LSL}{3\sigma}$$

Interpretasi nilai C_{pk} ditampilkan pada tabel berikut

No	Nilai C_{pk}	Interpretasi
1	$C_{pk} < 1$	Tidak Mampu. Proses pasti akan menghasilkan produk cacat. Proses harus diperbaiki (lokasi pusat dan/atau variabilitas).
2	$C_{pk} = 1$	Mampu Secara Marginal. Rata-rata proses terdekat dengan batas spesifikasi hanya berjarak 3σ . Risiko cacat tinggi jika ada pergeseran kecil.
3	$C_{pk} \geq 1,33$	Sangat Baik. Target minimum yang direkomendasikan untuk sebagian besar industri (berarti rata-rata terdekat 4σ dari batas spesifikasi).
4	$C_{pk} \geq 1,67$	Kelas Dunia/Enam Sigma (Level 5-Sigma). Kapabilitas tinggi, proses terpusat dengan baik, dan variabilitas rendah.

Hubungan antara C_p dan C_{pk} :

- Jika $C_{pk} = C_p$, proses terpusat sempurna di tengah rentang spesifikasi
- Jika $C_{pk} < C_p$, proses tidak terpusat. Semakin besar selisih antara C_p dan C_{pk} semakin jauh rata-rata proses ($\bar{\bar{x}}$) bergeser dari titik tengah spesifikasi.

E. Peta Kendali $\bar{x}$ dan s

Meskipun diagram $\bar{X}$ dan R banyak digunakan secara praktik, dalam beberapa kasus lebih disarankan untuk memperkirakan simpangan baku proses secara langsung daripada secara tidak langsung melalui Range (R). Pendekatan ini menghasilkan peta kendali $\bar{x}$ dan s , di mana s merupakan simpangan baku dari sampel. Secara umum, diagram $\bar{x}$ dan s , lebih disukai dibandingkan diagram $\bar{X}$ dan R , terutama ketika:

- Ukuran sampel (n) cukup besar, misalnya $n > 10$ atau 12, karena metode range kehilangan efisiensi statistik untuk ukuran sampel sedang hingga besar.
- Ukuran sampel (n) bervariasi dari satu pengamatan ke pengamatan lainnya.

Pembuatan dan penggunaan diagram kendali $\bar{x}$ dan s mengikuti langkah-langkah yang hampir sama dengan diagram kendali $\bar{X}$ dan R . Perbedaannya terletak pada perhitungan untuk setiap sampel, di mana kita harus menghitung rata-rata sampel ($\bar{X}$) dan simpangan baku sampel (s). Jika σ^2 adalah varians yang tidak diketahui dari suatu distribusi probabilitas, maka penaksir tak bias dari σ^2 adalah varian sampel:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \quad (10.15)$$

Namun, simpangan baku sampel s bukan merupakan penduga tak bias untuk σ . Jika data berasal dari distribusi normal, maka s sebenarnya menaksir $c_4\sigma$, dengan c_4 sebagai konstanta yang bergantung pada ukuran

sampel n . Selain itu, simpangan baku dari s adalah $\sigma\sqrt{1 - c_4^2}$. Informasi ini dapat digunakan untuk membangun diagram kendali $\bar{x}$ dan s .

Perhatikan kasus di mana nilai standar diberikan untuk σ . Karena $E(s) = c_4\sigma$, garis tengah grafik adalah $c_4\sigma$. Batas kendali tiga sigma untuk s adalah

$$UCL = c_4\sigma + 3\sigma\sqrt{1 - c_4^2}$$

$$LCL = c_4\sigma - 3\sigma\sqrt{1 - c_4^2}$$

Dengan definisi dua konstanta

$$B_5 = c_4 - 3\sqrt{1 - c_4^2} \text{ dan } B_6 = c_4 + 3\sqrt{1 - c_4^2} \quad (10.16)$$

Oleh karena itu, parameter grafik s dengan nilai standar untuk suatu keadaan tertentu menjadi

$$\begin{aligned} UCL &= B_6 \sigma \\ \text{Garis Tengah} &= c_4 \sigma \\ LCL &= B_5 \sigma \end{aligned} \quad (10.17)$$

Nilai B_5 dan B_6 untuk berbagai ukuran sampel dapat dilihat pada Lampiran Tabel VI.

Jika tidak terdapat standar yang diketahui untuk σ , maka nilai tersebut perlu diestimasi berdasarkan data historis atau data awal proses. Misalkan tersedia m sampel awal, masing-masing berukuran n , dan s adalah simpangan baku dari sampel ke- i . Maka, rata-rata dari m simpangan baku tersebut dapat dihitung sebagai berikut:

$$\bar{s} = \frac{1}{m} \sum_{i=1}^m s_i$$

Statistik $\bar{s}/c_4$ merupakan penduga tak bias dari σ . Oleh karena itu, parameter diagram s adalah

$$UCL = \bar{s} + 3 \frac{\bar{s}}{c_4} \sqrt{1 - c_4^2}$$

$$\text{Garis Tengah} = \bar{s}$$

$$LCL = \bar{s} - 3 \frac{\bar{s}}{c_4} \sqrt{1 - c_4^2}$$

Dengan definisi konstanta

$$B_3 = 1 - \frac{3}{c_4} \sqrt{1 - c_4^2} \text{ dan } B_4 = 1 + \frac{3}{c_4} \sqrt{1 - c_4^2} \quad (10.18)$$

Oleh karena itu, kita dapat menuliskan parameter grafik s sebagai

$$\begin{aligned} UCL &= B_4 \bar{s} \\ \text{Garis tengah} &= \bar{s} \\ LCL &= B_3 \bar{s} \end{aligned} \quad (10.19)$$

Perhatikan bahwa $B_4 = B_6/c_4$ dan $B_3 = B_5/c_4$.

Ketika $\bar{s}/c_4$ digunakan untuk memperkirakan σ , kita dapat mendefinisikan batas kontrol pada grafik $\bar{x}$ yang sesuai sebagai:

$$UCL = \bar{\bar{x}} + \frac{3\bar{s}}{c_4\sqrt{n}}$$

$$\text{Garis Tengah} = \bar{\bar{x}}$$

$$LCL = \bar{\bar{x}} - \frac{3\bar{s}}{c_4\sqrt{n}}$$

Misalkan konstanta $A_3 = 3/(c_4\sqrt{n})$. Maka parameter pada diagram $\bar{x}$ menjadi:

$$\begin{aligned} UCL &= \bar{\bar{x}} + A_3\bar{s} \\ \text{Garis Tengah} &= \bar{\bar{x}} \\ LCL &= \bar{\bar{x}} - A_3\bar{s} \end{aligned} \quad (10.20)$$

Konstanta B_3, B_4 , dan A_3 dalam membangun diagram $\bar{x}$ dan s tercantum dalam Lampiran Tabel VI untuk berbagai ukuran sampel. Perlu dicatat bahwa kita telah mengasumsikan bahwa simpangan baku sampel didefinisikan sebagai:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (10.21)$$

Contoh 10.2

Data berikut menunjukkan hasil pengukuran diameter cincin piston pada mesin mobil tempa, di mana setiap subkelompok (observasi) terdiri dari lima cincin piston. Buatlah dan interpretasikan diagram kendali $\bar{x}$ dan s berdasarkan data pengukuran diameter cincin piston tersebut.

Sampel	Observasi					$\bar{x}_i$	s_i
1	74,030	74,002	74,019	73,992	74,008	74,010	0,0148
2	73,995	73,992	74,001	74,011	74,004	74,001	0,0075
3	73,998	74,024	74,021	74,005	74,002	74,008	0,0147
4	74,002	73,996	73,993	74,015	74,009	74,003	0,0091
5	73,992	74,007	74,015	73,989	74,014	74,003	0,0122
6	74,009	73,994	73,997	73,985	73,993	73,996	0,0087
7	73,995	74,006	73,994	74,000	74,005	74,000	0,0055
8	73,985	74,003	73,993	74,015	73,988	73,997	0,0123
9	74,008	73,995	74,009	74,005	74,004	74,004	0,0055
10	73,998	74,000	73,990	74,007	73,995	73,998	0,0063
11	73,994	73,998	73,994	73,995	73,990	73,994	0,0029
12	74,004	74,000	74,007	74,000	73,996	74,001	0,0042
13	73,938	74,002	73,998	73,997	74,012	73,998	0,0105
14	74,006	73,967	73,994	74,000	73,984	73,990	0,0153
15	74,012	74,014	73,998	73,999	74,007	74,006	0,0073
16	74,000	73,984	74,005	73,998	73,996	73,997	0,0078
17	73,994	74,012	73,986	74,005	74,007	74,001	0,0106
18	74,006	74,010	74,018	74,003	74,000	74,007	0,0070
19	73,984	74,002	74,003	74,005	73,997	73,998	0,0085

20	74,000	74,010	74,013	74,020	74,003	74,009	0,0080
21	73,982	74,001	74,015	74,005	73,996	74,000	0,0122
22	74,004	73,999	73,990	74,006	74,009	74,002	0,0074
23	74,010	73,989	73,990	74,009	74,014	74,002	0,0119
24	74,015	74,008	73,993	74,000	74,010	74,005	0,0087
25	73,982	73,984	73,995	74,017	74,013	73,998	0,0162
							$\sum \bar{x}_i$ = 1850,028
							$\sum s_i = 0,2351$

Jawab:

Rata-rata keseluruhan dan rata-rata simpangan baku adalah:

$$\bar{\bar{x}} = \frac{1}{25} \sum_{i=1}^{25} \bar{x}_i = \frac{1}{25} (1850,028) = 74,001$$

$$\bar{s} = \frac{1}{25} \sum_{i=1}^{25} s_i = \frac{1}{25} (0,2351) = 0,0094$$

Parameter untuk diagram $\bar{x}$ adalah:

$$UCL = \bar{\bar{x}} + A_3 \bar{s} = 74,001 + (1,427)(0,0094) = 74,014$$

$$\text{Garis Tengah} = \bar{\bar{x}} = 74,001$$

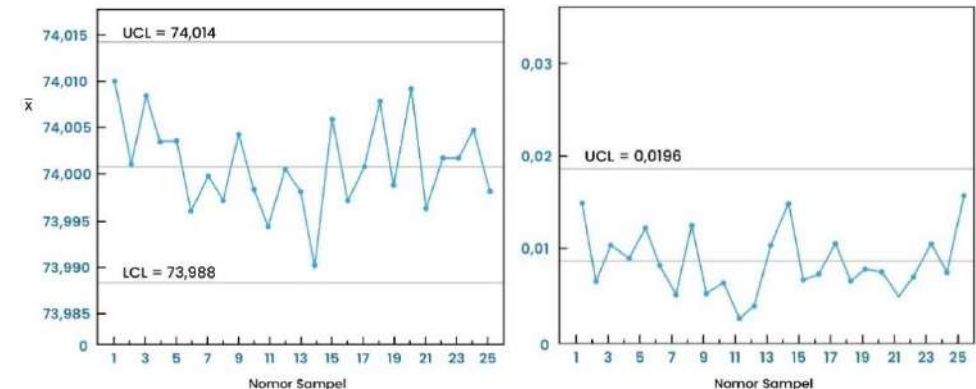
$$LCL = \bar{\bar{x}} - A_3 \bar{s} = 74,001 - (1,427)(0,0094) = 73,988$$

Dan untuk diagram s :

$$UCL = B_4 \bar{s} = (2,089)(0,0094) = 0,0196$$

$$\text{Garis Tengah} = \bar{s} = 0,0094$$

$$LCL = B_3 \bar{s} = (0)(0,0094) = 0$$



(a)

(b)

Grafik kendali $\bar{x}$ pada Gambar (a) memplot rata-rata dari setiap sampel terhadap batas kendali atas (UCL) yaitu 74,014 dan batas kendali bawah (LCL) yaitu 73,988. Semua titik data berada di antara UCL dan LCL. Sehingga proses terkait rata-rata sampel dalam kendali statistik (*in control*) yang menunjukkan rata-rata proses stabil dan tidak ada sebab khusus (*assignable causes*) yang signifikan yang memengaruhi lokasi pusat proses.

Grafik kendali s pada Gambar (b) memplot standar deviasi dari setiap sampel terhadap batas kendali atas (UCL). Grafik standar deviasi seringkali tidak memiliki batas kendali bawah (LCL) yang ditampilkan jika nilainya nol atau sangat dekat dengan nol. Semua titik data (standar deviasi sampel) berada **di bawah** UCL dan tidak ada titik yang melebihi batas kendali. Sehingga proses terkait penyebaran/variabilitas s juga dalam kendali statistik (*in control*) yang menunjukkan bahwa variabilitas proses (keberagaman antar sampel) stabil dan konsisten dari waktu ke waktu.

Estimasi nilai σ . Kita dapat memperkirakan simpangan baku proses dengan menggunakan fakta bahwa s/c_4 adalah estimasi tak bias untuk σ . Oleh karena itu, karena $c_4 = 0,9400$ untuk sampel berukuran 5, maka estimasi dari simpangan baku proses adalah:

$$\hat{\sigma} = \frac{\bar{s}}{c_4} = \frac{0,0094}{0,9400} = 0,01$$

F. Peta Kendali Individual (*I-Chart*) dan Moving Range (*MR-Chart*)

Ada banyak kasus di mana ukuran sampel untuk memantau proses hanya terdiri dari satu unit tunggal ($n = 1$). Ini berarti, setiap kali kita mengambil data, kita hanya mengamati satu barang atau satu pengukuran. Berikut adalah beberapa contoh situasi di mana pemantauan dengan sampel tunggal ini sering digunakan:

1. **Sistem Inspeksi Otomatis:** Ketika teknologi inspeksi dan pengukuran dilakukan secara otomatis, dan setiap unit produk langsung dianalisis. Dalam kondisi ini, tidak ada alasan kuat untuk mengelompokkan data (membentuk subgroup rasional), karena setiap unit sudah mewakili titik data individual.
2. **Data Datang Lambat:** Ketika data hasil produksi terkumpul dengan sangat lambat, sehingga tidak praktis menunggu hingga terkumpul cukup banyak unit untuk membentuk satu sampel (subgroup). Jeda waktu yang terlalu lama antar pengamatan akan membuat pembentukan subgroup rasional menjadi tidak relevan, baik itu di lingkungan manufaktur maupun non-manufaktur.
3. **Pengukuran Ulang Akibat Kesalahan:** Ketika pengukuran ulang pada suatu proses terjadi hanya karena adanya kesalahan teknis atau analisis (misalnya, di laboratorium kimia), bukan karena produk baru. Dalam kasus ini, setiap pengukuran ulang dianggap sebagai observasi tunggal.
4. **Pengukuran Berulang pada Bagian yang Sama:** Ketika beberapa pengukuran diambil pada satu bagian produk yang sama, dan setiap pengukuran tersebut dianggap sebagai data independen. Contohnya adalah mengukur ketebalan di beberapa titik berbeda pada satu wafer di industri semikonduktor.

Diagram Kendali Individual (*I-Chart*) dan Range Berpindah (*MR-Chart*) adalah alat Kendali Proses Statistik (KPS) yang digunakan ketika ukuran sampel hanya terdiri dari satu observasi ($n = 1$) per periode sampling. Oleh karena itu, pasangan ini sering disebut sebagai *I-MR Chart*.

Diagram Kendali Individual (*I-Chart*) memantau lokasi pusat (rata-rata) dari proses dari waktu ke waktu. Kunci dari *I-Chart* adalah

mengestimasi standar deviasi proses (σ) menggunakan $\overline{MR}$. Estimator σ yang tidak bias adalah:

$$\hat{\sigma} = \frac{\overline{MR}}{d_2}$$

Perhitungan batas kendali *I-Chart* dihitung sebagai Garis Tengah $\pm 3\hat{\sigma}$ (aturan 3-sigma Shewhart).

$$\begin{aligned} UCL &= \bar{x} + 3\hat{\sigma} = \bar{x} + 3 \frac{\overline{MR}}{d_2} \\ \text{Garis Tengah} &= \bar{x} \\ LCL &= \bar{x} - 3\hat{\sigma} = \bar{x} - 3 \frac{\overline{MR}}{d_2} \end{aligned} \quad (10.22)$$

Dalam banyak situasi, terutama pada Diagram Kendali Individual (*I-Chart*), kita perlu cara untuk memperkirakan seberapa bervariasi proses tersebut. Untuk melakukan ini, kita sering menggunakan **Range Berpindah (Moving Range)** yang dihitung dari dua pengamatan atau pengukuran yang berurutan sebagai dasar untuk memperkirakan variasi proses. Rentang berpindah (MR_i) didefinisikan sebagai:

$$MR_i = |x_i - x_{i-1}|$$

Diagram kendali khusus untuk memantau nilai Range Berpindah (*MR*) ini dikenal sebagai *MR-Chart*. Diagram kendali *MR* memantau variabilitas jangka pendek proses.

$$\begin{aligned} UCL &= D_4 \overline{MR} \\ \text{Garis Tengah} &= \overline{MR} \\ LCL &= D_3 \overline{MR} \end{aligned} \quad (10.23)$$

Contoh 10.3

Unit pembuatan komponen presisi dari sebuah pabrik memantau biaya pemrosesan untuk memproduksi *wafers* semikonduktor berkinerja tinggi. Kuantitas yang dilacak adalah **rata-rata biaya pemrosesan mingguan**, yang diperoleh dengan membagi total biaya mingguan dengan jumlah *wafers* yang diproses dalam seminggu. Data biaya pemrosesan selama 20 minggu terakhir ditunjukkan pada tabel berikut. Buatlah diagram kendali Pengukuran individual (*I-Chart*) dan range berpindah (*moving range*) untuk data ini.

Mingguan	Biaya(x)	Range Berpindah (MR)
1	310	
2	288	22
3	297	9
4	298	1
5	307	9
6	303	4
7	294	9
8	297	3
9	308	11
10	306	2
11	294	12
12	299	5
13	297	2
14	299	2
15	314	15
16	295	19
17	293	2
18	306	13
19	301	5
20	304	3
	$\bar{x} = 300,5$	$\overline{MR} = 7,79$

Jawab:

Untuk membuat diagram kontrol bagi **pengamatan individual**, diketahui bahwa rata-rata biaya dari 20 pengamatan adalah $\bar{x} = 300,5$ dan rata-rata dari *moving range* dua pengamatan berturut-turut adalah $\overline{MR} = 7,79$.

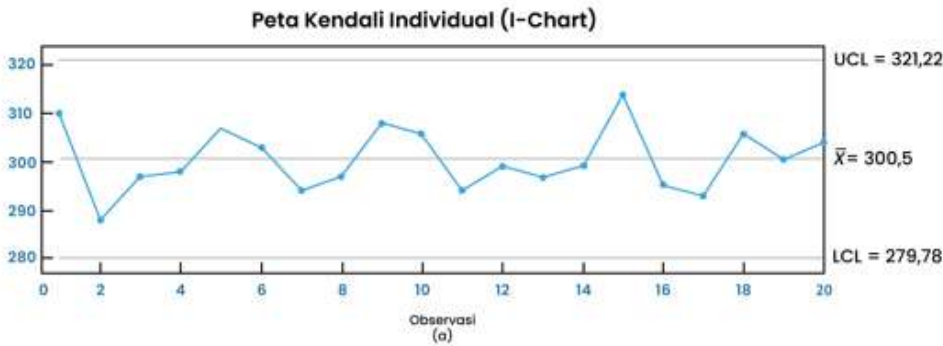
Dalam diagram kontrol Pengukuran individual (*I-Chart*) parameternya adalah:

$$UCL = \bar{x} + 3 \frac{\overline{MR}}{d_2} = 300,5 + 3 \left(\frac{7,79}{1,128} \right) = 321,22$$

Garis Tengah = $\bar{x} = 300,5$

$$LCL = \bar{x} - 3 \frac{\overline{MR}}{d_2} = 300,5 - 3 \frac{7,79}{1,128} = 279,78$$

Diagram kontrol untuk Pengukuran individual (*I-Chart*) ditunjukkan pada Gambar berikut dan tidak ada pengamatan yang berada di luar batas kontrol pada diagram kontrol individual tersebut. Berdasarkan kedua diagram kendali, proses pemrosesan *wafers* semikonduktor dalam hal biaya mingguan **berada dalam kendali statistik** (*in control*).



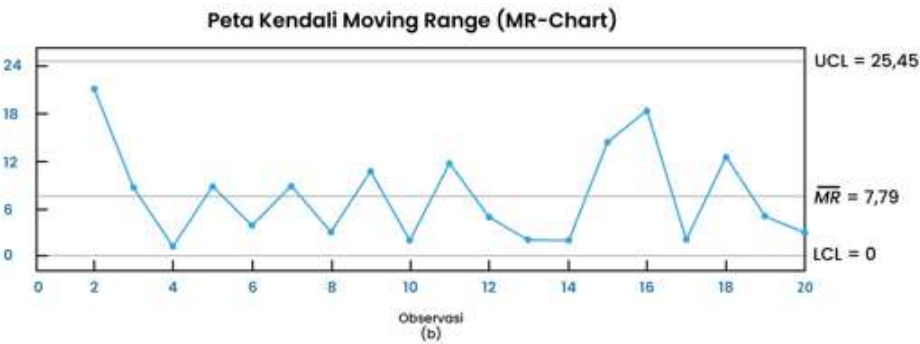
Selanjutnya, untuk menyusun diagram *moving range*, parameternya adalah:

$$UCL = D_4 \overline{MR} = (3,267)7,79 = 25,45$$

Garis Tengah = $\overline{MR} = 7,79$

$$LCL = D_3 \overline{MR} = (0)7,79 = 0$$

Diagram kendali Range Berpindah (*MR-Chart*) ditampilkan pada Gambar berikut dan tidak ada titik yang berada di luar batas kontrol.



TABEL RANGKUMAN.

Untuk mempermudah pemahaman prosedur Kendali Mutu Statistik, kami menyajikan ringkasan rumus-rumus penting untuk penetapan Garis Tengah dan Batas Kendali (UCL serta LCL) pada beragam jenis Diagram Kendali *Shewhart*. Konstanta yang relevan untuk perhitungan ini telah ditabulasi berdasarkan ukuran sampel tersedia pada Tabel VI Lampiran.

Jenis Chart	Batas Kendali Atas (UCL)	Garis Tengah (<i>Center Line</i>)	Batas Kendali Bawah (LCL)
$\bar{x}$ (menggunakan R)	$\bar{\bar{x}} + A_2 \bar{R}$	$\bar{\bar{x}}$	$\bar{\bar{x}} - A_2 \bar{R}$
R	$D_4 \bar{R}$	$\bar{R}$	$D_3 \bar{R}$
$\bar{x}$ (menggunakan s)	$\bar{\bar{x}} + A_3 \bar{s}$	$\bar{\bar{x}}$	$\bar{\bar{x}} - A_3 \bar{s}$
s	$B_4 \bar{s}$	$\bar{s}$	$B_3 \bar{s}$
I	$\bar{\bar{x}} + 3 \frac{\overline{MR}}{d_2}$	$\bar{\bar{x}}$	$\bar{\bar{x}} - 3 \frac{\overline{MR}}{d_2}$
MR	$D_4 \overline{MR}$	$\overline{MR}$	$D_3 \overline{MR}$

LATIHAN 10

Peta Kendali Untuk Variabel

- Sebuah pabrik manufaktur memproduksi suatu jenis komponen otomotif yang akan di *supply* ke perusahaan otomotif. Pemeriksaan dilakukan terhadap salah satu karakteristik kualitas komponen tersebut. Sebanyak 7 sampel diambil yang masing-masing terdiri dari 4 pengamatan dan hasil pengukurannya (dalam mm) adalah:

Sampel	Karakteristik Kualitas (mm)			
	X_1	X_2	X_3	X_4
1	145	159	146	152
2	148	159	149	148
3	147	146	150	151
4	156	151	153	150
5	155	146	152	145
6	150	152	160	143
7	157	150	161	146

- Sketsalah grafik kendali *Shewhart*: *R chart* dan *x-bar chart* berdasarkan data pada tabel
 - Hitung kapabilitas proses berdasarkan nilai C_p & C_{pk} dan simpulkan hasilnya.
- Dalam proses pembuatan Bearing dibuat peta kendali untuk diameter luar bearing dengan ukuran 5 subgroup. Dari 20 sampel diperoleh $\sum \bar{X} = 47,82$ dan $\sum R = 0,68$. Batas spesifikasi $2,4 \pm 0,05$ dengan biaya Rework 2000/pcs dan biaya Scrap 20.000/pcs Jika diasumsikan data berdistribusi normal. Maka hitunglah :
 - Capabilitas proses dan indeks nya beserta intrepetasinya
 - Total biaya Rework dan Scrap untuk produksi di atas.

3. Sebuah daya tegangan tinggi memiliki tegangan keluaran nominal yang seharusnya 350 V. Untuk tujuan pengendalian proses, setiap hari diambil satu sampel yang terdiri dari empat unit untuk diuji. Pada tabel berikut ditunjukkan nilai deviasi dari tegangan, lalu dikalikan sepuluh. Secara matematis, ini dihitung sebagai:

$$x_i = (\text{tegangan yang diamati} - 350) \times 10$$

Sampel	x_1	x_2	x_3	x_4
1	6	9	10	15
2	10	4	6	11
3	7	8	10	5
4	8	9	6	13
5	9	10	7	13
6	12	11	10	10
7	16	10	8	9
8	7	5	10	4
9	9	7	8	12
10	15	16	10	13
11	8	12	14	16
12	6	13	9	11
13	16	9	13	15
14	7	13	10	12
15	11	7	10	16
16	15	10	11	14
17	9	8	12	10
18	15	7	10	11
19	8	6	9	12
20	13	14	11	15

- Buat diagram $\bar{x}$ dan R untuk proses ini. Apakah prosesnya berada dalam kendali statistik?
- Jika spesifikasi berada pada $350 \text{ V} \pm 5 \text{ V}$, apa yang dapat kamu katakan tentang kapabilitas proses?

- Apakah ada bukti yang mendukung klaim bahwa tegangan terdistribusi secara normal?

4. Ketebalan papan sirkuit cetak (*printed circuit board*) merupakan parameter kualitas yang krusial. Data ketebalan papan yang disajikan dalam satuan inci, tercantum pada tabel di bawah. Data ini dikumpulkan dari 25 sampel, di mana setiap sampelnya terdiri dari tiga papan.

Sampel	x_1	x_2	x_3
1	0,0629	0,0636	0,0640
2	0,0630	0,0631	0,0622
3	0,0628	0,0631	0,0633
4	0,0634	0,0630	0,0631
5	0,0619	0,0628	0,0630
6	0,0613	0,0629	0,0634
7	0,0630	0,0639	0,0625
8	0,0628	0,0627	0,0622
9	0,0623	0,0626	0,0633
10	0,0631	0,0631	0,0633
11	0,0635	0,0630	0,0638
12	0,0623	0,0630	0,0630
13	0,0635	0,0631	0,0630
14	0,0645	0,0640	0,0631
15	0,0619	0,0644	0,0632
16	0,0631	0,0627	0,0630
17	0,0616	0,0623	0,0631
18	0,0630	0,0630	0,0626

19	0,0636	0,0631	0,0629
20	0,0640	0,0635	0,0629
21	0,0628	0,0625	0,0616
22	0,0615	0,0625	0,0619
23	0,0630	0,0632	0,0630
24	0,0635	0,0629	0,0635
25	0,0623	0,0629	0,0630

- Buat diagram kontrol $\bar{x}$ dan R . Apakah proses berada dalam kendali statistik?
 - Estimasikan simpangan baku proses.
 - Berapakah batas-batas yang diharapkan dapat mencakup hampir semua hasil pengukuran proses?
 - Jika spesifikasinya berada pada $0.0630 \text{ inci} \pm 0.0015 \text{ inci}$, berapakah nilai dari C_p ?
5. Untuk data pada no 3 dan 4, buatlah diagram kendali $\bar{x}$ dan s . Berikan interpretasi dari masing-masing diagram kendali yang diperoleh.

DAFTAR PUSTAKA

- [1] Alt, F. B. (1985). "Multivariate Quality Control," in Encyclopedia of Statistical Sciences, Vol. 6, N. L. Johnson and S. Kotz, (eds.) Wiley, New York.
- [2] Arief, F. (2011). *Pengantar Penelitian dalam Pendidikan*. Yogyakarta: Pustaka Pelajar
- [3] Automotive Industry Action Group (2002). *Measurement Systems Analysis*, 3rd ed., Detroit, MI.
- [4] Borror, C. M., D. C. Montgomery, and G. C. Runger (1997). "Confidence Intervals for Variance Components from Gauge Capability Studies," *Quality and Reliability Engineering International*, Vol. 13(6), pp. 361–369.
- [5] Bowker, A. H., and G. J. Lieberman (1972). *Engineering Statistics*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ.
- [6] Box, G. E. P., and A. Luceño (1997). *Statistical Control by Monitoring and Feedback Adjustment*, Wiley, New York.
- [7] Boyd, D. F. (1950). "Applying the Group Control Chart for $\bar{x}$ and R ," *Industrial Quality Control*, Vol. 7(3), pp. 22–25.
- [8] Burdick, R. K., C. M. Borror, and D. C. Montgomery (2003). "A Review of Methods for Measurement Systems Capability Analysis," *Journal of Quality Technology*, Vol. 35(4), pp. 342–354.
- [9] Del Castillo. E. (2002), *Statistical Process Adjustment for Quality Control*, John Wiley & Sons, New York.
- [10] Douglas, C. (2009). *Introduction to Statistical Quality Control – Sixth Edition*. John Wiley & Sons.
- [11] Duncan, A. J. (1986). *Quality Control and Industrial Statistics*, 5th ed., Irwin, Homewood, IL.

- [12] Edward, A.L. (1984). *An Introduction to Linier Regression and Correlation*. 2nd ed. New York: W.H.Freeman and Company.
- [13] Gardiner, J. S., and D. C. Montgomery (1987). "Using Statistical Control Charts for Software Quality Control," *Quality and Reliability Engineering International*, Vol. 3(1), pp. 15–20.
- [14] Grant, E. L., and R. S. Leavenworth (1980). *Statistical Quality Control*, 5th ed., McGraw-Hill, New York.
- [15] Hahn, G. J., and S. S. Shapiro (1967). *Statistical Models in Engineering*, Wiley, New York
- [16] Johnson, R. A.(2010). *Statistics: Principles and Methods*. (6rd ed).Hoboken, NJ: John Wiley & Sons
- [17] Johnson, R.A & Wichern, D.W .(2007). *Applied Multivariate Statistical Analysis*. London : Pearson Prenti ce Hall
- [18] Montgomery, D. C. (2005). *Design and Analysis of Experiments*, 6th ed., Wiley, New York
- [19] Montgomery, D. C., E. A. Peck, and G. G. Vining (2006). *Introduction to Linear Regression Analysis*, 4th ed., Wiley, New York.
- [20] Montgomery, D. C., and G. C. Runger (2007). *Applied Statistics and Probability for Engineers*, 4th ed., Wiley, New York.
- [21] Pramana, S., et.al. (2016). *Dasar-Dasar Statistika dengan Software R: Konsep dan Aplikasi*. In Media
- [22] Riduwan.(2008). *Dasar-dasar Statistika*. Bandung: Alfabeta.
- [23] Sartono, B., et.al. (2019). *Kumpulan Soal Hitung Peluang dan Teori Statistika*. IPB Press
- [24] Spiring, F., B. Leung, S. Cheng, and A. Yeung (2003). "A Bibliography of Process Capability Papers," *Quality and Reliability Engineering International*, Vol. 19(5), pp. 445–460
- [25] Stevens, J. (2002). *Applied multivariate statistics for the social sciences*. London: Lawrence Erlbaum Associates, Publishers.

- [26] Sudjana. (2005). *Metode Statistika*, Bandung : Tarsito.
- [27] Sugiyono. (2010). *Statistika untuk Penelitian*. Bandung: Alfabeta
- [28] Supranto, J. (2016). *Statistik: Teori dan Aplikasi*. Erlangga: Edisi Kedelapan
- [29] Supangat, A. (2010). *Statistika dalam Kajian Deskriptif, Inferensi dan Non Parametrik*. Kencana Media Group: Cetakan Ketiga
- [30] Wadsworth, H. M., K. S. Stephens, and A. B. Godfrey (2002). *Modern Methods for Quality Control and Improvement*, 2nd ed., Wiley, New York.
- [31] Walpole, R.E., et.al. (2007). *Probability and Statistics for Engineers & Scientists*. Pearson Education International.

GLOSARIUM

Akurasi	: Tingkat kedekatan hasil pengukuran dengan nilai sebenarnya.
Analisis Regresi	: Metode statistika untuk memodelkan hubungan antara satu atau lebih variabel bebas dengan variabel terikat.
Analisis Varians (ANOVA)	: Teknik untuk membandingkan rata-rata dari lebih dari dua kelompok
Batas Kendali (<i>Control Limits</i>)	: Garis atas dan bawah pada peta kendali yang menentukan apakah proses masih terkendali atau tidak.
Batas Spesifikasi (<i>Specification Limits</i>)	: Rentang yang ditetapkan oleh desain atau pelanggan sebagai toleransi produk (LSL–USL).
Bias	: Penyimpangan sistematis dalam proses pengukuran atau estimasi.
Distribusi Binomial	: Distribusi peluang untuk sejumlah keberhasilan dalam n percobaan Bernoulli.
<i>Confidence Interval</i>	: Rentang nilai yang digunakan untuk mengestimasi parameter populasi pada tingkat keyakinan tertentu.
Variabel Kontinu	: Variabel yang dapat mengambil nilai pada rentang tak terbatas.
Cp (Process Capability Ratio)	: Ukuran kemampuan proses terhadap batas spesifikasi.
Cpk (Process Capability Index)	: Ukuran kemampuan proses yang mempertimbangkan posisi rata-rata proses.
Estimasi Parameter	: Pendugaan nilai karakteristik populasi menggunakan data sampel.
Grafik Kendali / <i>Control Chart</i>	: Alat dalam SPC untuk memantau stabilitas proses dari waktu ke waktu

I-Chart (<i>Individual Chart</i>)	: Peta kendali untuk memantau nilai individu dari proses ketika ukuran sampel $n = 1$
IQR (<i>Interquartile Range</i>)	: Ukuran penyebaran yang dihitung dari $Q3 - Q1$.
LCL (<i>Lower Control Limit</i>)	: Batas kendali bawah pada grafik kontrol.
LSL (<i>Lower Specification Limit</i>)	: Batas spesifikasi bawah yang ditetapkan pelanggan/desain
MR-Chart (<i>Moving Range Chart</i>)	: Peta kendali untuk memantau ragam menggunakan rentang pergerakan antar data individu.
<i>Odds Ratio</i>	: Rasio kecenderungan suatu kejadian pada kelompok tertentu dibandingkan kelompok lain (digunakan dalam regresi logistik)
<i>P-Value</i>	: Probabilitas memperoleh data setidaknya se-ekstrem hasil observasi jika H_0 benar.
R Studio	: Perangkat lunak untuk analisis statistika dan visualisasi data.
<i>Skewness</i> (Kemencengan)	: Ukuran ketidaksimetrisan distribusi data.
SPC (<i>Statistical Process Control</i>)	: Metode statistik untuk memantau, mengendalikan, dan meningkatkan proses.
Teorema Limit Tengah (<i>Central Limit Theorem</i>)	: Menyatakan bahwa distribusi rata-rata sampel akan mendekati normal ketika ukuran sampel besar.
UCL (<i>Upper Control Limit</i>)	: Batas kendali atas pada peta kendali.
USL (<i>Upper Specification Limit</i>)	: Batas spesifikasi atas yang ditentukan oleh desain
Varians	: Ukuran penyebaran data berdasarkan kuadrat deviasi terhadap mean

LAMPIRAN

- Lampiran I : Distribusi F
- Lampiran II : Statistik *d* Durbin - Watson
- Lampiran III : Nilai Kritis Prosedur Uji Peringkat Bertanda Wilcoxon
- Lampiran IV : Distribusi *U* dalam Pengujian Mann-Whitney
- Lampiran V : Nilai Kritis *r* dalam Runs Test untuk Melihat Keacakan
- Lampiran VI : Faktor-faktor untuk Membangun Grafik Kontrol Variabel
- Lampiran VII : Faktor untuk Normal Dua Sisi Batas Toleransi
- Lampiran VIII : Kuantil Uji Kruskal-Wallis



■ LAMPIRAN I
Titik persentase dari distribusi *f*

		Derajat Kebebasan Untuk Pembilang (v_1)																				
		$F_{\alpha; v_1, v_2}$																				
$v_2 \backslash v_1$		1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞		
Derajat Kebebasan Untuk Pembilang (v_2)	1	4052	4999.5	5403	5625	5764	5859	5928	5982	6022	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366		
	2	98.5	99	99.17	99.25	99.3	99.33	99.36	99.37	99.39	99.4	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.5		
	3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.6	26.5	26.41	26.32	26.22	26.13		
	4	21.2	18	16.69	15.98	15.52	15.21	14.98	14.8	14.66	14.55	14.37	14.2	14.02	13.93	13.84	13.75	13.65	13.56	13.46		
	5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.2	9.11	9.02		
	6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.1	7.98	7.87	7.72	7.56	7.4	7.31	7.23	7.14	7.06	6.97	6.88		
	7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65		
	8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.2	5.12	5.03	4.95	4.86		
	9	10.56	8.02	6.99	6.42	6.06	5.8	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.4	4.31		
	10	10.04	7.56	6.55	5.99	5.64	5.39	5.2	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4	3.91		
	11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.4	4.25	4.1	4.02	3.94	3.86	3.78	3.69	3.6		
	12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.5	4.39	4.3	4.16	4.01	3.86	3.78	3.7	3.62	3.54	3.45	3.36		
	13	9.07	6.7	5.74	5.21	4.86	4.62	4.44	4.3	4.19	4.1	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17		
	14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.8	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3		
	15	8.68	6.36	5.42	4.89	4.52	4.32	4.14	4	3.89	3.8	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87		
	16	8.53	6.23	5.29	4.77	4.44	4.2	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.1	3.02	2.93	2.84	2.75		
	17	8.4	6.11	5.18	4.67	4.34	4.1	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3	2.92	2.83	2.75	2.65		
	18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.6	3.51	3.37	3.23	3.08	3	2.92	2.84	2.76	2.66	2.57		
	19	8.18	5.93	5.01	4.5	4.17	3.94	3.77	3.63	3.52	3.43	3.3	3.15	3	2.92	2.84	2.76	2.67	2.58	2.59		
	20	8.1	5.85	4.94	4.43	4.1	3.87	3.7	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42		
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.4	3.31	3.17	3.03	2.88	2.8	2.72	2.64	2.55	2.46	2.36			
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.5	2.4	2.31			
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.3	3.21	3.07	2.93	2.78	2.7	2.62	2.54	2.45	2.35	2.26			
24	7.82	5.61	4.72	4.22	3.9	3.67	3.5	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.4	2.31	2.21			
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.7	2.62	2.54	2.45	2.36	2.27	2.17			
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.81	2.66	2.58	2.5	2.42	2.33	2.23	2.13			
27	7.68	5.49	4.6	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.2	2.1			
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.9	2.75	2.6	2.52	2.44	2.35	2.26	2.17	2.06			
29	7.6	5.42	4.54	4.04	3.73	3.5	3.33	3.2	3.09	3	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03			
30	7.56	5.39	4.51	4.02	3.7	3.47	3.3	3.17	3.07	2.98	2.84	2.7	2.55	2.47	2.39	2.3	2.21	2.11	2.01			
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.8	2.66	2.52	2.37	2.29	2.2	2.11	2.02	1.92	1.8			
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.5	2.35	2.2	2.12	2.03	1.94	1.84	1.73	1.6			
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38			
∞	6.63	4.61	3.78	3.32	3.02	2.8	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.7	1.59	1.47	1.32	1			

Catatan: *F*_{0.75, *v*₁, *v*₂} = 1/*F*_{0.25, *v*₂, *v*₁}

Sumber: Diadaptasi dengan izin dari Tabel Biometrika untuk Ahli Statistik, Vol. 1, 3rd ed., by E.S. Pearson and H.O. Hartley, Cambridge University Press, Cambridge, 19

LAMPIRAN I
Titik persentase dari distribusi f (lanjutan)

		F_{α, v_1, v_2}																			
		Derajat Kebebasan Untuk Pembilang (v_1)																			
$v_2 \backslash v_1$		1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞	
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3	968.6	976.7	984.9	993.1	997.2	1001	1006	1010	1014	1018		
2	38.51	39	39.17	39.25	39.3	39.33	39.36	39.37	39.39	39.4	39.41	39.43	39.45	39.46	39.46	39.47	39.48	39.49	39.5		
3	17.44	16.04	15.44	15.1	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	13.9		
4	12.22	10.65	9.98	9.6	9.36	9.2	9.07	8.98	8.9	8.84	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26		
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02		
6	8.81	7.26	6.6	6.23	5.99	5.82	5.7	5.6	5.52	5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.9	4.85		
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.9	4.82	4.76	4.67	4.57	4.47	4.42	4.36	4.31	4.25	4.2	4.14		
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.3	4.2	4.1	4	3.95	3.89	3.84	3.78	3.73	3.67		
9	7.21	5.71	5.08	4.72	4.48	4.32	4.2	4.1	4.03	3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33		
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.2	3.14	3.08		
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.43	3.33	3.23	3.17	3.12	3.06	3	2.94	2.88		
12	6.55	5.1	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72		
13	6.41	4.97	4.35	4	3.77	3.6	3.48	3.39	3.31	3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.6		
14	6.3	4.86	4.24	3.89	3.66	3.5	3.38	3.29	3.21	3.15	3.06	2.96	2.86	2.76	2.7	2.64	2.59	2.52	2.46		
15	6.2	4.77	4.15	3.8	3.58	3.41	3.29	3.2	3.12	3.06	2.96	2.89	2.79	2.68	2.63	2.57	2.51	2.44	2.38		
16	6.12	4.69	4.08	3.73	3.5	3.34	3.22	3.12	3.05	2.99	2.89	2.79	2.68	2.62	2.56	2.5	2.44	2.38	2.32		
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.82	2.72	2.62	2.56	2.5	2.44	2.38	2.32	2.25		
18	5.98	4.56	3.95	3.61	3.38	3.22	3.1	3.01	2.93	2.87	2.77	2.67	2.56	2.5	2.44	2.38	2.32	2.26	2.19		
19	5.92	4.51	3.9	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.2	2.13		
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09		
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.8	2.73	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	2.04		
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.7	2.6	2.5	2.39	2.33	2.27	2.21	2.14	2.08	2		
23	5.75	4.35	3.75	3.41	3.18	3.02	2.9	2.81	2.73	2.67	2.57	2.47	2.36	2.3	2.24	2.18	2.12	2.05	1.98		
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.7	2.64	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94		
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.51	2.41	2.3	2.24	2.18	2.12	2.05	1.98	1.91		
26	5.66	4.27	3.67	3.33	3.1	2.94	2.82	2.73	2.65	2.59	2.49	2.39	2.28	2.22	2.16	2.09	2.03	1.95	1.88		
27	5.63	4.24	3.65	3.31	3.08	2.92	2.8	2.71	2.63	2.57	2.47	2.36	2.25	2.19	2.13	2.07	2	1.93	1.85		
28	5.61	4.22	3.63	3.29	3.06	2.9	2.78	2.69	2.61	2.55	2.45	2.34	2.23	2.17	2.11	2.05	1.98	1.91	1.83		
29	5.59	4.2	3.61	3.27	3.04	2.88	2.76	2.67	2.59	2.53	2.43	2.32	2.21	2.15	2.09	2.03	1.96	1.89	1.81		
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31	2.2	2.14	2.07	2.01	1.94	1.87	1.79		
40	5.42	4.05	3.46	3.13	2.9	2.74	2.62	2.53	2.45	2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.8	1.72	1.64		
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48		
120	5.15	3.8	3.23	2.89	2.67	2.52	2.39	2.3	2.22	2.16	2.05	1.94	1.82	1.76	1.69	1.61	1.53	1.43	1.31		
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1		

Catatan: $F_{0.90, v_1, v_2} = 1/F_{0.10, v_2, v_1}$

Catatan: $F_{0.90, v_1, v_2} = 1/F_{0.10, v_2, v_1}$

LAMPIRAN I
(Lanjutan)

		$F_{0.05, v_1, v_2}$																					
v_2	v_1	Derajat Kebebasan Untuk Pembilang (v_1)																				∞	∞
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120					
1	161.4	199.5	215.7	224.6	230.2	234	236.8	238.9	240.5	241.9	243.9	245.9	248	249.1	250.1	251.1	252.2	253.3	254.3	254.3			
2	18.51	19	19.16	19.25	19.3	19.33	19.35	19.37	19.38	19.4	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.5	19.5			
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.7	8.66	8.64	8.62	8.59	8.57	8.55	8.53	8.53			
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6	5.96	5.91	5.86	5.8	5.77	5.75	5.72	5.69	5.66	5.63	5.63			
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.5	4.46	4.43	4.4	4.36	4.36			
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.1	4.06	4	3.94	3.87	3.84	3.81	3.77	3.74	3.7	3.67	3.67			
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.3	3.27	3.23	3.23			
8	5.32	4.46	4.07	3.84	3.69	3.58	3.5	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93	2.93			
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.9	2.86	2.83	2.79	2.75	2.71	2.71			
10	4.96	4.1	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.7	2.66	2.62	2.58	2.54	2.54			
11	4.84	3.98	3.59	3.36	3.2	3.09	3.01	2.95	2.9	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.4	2.4			
12	4.75	3.89	3.49	3.26	3.11	3	2.91	2.85	2.8	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.3	2.3			
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.6	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13			
14	4.6	3.74	3.34	3.11	2.96	2.85	2.76	2.7	2.65	2.6	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13	2.13			
15	4.54	3.68	3.29	3.06	2.9	2.79	2.71	2.64	2.59	2.54	2.48	2.4	2.33	2.29	2.25	2.2	2.16	2.11	2.07	2.07			
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01	1.96			
17	4.45	3.59	3.2	2.96	2.81	2.7	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.1	2.06	2.01	1.96	1.96			
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92	1.92			
19	4.38	3.52	3.13	2.9	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88	1.88			
20	4.35	3.49	3.1	2.87	2.71	2.6	2.51	2.45	2.39	2.35	2.28	2.2	2.12	2.08	2.04	1.99	1.95	1.9	1.84	1.84			
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.1	2.05	2.01	1.96	1.92	1.87	1.81	1.81			
22	4.3	3.44	3.05	2.82	2.66	2.55	2.46	2.4	2.34	2.3	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78	1.78			
23	4.28	3.42	3.03	2.8	2.64	2.53	2.44	2.37	2.32	2.27	2.2	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76	1.76			
24	4.26	3.4	3.01	2.78	2.62	2.51	2.42	2.36	2.3	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73	1.73			
25	4.24	3.39	2.99	2.76	2.6	2.49	2.4	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71	1.71			
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.9	1.85	1.8	1.75	1.69	1.69			
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.2	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67	1.67			
28	4.2	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65	1.65			
29	4.18	3.33	2.93	2.7	2.55	2.43	2.35	2.28	2.22	2.18	2.1	2.03	1.94	1.9	1.85	1.81	1.75	1.7	1.64	1.64			
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62	1.62			
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51	1.51			
60	4	3.15	2.76	2.53	2.37	2.25	2.17	2.1	2.04	1.99	1.92	1.84	1.75	1.7	1.65	1.59	1.53	1.47	1.39	1.39			
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.55	1.43	1.35	1.25	1.25			
∞	3.84	3	2.6	2.37	2.21	2.1	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1	1			
Catatan: $F_{0.95, v_1, v_2} = 1/F_{0.05, v_2, v_1}$																							

LAMPIRAN I

Titik persentase dari distribusi f (Lanjutan)

$v_1 \backslash v_2$		$F_{0.025, v_1, v_2}$																			
		Derajat Kebebasan Untuk Pembilang (v_1)																			
v_2	v_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞	
1	39,86	49,5	53,59	55,83	57,24	58,2	58,91	59,44	59,86	60,19	60,71	61,22	61,74	62	62,26	62,53	62,79	63,06	63,33	63,33	
2	8,53	9	9,16	9,24	9,29	9,33	9,35	9,37	9,38	9,39	9,41	9,42	9,44	9,45	9,46	9,47	9,47	9,48	9,49	9,49	
3	5,54	5,46	5,39	5,34	5,31	5,28	5,27	5,25	5,24	5,23	5,22	5,2	5,18	5,18	5,17	5,16	5,15	5,14	5,13	5,13	
4	4,54	4,32	4,19	4,11	4,05	4,01	3,98	3,95	3,94	3,92	3,9	3,87	3,84	3,83	3,82	3,8	3,79	3,78	3,76	3,76	
5	4,06	3,78	3,62	3,52	3,45	3,4	3,37	3,34	3,32	3,3	3,27	3,24	3,21	3,19	3,17	3,16	3,14	3,12	3,1	3,1	
6	3,78	3,46	3,29	3,18	3,11	3,05	3,01	2,98	2,96	2,94	2,9	2,87	2,84	2,82	2,8	2,78	2,76	2,74	2,72	2,72	
7	3,59	3,26	3,07	2,96	2,88	2,83	2,78	2,75	2,72	2,7	2,67	2,63	2,59	2,58	2,56	2,54	2,51	2,49	2,47	2,47	
8	3,46	3,11	2,92	2,81	2,73	2,67	2,62	2,59	2,56	2,54	2,5	2,46	2,42	2,4	2,38	2,36	2,34	2,32	2,29	2,29	
9	3,36	3,01	2,81	2,69	2,61	2,55	2,51	2,47	2,44	2,42	2,38	2,34	2,3	2,28	2,25	2,23	2,21	2,18	2,16	2,16	
10	3,29	2,92	2,73	2,61	2,52	2,46	2,41	2,38	2,35	2,32	2,28	2,24	2,2	2,18	2,16	2,13	2,11	2,08	2,06	2,06	
11	3,23	2,86	2,66	2,54	2,45	2,39	2,34	2,3	2,27	2,25	2,21	2,17	2,12	2,1	2,08	2,05	2,03	2	1,97	1,97	
12	3,18	2,81	2,61	2,48	2,39	2,33	2,28	2,24	2,21	2,19	2,15	2,1	2,05	2,01	1,98	1,96	1,93	1,9	1,88	1,85	
13	3,14	2,76	2,56	2,43	2,35	2,28	2,23	2,2	2,16	2,14	2,1	2,05	2,01	1,96	1,94	1,91	1,89	1,86	1,83	1,8	
14	3,1	2,73	2,52	2,39	2,31	2,24	2,19	2,15	2,12	2,1	2,05	2,01	1,96	1,92	1,89	1,87	1,85	1,82	1,79	1,76	
15	3,07	2,7	2,49	2,36	2,27	2,21	2,16	2,12	2,09	2,06	2,02	1,97	1,94	1,91	1,88	1,86	1,84	1,81	1,78	1,75	
16	3,05	2,67	2,46	2,33	2,24	2,18	2,13	2,09	2,06	2,03	1,99	1,94	1,89	1,86	1,84	1,81	1,78	1,75	1,72	1,69	
17	3,03	2,64	2,44	2,31	2,22	2,15	2,1	2,06	2,03	2	1,96	1,93	1,89	1,86	1,84	1,81	1,78	1,75	1,72	1,69	
18	3,01	2,62	2,42	2,29	2,2	2,13	2,08	2,04	2	1,98	1,93	1,89	1,86	1,84	1,81	1,78	1,75	1,72	1,69	1,66	
19	2,99	2,61	2,4	2,27	2,18	2,11	2,06	2,02	1,98	1,93	1,89	1,86	1,84	1,81	1,79	1,76	1,73	1,7	1,67	1,63	
20	2,97	2,59	2,38	2,25	2,16	2,09	2,04	2	1,96	1,94	1,89	1,86	1,84	1,81	1,79	1,74	1,71	1,68	1,64	1,61	
21	2,96	2,57	2,36	2,23	2,14	2,08	2,02	1,98	1,93	1,95	1,92	1,87	1,83	1,78	1,75	1,72	1,69	1,66	1,62	1,59	
22	2,95	2,56	2,35	2,22	2,13	2,06	2,01	1,97	1,93	1,9	1,86	1,81	1,76	1,73	1,7	1,67	1,64	1,6	1,57	1,55	
23	2,94	2,55	2,34	2,21	2,11	2,05	1,99	1,95	1,92	1,89	1,84	1,8	1,74	1,72	1,69	1,66	1,62	1,59	1,55	1,53	
24	2,93	2,54	2,33	2,19	2,1	2,04	1,98	1,94	1,91	1,88	1,83	1,78	1,73	1,7	1,67	1,64	1,61	1,57	1,53	1,51	
25	2,92	2,53	2,32	2,18	2,09	2,02	1,97	1,93	1,89	1,87	1,82	1,77	1,72	1,69	1,66	1,63	1,59	1,56	1,52	1,5	
26	2,91	2,52	2,31	2,17	2,08	2,01	1,96	1,92	1,88	1,86	1,81	1,76	1,71	1,68	1,65	1,61	1,58	1,54	1,5	1,49	
27	2,9	2,51	2,3	2,17	2,07	2	1,95	1,91	1,87	1,85	1,8	1,75	1,7	1,67	1,64	1,6	1,57	1,53	1,49	1,48	
28	2,89	2,5	2,29	2,16	2,06	2	1,94	1,9	1,87	1,84	1,79	1,74	1,69	1,66	1,63	1,59	1,56	1,52	1,48	1,48	
29	2,89	2,5	2,28	2,15	2,06	1,99	1,93	1,89	1,86	1,83	1,78	1,73	1,68	1,65	1,62	1,58	1,55	1,51	1,47	1,46	
30	2,88	2,49	2,28	2,14	2,03	1,98	1,93	1,88	1,85	1,82	1,77	1,72	1,67	1,64	1,61	1,57	1,54	1,5	1,46	1,46	
40	2,84	2,44	2,23	2,09	2	1,93	1,87	1,83	1,79	1,76	1,71	1,66	1,61	1,57	1,54	1,51	1,47	1,42	1,38	1,38	
60	2,79	2,39	2,18	2,04	1,95	1,87	1,82	1,77	1,74	1,71	1,66	1,6	1,54	1,51	1,48	1,44	1,4	1,35	1,29	1,29	
120	2,75	2,35	2,13	1,99	1,9	1,82	1,77	1,72	1,68	1,65	1,6	1,55	1,48	1,45	1,41	1,37	1,32	1,26	1,2	1,19	
∞	2,71	2,3	2,08	1,94	1,85	1,77	1,72	1,67	1,63	1,6	1,55	1,49	1,42	1,38	1,34	1,3	1,24	1,17	1,1	1,1	

<

Catatan: $F_{0.975, v_1, v_2} = 1/F_{0.025, v_2, v_1}$

LAMPIRAN I

(lanjutan)

		$F_{0.01, v_1, v_2}$																			
		Derajat Kebebasan Untuk Pembilang (v_1)																			
$v_2 \backslash v_1$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞		
1	5,83	7,5	8,2	8,58	8,82	8,98	9,1	9,19	9,26	9,32	9,41	9,49	9,58	9,63	9,67	9,71	9,76	9,8	9,85		
2	2,57	3	3,15	3,23	3,28	3,31	3,34	3,35	3,37	3,38	3,39	3,41	3,43	3,43	3,44	3,45	3,46	3,47	3,48		
3	2,02	2,28	2,36	2,39	2,41	2,42	2,43	2,44	2,44	2,44	2,45	2,46	2,46	2,47	2,47	2,47	2,47	2,47	2,47		
4	1,81	2	2,05	2,06	2,07	2,08	2,08	2,08	2,08	2,08	2,08	2,08	2,08	2,08	2,08	2,08	2,08	2,08	2,08		
5	1,69	1,85	1,88	1,89	1,89	1,89	1,89	1,89	1,89	1,89	1,89	1,89	1,88	1,88	1,88	1,88	1,87	1,87	1,87		
6	1,62	1,76	1,78	1,79	1,79	1,78	1,78	1,78	1,77	1,77	1,77	1,76	1,76	1,75	1,75	1,75	1,74	1,74	1,74		
7	1,57	1,7	1,72	1,72	1,71	1,71	1,7	1,7	1,7	1,69	1,68	1,68	1,67	1,67	1,66	1,66	1,65	1,65	1,65		
8	1,54	1,66	1,67	1,66	1,66	1,65	1,64	1,64	1,63	1,63	1,62	1,62	1,61	1,6	1,6	1,59	1,59	1,58	1,58		
9	1,51	1,62	1,63	1,63	1,62	1,61	1,6	1,6	1,59	1,59	1,58	1,57	1,56	1,56	1,55	1,55	1,54	1,53	1,53		
10	1,49	1,6	1,6	1,59	1,59	1,58	1,57	1,56	1,56	1,55	1,54	1,53	1,52	1,52	1,51	1,51	1,5	1,49	1,48		
11	1,47	1,58	1,58	1,57	1,56	1,55	1,54	1,53	1,53	1,52	1,51	1,5	1,49	1,48	1,48	1,47	1,47	1,46	1,45		
12	1,46	1,56	1,56	1,55	1,54	1,53	1,52	1,51	1,51	1,5	1,49	1,48	1,47	1,46	1,45	1,45	1,44	1,43	1,42		
13	1,45	1,55	1,55	1,53	1,52	1,51	1,5	1,49	1,49	1,48	1,47	1,46	1,45	1,44	1,43	1,42	1,42	1,41	1,4		
14	1,44	1,53	1,53	1,52	1,51	1,5	1,49	1,48	1,47	1,46	1,45	1,44	1,43	1,42	1,41	1,41	1,4	1,39	1,38		
15	1,43	1,52	1,52	1,51	1,49	1,48	1,47	1,46	1,46	1,45	1,44	1,43	1,41	1,4	1,39	1,38	1,37	1,36	1,36		
16	1,42	1,51	1,51	1,5	1,48	1,47	1,46	1,45	1,44	1,44	1,43	1,41	1,4	1,39	1,38	1,37	1,36	1,35	1,34		
17	1,42	1,51	1,5	1,49	1,47	1,46	1,45	1,44	1,43	1,43	1,41	1,4	1,39	1,38	1,37	1,36	1,35	1,34	1,33		
18	1,41	1,5	1,49	1,48	1,46	1,45	1,44	1,43	1,42	1,42	1,4	1,39	1,38	1,37	1,36	1,35	1,34	1,33	1,32		
19	1,41	1,49	1,49	1,47	1,46	1,44	1,43	1,42	1,41	1,41	1,4	1,38	1,37	1,36	1,35	1,34	1,33	1,32	1,3		
20	1,4	1,49	1,48	1,47	1,45	1,44	1,43	1,42	1,41	1,4	1,39	1,37	1,36	1,35	1,34	1,33	1,32	1,31	1,29		
21	1,4	1,48	1,48	1,46	1,44	1,43	1,42	1,41	1,4	1,39	1,38	1,37	1,35	1,34	1,33	1,32	1,31	1,3	1,28		
22	1,4	1,48	1,47	1,45	1,44	1,42	1,41	1,4	1,39	1,38	1,37	1,36	1,34	1,33	1,32	1,31	1,3	1,29	1,28		
23	1,39	1,47	1,47	1,45	1,43	1,42	1,41	1,4	1,39	1,38	1,37	1,35	1,34	1,33	1,32	1,31	1,29	1,28	1,27		
24	1,39	1,47	1,46	1,44	1,43	1,41	1,4	1,39	1,38	1,38	1,36	1,35	1,33	1,32	1,31	1,3	1,29	1,28	1,26		
25	1,39	1,47	1,46	1,44	1,42	1,41	1,4	1,39	1,38	1,37	1,36	1,34	1,33	1,32	1,31	1,29	1,28	1,27	1,25		
26	1,38	1,46	1,45	1,44	1,42	1,41	1,39	1,38	1,37	1,37	1,35	1,34	1,32	1,31	1,3	1,29	1,28	1,26	1,25		
27	1,38	1,46	1,45	1,43	1,42	1,4	1,39	1,38	1,37	1,36	1,35	1,33	1,32	1,31	1,3	1,28	1,27	1,26	1,24		
28	1,38	1,46	1,45	1,43	1,41	1,4	1,39	1,38	1,37	1,36	1,34	1,33	1,31	1,3	1,29	1,28	1,27	1,25	1,24		
29	1,38	1,45	1,45	1,43	1,41	1,4	1,38	1,37	1,36	1,35	1,34	1,32	1,31	1,3	1,29	1,27	1,26	1,25	1,23		
30	1,38	1,45	1,44	1,42	1,41	1,39	1,38	1,37	1,36	1,35	1,34	1,32	1,3	1,29	1,28	1,27	1,26	1,24	1,23		
40	1,36	1,44	1,42	1,4	1,39	1,37	1,36	1,35	1,34	1,33	1,31	1,3	1,28	1,26	1,25	1,24	1,22	1,21	1,19		
60	1,35	1,42	1,41	1,38	1,37	1,35	1,33	1,32	1,31	1,3	1,29	1,27	1,25	1,24	1,22	1,21	1,19	1,17	1,15		
120	1,34	1,4	1,39	1,37	1,35	1,33	1,31	1,3	1,29	1,28	1,26	1,24	1,22	1,21	1,19	1,18	1,16	1,13	1,1		
∞	1,32	1,39	1,37	1,35	1,33	1,31	1,29	1,28	1,27	1,25	1,24	1,22	1,19	1,18	1,16	1,14	1,12	1,08	1		
Catatan: $F_{0.99, v_1, v_2} = 1/F_{0.01, v_2, v_1}$																					

LAMPIRAN II

Statistik d Durbin-Watson

Taraf nyata untuk d_L dan d_U 5%

n	$k' = 1$		$k' = 2$		$k' = 3$		$k' = 4$		$k' = 5$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_u
15	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21
16	1,1	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,9	1,71	0,78	1,9	0,67	2,1
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06
19	1,18	1,4	1,08	1,53	0,97	1,68	0,86	1,85	0,75	2,02
20	1,2	1,41	1,1	1,54	1	1,68	0,9	1,83	0,79	1,99
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,97
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,8	0,86	1,94
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,9	1,92
24	1,27	1,45	1,19	1,55	1,1	1,66	1,01	1,78	0,93	1,9
25	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
26	1,3	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,88
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	1,86
28	1,33	1,48	1,26	1,56	1,18	1,65	1,1	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,2	1,65	1,12	1,74	1,05	1,84
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83
31	1,36	1,5	1,3	1,57	1,23	1,65	1,16	1,74	1,09	1,83
32	1,37	1,5	1,31	1,57	1,24	1,65	1,18	1,73	1,11	1,82
33	1,38	1,51	1,32	1,58	1,26	1,65	1,19	1,73	1,13	1,81
34	1,39	1,51	1,33	1,58	1,27	1,65	1,21	1,73	1,15	1,81
35	1,4	1,52	1,34	1,58	1,28	1,65	1,22	1,73	1,16	1,8
36	1,41	1,52	1,35	1,59	1,29	1,65	1,24	1,73	1,18	1,8
37	1,42	1,53	1,36	1,59	1,31	1,66	1,25	1,72	1,19	1,8
38	1,43	1,54	1,37	1,59	1,32	1,66	1,26	1,72	1,21	1,79
39	1,43	1,54	1,38	1,6	1,33	1,66	1,27	1,72	1,22	1,79
40	1,44	1,54	1,39	1,6	1,34	1,66	1,29	1,72	1,23	1,79
45	1,48	1,57	1,43	1,62	1,38	1,67	1,34	1,72	1,29	1,78
50	1,5	1,59	1,46	1,63	1,42	1,67	1,38	1,72	1,34	1,77
55	1,53	1,6	1,49	1,64	1,45	1,68	1,41	1,72	1,38	1,77
60	1,55	1,62	1,51	1,65	1,48	1,69	1,44	1,73	1,41	1,77
65	1,57	1,63	1,54	1,66	1,5	1,7	1,47	1,73	1,44	1,77
70	1,58	1,64	1,55	1,67	1,52	1,7	1,49	1,74	1,46	1,77
75	1,6	1,65	1,57	1,68	1,54	1,71	1,51	1,74	1,49	1,77
80	1,61	1,66	1,59	1,69	1,56	1,72	1,53	1,74	1,51	1,77
85	1,62	1,67	1,6	1,7	1,57	1,72	1,55	1,75	1,52	1,77
90	1,63	1,68	1,61	1,7	1,59	1,73	1,57	1,75	1,54	1,78
95	1,64	1,69	1,62	1,71	1,6	1,73	1,58	1,75	1,56	1,78
100	1,65	1,69	1,63	1,72	1,61	1,74	1,59	1,76	1,57	1,78

Taraf nyata untuk d_L dan d_U 2,5%

n	$k' = 1$		$k' = 2$		$k' = 3$		$k' = 4$		$k' = 5$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_u
15	0,95	1,23	0,83	1,4	0,71	1,61	0,59	1,84	0,48	2,09
16	0,98	1,24	0,86	1,4	0,75	1,59	0,64	1,8	0,53	2,03
17	1,01	1,25	0,9	1,4	0,79	1,58	0,68	1,77	0,57	1,98
18	1,03	1,26	0,93	1,4	0,82	1,56	0,72	1,74	0,62	1,93
19	1,06	1,28	0,96	1,41	0,86	1,55	0,76	1,72	0,66	1,9
20	1,08	1,28	0,99	1,41	0,89	1,55	0,79	1,7	0,7	1,87
21	1,1	1,3	1,01	1,41	0,92	1,54	0,83	1,69	0,73	1,84
22	1,12	1,31	1,04	1,42	0,95	1,54	0,86	1,68	0,77	1,82
23	1,14	1,32	1,06	1,42	0,97	1,54	0,89	1,67	0,8	1,8
24	1,16	1,33	1,08	1,43	1	1,54	0,91	1,66	0,83	1,79
25	1,18	1,34	1,1	1,43	1,02	1,54	0,94	1,65	0,86	1,77
26	1,19	1,35	1,12	1,44	1,04	1,54	0,96	1,65	0,88	1,76
27	1,21	1,36	1,13	1,44	1,06	1,54	0,99	1,64	0,91	1,75
28	1,22	1,37	1,15	1,45	1,08	1,54	1,01	1,64	0,93	1,74
29	1,24	1,38	1,17	1,45	1,1	1,54	1,03	1,63	0,96	1,73
30	1,25	1,38	1,18	1,46	1,12	1,54	1,05	1,63	0,98	1,73
31	1,26	1,39	1,2	1,47	1,13	1,55	1,07	1,63	1	1,72
32	1,27	1,4	1,21	1,47	1,15	1,55	1,08	1,63	1,02	1,71
33	1,28	1,41	1,22	1,48	1,16	1,55	1,1	1,63	1,04	1,71
34	1,29	1,41	1,24	1,48	1,17	1,55	1,12	1,63	1,06	1,7
35	1,3	1,42	1,25	1,48	1,19	1,55	1,13	1,63	1,07	1,7
36	1,31	1,43	1,26	1,49	1,2	1,56	1,15	1,63	1,09	1,7
37	1,32	1,43	1,27	1,49	1,21	1,56	1,16	1,62	1,1	1,7
38	1,33	1,44	1,28	1,5	1,23	1,56	1,17	1,62	1,12	1,7
39	1,34	1,44	1,29	1,5	1,24	1,56	1,19	1,63	1,13	1,69
40	1,35	1,45	1,3	1,51	1,25	1,57	1,2	1,63	1,15	1,69
45	1,39	1,48	1,34	1,53	1,3	1,58	1,25	1,63	1,21	1,69
50	1,42	1,5	1,38	1,54	1,34	1,59	1,3	1,64	1,26	1,69
55	1,45	1,52	1,41	1,56	1,37	1,6	1,33	1,64	1,3	1,69
60	1,47	1,54	1,44	1,57	1,4	1,61	1,37	1,65	1,33	1,69
65	1,49	1,55	1,46	1,59	1,43	1,62	1,4	1,66	1,36	1,69
70	1,51	1,57	1,48	1,6	1,45	1,63	1,42	1,66	1,39	1,7
75	1,53	1,58	1,5	1,61	1,47	1,64	1,45	1,67	1,42	1,7
80	1,54	1,59	1,52	1,62	1,49	1,65	1,47	1,67	1,44	1,7
85	1,56	1,6	1,53	1,63	1,51	1,65	1,49	1,68	1,46	1,71
90	1,57	1,61	1,55	1,64	1,53	1,66	1,5	1,69	1,48	1,71
95	1,58	1,62	1,56	1,65	1,54	1,67	1,52	1,69	1,5	1,71
100	1,59	1,63	1,57	1,65	1,55	1,67	1,53	1,7	1,51	1,72

Taraf nyata untuk d_L dan d_U 1%

n	$k' = 1$		$k' = 2$		$k' = 3$		$k' = 4$		$k' = 5$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	0,81	1,07	0,7	1,25	0,59	1,46	0,49	1,7	0,39	1,96
16	0,84	1,09	0,74	1,25	0,63	1,44	0,53	1,66	0,44	1,9
17	0,87	1,1	0,77	1,25	0,67	1,43	0,57	1,63	0,48	1,85
18	0,9	1,12	0,8	1,26	0,71	1,42	0,61	1,6	0,52	1,8
19	0,93	1,13	0,83	1,26	0,74	1,41	0,65	1,58	0,56	1,77
20	0,95	1,15	0,86	1,27	0,77	1,41	0,68	1,57	0,6	1,74
21	0,97	1,16	0,89	1,27	0,8	1,41	0,72	1,55	0,63	1,71
22	1	1,17	0,91	1,28	0,83	1,4	0,75	1,54	0,66	1,69
23	1,02	1,19	0,94	1,29	0,86	1,4	0,77	1,53	0,7	1,67
24	1,04	1,2	0,96	1,3	0,88	1,41	0,8	1,53	0,72	1,66
25	1,05	1,21	0,98	1,3	0,9	1,41	0,83	1,52	0,75	1,65
26	1,07	1,22	1	1,31	0,93	1,41	0,85	1,52	0,78	1,64
27	1,09	1,23	1,02	1,32	0,95	1,41	0,88	1,51	0,81	1,63
28	1,1	1,24	1,04	1,32	0,97	1,41	0,9	1,51	0,83	1,62
29	1,12	1,25	1,05	1,33	0,99	1,42	0,92	1,51	0,85	1,61
30	1,13	1,26	1,07	1,34	1,01	1,42	0,94	1,51	0,88	1,61
31	1,15	1,27	1,08	1,34	1,02	1,42	0,96	1,51	0,9	1,6
32	1,16	1,28	1,1	1,35	1,04	1,43	0,98	1,51	0,92	1,6
33	1,17	1,29	1,11	1,36	1,05	1,43	1	1,51	0,94	1,59
34	1,18	1,3	1,13	1,36	1,07	1,43	1,01	1,51	0,95	1,58
35	1,19	1,31	1,14	1,37	1,08	1,44	1,03	1,51	0,97	1,59
36	1,21	1,32	1,15	1,38	1,1	1,44	1,04	1,51	0,99	1,59
37	1,22	1,32	1,16	1,38	1,11	1,45	1,06	1,51	1	1,59
38	1,23	1,33	1,18	1,39	1,12	1,45	1,07	1,52	1,02	1,58
39	1,24	1,34	1,19	1,39	1,14	1,45	1,09	1,52	1,03	1,58
40	1,25	1,34	1,2	1,4	1,15	1,46	1,1	1,52	1,05	1,58
45	1,29	1,38	1,24	1,42	1,2	1,48	1,16	1,53	1,11	1,58
50	1,32	1,4	1,28	1,45	1,24	1,49	1,2	1,54	1,16	1,59
55	1,36	1,43	1,32	1,47	1,28	1,51	1,25	1,55	1,21	1,59
60	1,38	1,45	1,35	1,48	1,32	1,52	1,28	1,56	1,25	1,6
65	1,41	1,47	1,38	1,5	1,35	1,53	1,31	1,57	1,28	1,61
70	1,43	1,49	1,4	1,52	1,37	1,55	1,34	1,58	1,31	1,61
75	1,45	1,5	1,42	1,53	1,39	1,56	1,37	1,59	1,34	1,62
80	1,47	1,52	1,44	1,54	1,42	1,57	1,39	1,6	1,36	1,62
85	1,48	1,53	1,46	1,55	1,43	1,58	1,41	1,6	1,39	1,63
90	1,5	1,54	1,47	1,56	1,45	1,59	1,43	1,61	1,41	1,64
95	1,51	1,55	1,49	1,57	1,47	1,6	1,45	1,62	1,42	1,64
100	1,52	1,56	1,5	1,58	1,48	1,6	1,46	1,63	1,44	1,65

LAMPIRAN III

Nilai kritis t untuk $\alpha = 0,05$ dan $\alpha = 0,01$ dalam prosedur Uji Peringat Bertanda Wilcoxon

n	Pengujian dua arah		Pengujian satu arah	
	0,05	0,01	0,05	0,01
4				
5			0	
6	0		2	
7	2		3	0
8	3	0	5	1
9	5	1	8	3
10	8	3	10	5
11	10	5	13	7
12	13	7	17	9
13	17	9	21	12
14	21	12	25	15
15	25	15	30	19
16	29	19	35	23
17	34	23	41	27
18	40	27	47	32
19	46	32	53	37
20	52	37	60	43
21	58	42	67	49
22	65	48	75	55
23	73	54	83	62
24	81	61	91	69
25	89	68	100	76
26	98	75	110	84
27	107	83	119	92
28	116	91	130	101
29	126	100	140	110
30	137	109	151	120
31	147	118	163	130
32	159	128	175	140
33	170	138	187	151
34	182	148	200	162
35	195	159	213	173
40	264	220	286	238
50	434	373	466	397
60	648	567	690	600
70	907	805	960	846
80	1211	1086	1276	1136
90	1560	1410	1638	1471
100	1955	1779	2045	1850

LAMPIRAN IV

Distribusi v dalam pengujian Mann-Whitney

Tabel pengujian satu arah

Nilai v kritis: $\alpha = 0,05$ untuk pengujian satu-arah (dan $\alpha = 0,10$ untuk pengujian dua-arah)

n_1	n_2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1																				0	0
2						0	0	0	1	1	1	1	2	2	2	3	3	3	4	4	4
3				0	0	1	2	2	3	3	4	5	5	6	7	7	8	9	9	10	11
4				0	1	2	3	4	5	6	7	8	9	10	11	12	13	15	16	17	18
5		0	1	2	4	5	6	8	9	11	12	13	15	16	18	19	20	22	23	25	
6		0	2	3	5	7	8	10	12	14	16	17	19	21	23	25	26	28	30	32	
7		0	2	4	6	8	11	13	15	17	19	21	24	26	28	30	33	35	37	39	
8		1	3	5	8	10	13	15	18	20	23	26	28	31	33	36	39	41	44	47	
9		1	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51	54	
10		1	4	7	11	14	17	20	24	27	31	34	37	41	44	48	51	55	58	62	
11		1	5	8	12	16	19	23	27	31	34	38	42	46	50	54	57	61	65	69	
12		2	5	9	13	17	21	26	30	34	38	42	47	51	55	59	64	68	72	77	
13		2	6	10	15	19	24	28	33	37	42	47	51	56	61	65	70	75	80	84	
14		2	7	11	16	21	26	31	36	41	46	51	56	61	66	71	77	82	87	92	
15		3	7	12	18	23	28	33	39	44	50	55	61	66	72	77	83	88	94	100	
16		3	8	14	19	25	30	36	42	48	54	60	65	71	77	83	89	95	101	107	
17		3	9	15	20	26	33	39	45	51	57	64	70	77	82	89	96	102	109	115	
18		4	9	16	22	28	35	41	48	55	61	68	75	82	88	95	102	109	116	123	
19		0	4	10	17	23	30	37	44	51	58	65	72	80	87	94	101	109	116	123	130
20		0	4	11	18	25	32	39	47	54	62	69	77	84	92	100	107	115	123	130	138

Nilai v kritis: $\alpha = 0,01$ untuk pengujian satu-arah (dan $\alpha = 0,02$ untuk pengujian dua-arah)

n_1	n_2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1																					
2														0	0	0	0	0	0	1	1
3								0	0	1	1	1	2	2	2	3	3	4	4	4	5
4					0	1	1	2	3	3	4	5	5	6	7	7	8	9	9	9	10
5				0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
6			1	2	3	4	6	7	8	9	11	12	13	15	16	18	19	20	22		
7		0	1	3	4	6	7	9	11	12	14	16	17	19	21	23	24	26	28		
8		0	2	4	6	7	9	11	13	15	17	20	22	24	26	28	30	32	34		
9		1	3	5	7	9	11	14	16	18	21	23	26	28	31	33	36	38	40		
10		1	3	6	8	11	13	16	19	22	24	27	30	33	36	38	41	44	47		
11		1	4	7	9	12	15	18	22	25	28	31	34	37	41	44	47	50	53		
12			2	5	8	11	14	17	21	24	28	31	35	38	42	46	49	53	56	60	
13		0	2	5	9	12	16	20	23	27	31	35	39	43	47	51	55	59	63	67	
14		0	2	6	10	13	17	22	26	30	34	38	43	47	51	56	60	65	69	73	
15		0	3	7	11	15	19	24	28	33	37	42	47	51	56	61	66	70	75	80	
16		0	3	7	12	16	21	26	31	36	41	46	51	56	61	66	71	76	82	87	
17		0	4	8	13	18	23	28	33	38	44	49	55	60	66	71	77	82	88	93	
18		0	4	6	14	19	24	30	36	41	47	53	59	65	70	76	82	88	94	100	
19		1	4	6	15	20	26	32	38	44	50	56	63	69	75	82	88	94	101	107	
20		1	5	10	16	22	28	34	40	47	53	60	67	73	80	87	93	100	107	114	

Nilai v kritis: $\alpha = 0,05$ untuk pengujian dua-arah (dan $\alpha = 0,025$ untuk pengujian satu-arah)

n_1	n_2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1																					
2																					
3						0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8
4					0	1	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19
5				0	1	2	3	5	6	8	10	11	13	14	16	17	19	21	22	24	25
6				1	2	3	5	6	8	10	12	14	16	18	20	22	24	26	28	30	32
7				1	3	5	6	8	10	12	15	17	19	22	24	26	29	31	34	36	38
8		0	2	4	6	8	10	13	15	17	20	23	26	29	31	34	37	39	42	45	48
9		0	2	4	7	10	12	15	17	20	23	26	29	33	36	39	42	45	48	52	55
10		0	3	5	8	11	14	17	20	23	26	29	33	36	39	42	45	48	51	55	58
11		0	3	6	9	13	16	20	23	26	30	33	37	40	44	47	51	55	58	62	67
12		1	4	7	11	14	18	22	26	29	33	37	41	45	49	53	57	61	65	69	
13		1	4	8	12	16	20	24	28	33	37	41	45	50	54	59	63	67	72	76	
14		1	5	9	13	17	22	26	31	36	40	45	50	55	59	64	67	74	78	83	
15		1	5	10	14	16	24	29	34	39	44	49	54	59	64	70	75	80	85	90	
16		1	6	11	15	21	26	31	37	42	47	53	59	64	70	75	81	86	92	98	
17		2	6	11	17	22	28	34	39	45	51	57	63	67	75	81	87	93	99	105	
18		2	7	12	18	24	30	36	42	48	55	61	67	74	81	86	93	99	106	112	
19		2	7	13	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	119	
20		2	8	13	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127	

Nilai v kritis: $\alpha = 0,01$ untuk pengujian satu-arah (dan $\alpha = 0,02$ untuk pengujian dua-arah)

n_1	n_2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1																					
2																				0	0
3										0	0	1	1	1	1	2	2	2	2	3	3
4							0	0	1	1	2	3	4	5	6	7	8	9	10	11	12
5						0	1	2	3	4	5	6	7	9	10	11	12	13	15	16	17
6					0	1	3	4	6	7	9	10	12	13	15	16	18	19	21	22	24
7					0	2	4	6	7	9	11	13	15	17	18	20	22	24	26	28	30
8					1	3	5	7	9	11	13	16	18	20	22	24	27	29	31	33	36
9			0	1	3	5	7	9	11	13	16	18	21	24	27	30	33	36	39	42	45
10			0	2	4	6	9	11	13	16	18	21	24	27	30	33	36	39	42	45	48
11			0	2	5	7	10	13	16	18	21	24	27	31	34	37	41	44	47	51	54
12			1	3	6	9	12	15	18	21	24	27	31	34	38	42	45	49	53	56	60
13			1	3	7	10	13	17	20	24	27	31	34	38	42	46	50	54	58	63	67
14			1	4	7	11	15	18	22	26	30	34	38	42	46	51	55	60	64	69	73
15			2	5	8	12	16	20	24	29	33	37	42	46	51	55	60	64	69	74	79
16			2	5	9	13	18	22	27	31	36	41	45	50	55	60	65	70	75	81	86
17			2	6	10	15	19	24	29	34	39	44	49	54	60	65	70	75	81	87	92
18			2	6	11	16	21	26	31	37	42	47	53	58	64	70	75	81	87	93	99
19		0	3	7	12	17	22	28	33	39	45	51	56	63	69	74	81	87	93	99	
20		0	3	8	13	18	24	30	36	42	48	54	60	67	73	79	86	92	99	105	

LAMPIRAN V

Nilai Kritis r dalam *Runs Test* untuk Melihat keacakan

Setiap nilai r sampel yang sama atau lebih atau lebih kecil dari angka table (a); atau nilai r sample yang sama atau lebih besar dari angka pada table (b) menyebabkan penolakan H_0 pada taraf nyata sebesar 0,05.

Tabel r (a)

	n_2																			
n_1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
2											2	2	2	2	2	2	2	2	2	
3					2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	
4				1	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4	
5			2	2	3	3	3	3	3	4	4	4	4	4	4	4	5	5	5	
6		2	2	3	3	3	3	4	4	4	4	5	5	5	5	5	5	6	6	
7		2	2	3	3	3	4	4	5	5	5	5	5	6	6	6	6	6	6	
8		2	3	3	3	4	4	5	5	5	6	6	6	6	6	7	7	7	7	
9		2	3	3	4	4	5	5	5	6	6	6	7	7	7	7	8	8	8	
10		2	3	3	4	5	5	5	6	6	7	7	7	7	8	8	8	8	9	
11		2	3	4	4	5	5	6	6	7	7	7	8	8	8	9	9	9	9	
12	2	2	3	4	4	5	6	6	7	7	7	8	8	8	9	9	9	10	10	
13	2	2	3	4	5	5	6	6	7	7	8	8	9	9	9	10	10	10	10	
14	2	2	3	4	5	5	6	7	7	8	8	9	9	9	10	10	10	11	11	
15	2	3	3	4	5	6	6	7	7	8	8	9	9	10	10	11	11	11	12	
16	2	3	4	4	5	6	6	7	8	8	9	9	10	10	11	11	11	12	12	
17	2	3	4	4	5	6	7	7	8	9	9	10	10	11	11	11	12	12	13	
18	2	3	4	5	5	6	7	8	8	9	9	10	10	11	11	12	12	13	13	
19	2	3	4	5	6	6	7	8	8	9	10	10	11	11	12	12	13	13	13	
20	2	3	4	5	6	6	7	8	9	9	10	10	11	12	12	13	13	13	14	

LAMPIRAN V

Nilai Kritis r dalam *Runs Test* untuk Melihat keacakan (Lanjutan)

Tabel r (b)

	n_2																			
n_1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
2											2	2	2	2	2	2	2	2	2	
3					2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	
4				1	2	2	3	3	3	3	3	3	3	3	3	4	4	4	4	
5			2	2	3	3	3	3	3	4	4	4	4	4	4	4	5	5	5	
6		2	2	3	3	3	3	4	4	4	4	5	5	5	5	5	5	6	6	
7		2	2	3	3	3	4	4	5	5	5	5	5	6	6	6	6	6	6	
8		2	3	3	3	4	4	5	5	5	6	6	6	6	6	7	7	7	7	
9		2	3	3	4	4	5	5	5	6	6	6	7	7	7	7	8	8	8	
10		2	3	3	4	5	5	5	6	6	7	7	7	7	8	8	8	8	9	
11		2	3	4	4	5	5	6	6	7	7	7	8	8	8	9	9	9	9	
12	2	2	3	4	4	5	6	6	7	7	7	8	8	8	9	9	9	10	10	
13	2	2	3	4	5	5	6	6	7	7	8	8	9	9	9	10	10	10	10	
14	2	2	3	4	5	5	6	7	7	8	8	9	9	9	10	10	10	11	11	
15	2	3	3	4	5	6	6	7	7	8	8	9	9	10	10	11	11	11	12	
16	2	3	4	4	5	6	6	7	8	8	9	9	10	10	11	11	11	12	12	
17	2	3	4	4	5	6	7	7	8	9	9	10	10	11	11	11	12	12	13	
18	2	3	4	5	5	6	7	8	8	9	9	10	10	11	11	12	12	13	13	
19	2	3	4	5	6	6	7	8	8	9	10	10	11	11	12	12	13	13	13	
20	2	3	4	5	6	6	7	8	9	9	10	10	11	12	12	13	13	13	14	

Sub- grup	Bagan Rata-rata					Bagan untuk Simpangan Baku					Faktor untuk Garis Tengah		Bagan untuk Range				
	Faktor-faktor untuk Batas Kontrol			Faktor-faktor untuk Garis Tengah		Faktor-faktor untuk Batas Kontrol					d_2	$1/d_2$	d_3	D_1	D_2	D_3	D_4
	A	A_2	A_3	c_4	$1/c_4$	B_3	B_4	B_5	B_6								
2	2,121	1,880	2,659	0,7979	1,2533	0	3,267	0	2,606	1,128	0,8865	0,853	0	3,686	0	3,267	
3	1,732	1,023	1,954	0,8862	1,1284	0	2,568	0	2,276	1,693	0,5907	0,888	0	4,358	0	2,574	
4	1,500	0,729	1,628	0,9213	1,0854	0	2,266	0	2,088	2,059	0,4857	0,880	0	4,698	0	2,282	
5	1,342	0,577	1,427	0,9400	1,0638	0	2,089	0	1,964	2,326	0,4299	0,864	0	4,918	0	2,114	
6	1,225	0,483	1,287	0,9515	1,0510	0,030	1,970	0,029	1,874	2,534	0,3946	0,848	0	5,078	0	2,004	
7	1,134	0,419	1,182	0,9594	1,0423	0,118	1,882	0,113	1,806	2,704	0,3698	0,833	0,204	5,204	0,076	1,924	
8	1,061	0,373	1,099	0,9650	1,0363	0,185	1,815	0,179	1,751	2,847	0,3512	0,820	0,388	5,306	0,136	1,864	
9	1,000	0,337	1,032	0,9693	1,0317	0,239	1,761	0,232	1,707	2,970	0,3367	0,808	0,547	5,393	0,184	1,816	
10	0,949	0,308	0,975	0,9727	1,0281	0,284	1,716	0,276	1,669	3,078	0,3249	0,797	0,687	5,469	0,223	1,777	
11	0,905	0,285	0,927	0,9754	1,0252	0,321	1,679	0,313	1,637	3,173	0,3152	0,787	0,811	5,535	0,256	1,744	
12	0,866	0,266	0,886	0,9776	1,0229	0,354	1,646	0,346	1,610	3,258	0,3069	0,778	0,922	5,594	0,283	1,717	
13	0,832	0,249	0,850	0,9794	1,0210	0,382	1,618	0,374	1,585	3,336	0,2998	0,770	1,025	5,647	0,307	1,693	
14	0,802	0,235	0,817	0,9810	1,0194	0,406	1,594	0,399	1,563	3,407	0,2935	0,763	1,118	5,696	0,328	1,672	
15	0,775	0,223	0,789	0,9823	1,0180	0,428	1,572	0,421	1,544	3,472	0,2880	0,756	1,203	5,741	0,347	1,653	
16	0,750	0,212	0,763	0,9835	1,0168	0,448	1,552	0,440	1,526	3,532	0,2831	0,750	1,282	5,782	0,363	1,637	
17	0,728	0,203	0,739	0,9845	1,0157	0,466	1,534	0,458	1,511	3,588	0,2787	0,744	1,356	5,820	0,378	1,622	
18	0,707	0,194	0,718	0,9854	1,0148	0,482	1,518	0,475	1,496	3,640	0,2747	0,739	1,424	5,856	0,391	1,608	
19	0,688	0,187	0,698	0,9862	1,0140	0,497	1,503	0,490	1,483	3,689	0,2711	0,734	1,487	5,891	0,403	1,597	
20	0,671	0,180	0,680	0,9869	1,0133	0,510	1,490	0,504	1,470	3,735	0,2677	0,729	1,549	5,921	0,415	1,585	
21	0,655	0,173	0,663	0,9876	1,0126	0,523	1,477	0,516	1,459	3,778	0,2647	0,724	1,605	5,951	0,425	1,575	
22	0,640	0,167	0,647	0,9882	1,0119	0,534	1,466	0,528	1,448	3,819	0,2618	0,720	1,659	5,979	0,434	1,567	
23	0,626	0,162	0,633	0,9887	1,0114	0,545	1,455	0,539	1,438	3,858	0,2592	0,716	1,710	6,006	0,443	1,557	
24	0,612	0,157	0,619	0,9892	1,0109	0,555	1,445	0,549	1,429	3,895	0,2567	0,712	1,759	6,031	0,451	1,548	
25	0,600	0,153	0,606	0,9896	1,0105	0,565	1,435	0,559	1,420	3,931	0,2544	0,708	1,806	6,056	0,459	1,541	

LAMPIRAN VII
Faktor untuk Batas Toleransi Normal Dua Sisi

n	90% Keyakinan Bahwa Persentase Populasi Antara Batas			95% Keyakinan Bahwa Persentase Populasi Antara Batas			99% Keyakinan Bahwa Persentase Populasi Antara Batas A		
	90%	95%	99%	90%	95%	99%	90%	95%	99%
2	15,98	18,80	24,17	32,02	37,67	48,43	160,2	188,5	242,3
3	5,847	6,919	8,974	8,380	9,916	12,86	18,93	22,40	29,06
4	4,166	4,943	6,440	5,369	6,370	8,299	9,398	11,15	14,53
5	3,494	4,152	5,423	4,275	5,079	6,634	6,612	7,855	10,26
6	3,131	3,723	4,870	3,712	4,414	5,775	5,337	6,345	8,301
7	2,902	3,452	4,521	3,369	4,007	5,248	4,613	5,448	7,187
8	2,743	3,264	4,278	3,136	3,732	4,891	4,147	4,936	6,468
9	2,626	3,125	4,098	2,967	3,532	4,633	3,822	4,550	5,966
10	2,535	3,018	3,959	2,829	3,379	4,433	3,582	4,265	5,594
11	2,463	2,933	3,849	2,737	3,259	4,277	3,397	4,045	5,308
12	2,404	2,863	3,758	2,655	3,162	4,150	3,250	3,870	5,079
13	2,355	2,805	3,682	2,587	3,081	4,044	3,130	3,727	4,893
14	2,314	2,756	3,618	2,529	3,012	3,955	3,029	3,608	4,737
15	2,278	2,713	3,562	2,480	2,954	3,878	2,945	3,507	4,605
16	2,246	2,676	3,514	2,437	2,903	3,812	2,872	3,421	4,492
17	2,219	2,643	3,471	2,400	2,858	3,754	2,808	3,345	4,393
18	2,194	2,614	3,433	2,366	2,819	3,702	2,753	3,279	4,307
19	2,172	2,588	3,399	2,337	2,784	3,656	2,703	3,221	4,230
20	2,152	2,564	3,368	2,310	2,752	3,615	2,659	3,168	4,161
21	2,135	2,543	3,340	2,286	2,723	3,577	2,620	3,121	4,100
22	2,118	2,524	3,315	2,264	2,697	3,543	2,584	3,078	4,044
23	2,103	2,506	3,292	2,244	2,673	3,512	2,551	3,040	3,993
24	2,089	2,489	3,270	2,225	2,651	3,483	2,522	3,004	3,947
25	2,077	2,474	3,251	2,208	2,631	3,457	2,494	2,972	3,904
26	2,065	2,460	3,232	2,193	2,612	3,432	2,469	2,941	3,865
27	2,054	2,447	3,215	2,178	2,595	3,409	2,446	2,914	3,828
28	2,044	2,435	3,199	2,164	2,579	3,388	2,424	2,888	3,794
29	2,034	2,424	3,184	2,152	2,554	3,368	2,404	2,864	3,763
30	2,025	2,413	3,170	2,140	2,549	3,350	2,385	2,841	3,733
35	1,988	2,368	3,112	2,090	2,490	3,272	2,306	2,748	3,611
40	1,959	2,334	3,066	2,052	2,445	3,213	2,247	2,677	3,518
50	1,916	2,284	3,001	1,996	2,379	3,126	2,162	2,576	3,385
60	1,887	2,248	2,955	1,958	2,333	3,066	2,103	2,506	3,293
80	1,848	2,202	2,894	1,907	2,272	2,986	2,026	2,414	3,173
100	1,822	2,172	2,854	1,874	2,233	2,934	1,977	2,355	3,096
200	1,764	2,102	2,762	1,798	2,143	2,816	1,865	2,222	2,921
500	1,717	2,046	2,689	1,737	2,070	2,721	1,777	2,117	2,783
1000	1,695	2,019	2,654	1,709	2,036	2,676	1,736	2,068	2,718
∞	1,645	1,960	2,576	1,645	1,960	2,576	1,645	1,960	2,576

LAMPIRAN VII (lanjutan)

Faktor-faktor Batas Toleransi Normal Satu Sisi

<i>n</i>	90% Keyakinan Bahwa Persentase Penduduk di Bawah (Di Atas) Batas			95% Keyakinan Bahwa Persentase Penduduk di Bawah (Di Atas) Batas			95% Keyakinan Bahwa Persentase Penduduk di Bawah (Di Atas) Batas		
	90%	95%	99%	90%	95%	99%	90%	95%	99%
3	4,258	5,310	7,340	6,158	7,655	10,552			
4	3,187	3,957	5,437	4,163	5,145	7,042			
5	2,742	3,400	4,666	3,407	4,202	5,741			
6	2,494	3,091	4,242	3,006	3,707	5,062	4,408	5,409	7,334
7	2,333	2,894	3,972	2,755	3,399	4,641	3,856	4,730	6,411
8	2,219	2,755	3,783	2,582	3,188	4,353	3,496	4,287	5,811
9	2,133	2,649	3,641	2,454	3,031	4,143	3,242	3,971	5,389
10	2,065	2,568	3,532	2,355	2,911	3,981	3,048	3,739	5,075
11	2,012	2,503	3,444	2,275	2,815	3,852	2,897	3,557	4,828
12	1,966	2,448	3,371	2,210	2,736	3,747	2,773	3,410	4,633
13	1,928	2,403	3,310	2,155	2,670	3,659	2,677	3,290	4,472
14	1,895	2,363	3,257	2,108	2,614	3,585	2,592	3,189	4,336
15	1,866	2,329	3,212	2,068	2,566	3,520	2,521	3,102	4,224
16	1,842	2,299	3,172	2,032	2,523	3,463	2,458	3,028	4,124
17	1,820	2,272	3,136	2,001	2,486	3,415	2,405	2,962	4,038
18	1,800	2,249	3,106	1,974	2,453	3,370	2,357	2,906	3,961
19	1,781	2,228	3,078	1,949	2,423	3,331	2,315	2,855	3,893
20	1,765	2,208	3,052	1,926	2,396	3,295	2,275	2,807	3,832
21	1,750	2,190	3,028	1,905	2,371	3,262	2,241	2,768	3,776
22	1,736	2,174	3,007	1,887	2,350	3,233	2,208	2,729	3,727
23	1,724	2,159	2,987	1,869	2,329	3,206	2,179	2,693	3,680
24	1,712	2,145	2,969	1,853	2,309	3,181	2,154	2,663	3,638
25	1,702	2,132	2,952	1,838	2,292	3,158	2,129	2,632	3,601
30	1,657	2,080	2,884	1,778	2,220	3,064	2,029	2,516	3,446
35	1,623	2,041	2,833	1,732	2,166	2,994	1,957	2,431	3,334
40	1,598	2,010	2,793	1,697	2,126	2,941	1,902	2,365	3,250
45	1,577	1,986	2,762	1,669	2,092	2,897	1,857	2,313	3,181
50	1,560	1,965	2,735	1,646	2,065	2,863	1,821	2,296	3,124

LAMPIRAN VIII

Tabel Kuantil Uji Kruskal-Wallis untuk Ukuran Sampel Kecil

Ukuran Sampel	$W_{0,9}$	$W_{0,95}$	$W_{0,99}$
2, 2, 2	3,7143	4,5714	4,5714
3, 2, 1	3,8571	4,2857	4,2857
3, 2, 2	4,4643	4,5	5,3571
3, 3, 1	4	4,5714	5,1429
3, 3, 2	4,25	5,1389	6,25
3, 3, 3	4,6	5,0667	6,4889
4, 2, 1	4,0179	4,8214	4,8214
4, 2, 2	4,1667	5,125	6
4, 3, 1	3,8889	5	5,8333
4, 3, 2	4,4444	5,4	6,3
4, 3, 3	4,7	5,7273	6,7091
4, 4, 1	4,0667	4,8667	6,1667
4, 4, 2	4,4455	5,2364	6,8727
4, 4, 3	4,773	5,5758	7,1364
4, 4, 4	4,5	5,6538	7,5385
5, 2, 1	4,05	4,45	5,25
5, 2, 2	4,2933	5,04	6,1333
5, 3, 1	3,84	4,8711	6,4
5, 3, 2	4,4946	5,1055	6,8218
5, 3, 3	4,4121	5,5152	6,9818
5, 4, 1	3,96	4,86	6,84
5, 4, 2	4,5182	5,2682	7,1182
5, 4, 3	4,5231	5,6308	7,3949

- **Ukuran Sampel** menunjukkan jumlah pengamatan pada setiap kelompok.
- $W_{0,9}$, $W_{0,95}$, dan $W_{0,99}$ masing-masing menunjukkan nilai kuantil untuk taraf signifikansi 10%, 5%, dan 1%.

BIOGRAFI PENULIS



Sofia Debi Puspa, S.Pd, M.Si

dosen tetap di Fakultas Teknologi Industri, Universitas Trisakti. Ia menyelesaikan pendidikan sarjana pada bidang Matematika pada tahun 2016 dan melanjutkan studi Magister pada Program Studi Matematika, FMIPA, Universitas Indonesia, yang diselesaikannya pada tahun 2018 dengan fokus pada analisis data, statistika, dan pemodelan matematika. Dalam karir akademiknya, ia mengampu berbagai mata kuliah terkait matematika, statistika, dan *data mining*. Saat ini ia bertanggung jawab sebagai Koordinator Mata Kuliah Statistik dan Kontrol Kualitas, Kalkulus 1, dan Analisa Numerik. Selain mengajar, penulis aktif melakukan penelitian di bidang *data science* dan *machine learning*. Dalam bidang struktural, ia pernah bergabung dalam Tim Koordinasi Renbang FTI USAKTI dan saat ini berperan sebagai Sekretaris Pusat Studi Industri Berkelanjutan FTI USAKTI.



Drs. Joko Riyono, M.Si

meraih gelar M.Si. di bidang matematika dari Universitas Gajah Mada, Yogyakarta, Indonesia, pada tahun 1998. Ia juga meraih gelar Drs. (Matematika) dari Universitas Diponegoro, Semarang, Indonesia, pada tahun 1991. Saat ini, menjadi dosen matematika dan statistika di Departemen Teknik Mesin Universitas Trisakti sejak tahun 1992. Minat penelitiannya meliputi meta-heuristik, global optimizer, pembelajaran mesin, dan penambangan data, dan Aplikasi statistik. Untuk korespondensi dapat dihubungi melalui email: jokoriyono@trisakti.ac.id



Dra. Christina Eni Pujiastuti, M.Si

lahir di Lampung Selatan pada 22 Juli 1963. Ia menempuh pendidikan dasar hingga menengah di Lampung, kemudian melanjutkan pendidikan tinggi di Universitas Gadjah Mada, Yogyakarta. Pendidikan S1 diselesaikannya pada tahun 1988, dan pendidikan S2 diselesaikan pada tahun 1996 di universitas yang sama. Ia mulai berkarir sebagai pengajar di Universitas Trisakti pada tahun 1989. Selain itu, ia juga pernah menjadi dosen tidak tetap di Universitas Pelita Harapan pada periode 2000-2010. Selama menjalani profesinya sebagai pendidik, ia telah mengampu berbagai mata kuliah di bidang matematika, antara lain Kalkulus I, Kalkulus II, Kalkulus III, Matematika Teknik I-IV, Logika Matematika serta Probabilitas dan Statistik.



Aina Latifa Riyana Putri, S.Si, M.Mat

Meraih gelar M.Mat. di bidang matematika dari Universitas Diponegoro, Semarang, Indonesia pada tahun 2023. Ia juga meraih gelar S.Si (Matematika) dari Universitas Negeri Semarang, Indonesia, pada tahun 2021. Saat ini, menjadi dosen pada Prodi Sains Data di Universitas Telkom sejak tahun 2024. Penelitiannya meliputi meta-heuristik, optimasi global, pembelajaran mesin, dan penambangan data. Korespondensi dapat hubungi melalui email: ainaqp@telkomuniversity.ac.id