DASAR DASAR STATISTIKA MATEMATIKA KONSEP PENDUGAAN PARAMETER DAN PENGUJIAN HIPOTESIS

- Samsul Bahri Loklomin
- Muh. Khaedir Lutfi
- Hanna Hilyati Aulia
- Yusup Junaedi
- Agus Winarji
- Ridha Yuniara
- Mia Audina Musyadad
- Fitri Anisa Kusumastuti
- Novela Wulandari
- Tedi Hartoyo
- Rektor Sianturi

ISBN 978-623-125-420-1

DASAR DASAR STATISTIKA MATEMATIKA KONSEP PENDUGAAN PARAMETER DAN PENGUJIAN HIPOTESIS

Samsul Bahri Loklomin
Muh. Khaedir Lutfi
Hanna Hilyati Aulia
Yusup Junaedi
Agus Winarji
Ridha Yuniara
Mia Audina Musyadad
Fitri Anisa Kusumastuti
Novela Wulandari
Tedi Hartoyo
Rektor Sianturi

GET PRESS INDONESIA

DASAR DASAR STATISTIKA MATEMATIKA KONSEP PENDUGAAN PARAMETER DAN PENGUJIAN HIPOTESIS

Penulis :
Samsul Bahri Loklomin
Muh. Khaedir Lutfi
Hanna Hilyati Aulia
Yusup Junaedi
Agus Winarji
Ridha Yuniara
Mia Audina Musyadad
Fitri Anisa Kusumastuti
Novela Wulandari
Tedi Hartoyo
Rektor Sianturi

ISBN : 978-623-125-620-1

Editor : Ari Yanto M.Pd
Penyunting : Tri Putri Wahyuni, S.Pd
Desain Sampul dan Tata Letak : Atyka Trianisa, S.Pd

Penerbit : GET PRESS INDONESIA
Anggota IKAPI No. 033/SBA/2022

Redaksi :
Jln. Palarik Air Pacah No 26 Kel. Air Pacah
Kec. Koto Tangah Kota Padang Sumatera Barat
Website : www. getpress.co.id
Email : adm.getpress@gmail.com

Cetakan pertama, Februari 2025

KATA PENGANTAR

Segala Puji dan syukur atas kehadirat Allah SWT dalam segala kesempatan. Sholawat beriring salam dan doa kita sampaikan kepada Nabi Muhammad SAW. Alhamdulillah atas Rahmat dan Karunia-Nya penulis telah menyelesaikan Buku Dasar Dasar Statistika Matematika Konsep Pendugaan Parameter Dan Pengujian Hipotesis ini.

Buku Ini Membahas Pengantar Statistika Matematika: Konsep Dasar Dan Ruang Sampel, Variabel Acak: Definisi, Sifat, Dan Jenis, Distribusi Peluang: Contoh Dan Aplikasi Dalam Berbagai Bidang, Metode Pemusatan Dan Penyebaran Data: Nilai Harapan Dan Varians, Transformasi Variabel Acak: Menyederhanakan Dan Memahami Distribusi, Metode Pendugaan Parameter: Estimasi Karakteristik Populasi, Uji Hipotesis: Mengambil Keputusan Berdasarkan Data, Analisis Regresi: Hubungan Antara Variabel Dalam Model Matematis, Analisis Variansi: Memahami Perbedaan Antara Kelompok, Model Linier: Prediksi Dan Interpretasi Dalam Konteks Data, Statistika Non-parametrik: Alternatif Untuk Data Yang Tidak Terdistribusi Normal.

Proses penulisan buku ini berhasil diselesaikan atas kerjasama tim penulis. Demi kualitas yang lebih baik dan kepuasan para pembaca, saran dan masukan yang membangun dari pembaca sangat kami harapkan.

Penulis ucapkan terima kasih kepada semua pihak yang telah mendukung dalam penyelesaian buku ini. Terutama pihak yang telah membantu terbitnya buku ini dan telah mempercayakan mendorong, dan menginisiasi terbitnya buku ini. Semoga buku ini dapat bermanfaat bagi masyarakat Indonesia.

Padang, Januari 2025

Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

BAB 1
PENGANTAR STATISTIKA MATEMATIKA: KONSEP DASAR DAN RUANG SAMPEL

Oleh Samsul Bahri Loklomin

Statistika matematika adalah cabang ilmu yang menggabungkan prinsip matematika dan teori statistik untuk menganalisis dan menafsirkan data. Statistika matematika memegang peranan penting dalam berbagai bidang seperti ekonomi, ilmu alam, teknologi, dan ilmu sosial sebagai dasar berbagai metode statistik. Bab ini menjelaskan konsep dasar statistika matematika dan ruang sampel yang menjadi dasar analisis statistik.

1.1 Konsep Dasar Statistika Matematika

Statistika matematika berfokus pada pengembangan dan penerapan teori probabilitas, estimasi parameter, pengujian hipotesis, dan inferensi statistik lainnya. Tujuan utamanya adalah menggunakan informasi yang diperoleh dari sampel untuk memahami karakteristik populasi. Statistik matematika juga memberikan landasan teori bagi berbagai metode analisis data yang digunakan dalam praktik statistik. Statistika matematika berbeda dengan statistika terapan, yang berfokus pada penerapan metode analisis data. Statistika matematika memberikan kerangka teoritis untuk pengembangan metode tersebut.

1.2 Komponen Utama dan Manfaat Statistika Matematika

Statistika Matematika mencakup beberapa komponen besar, yaitu :

1. Probabilitas: Probabilitas adalah dasar statistika matematika yang digunakan untuk memodelkan ketidakpastian. Teori

probabilitas menyediakan alat untuk menganalisis fenomena acak dan menentukan probabilitas suatu kejadian akan terjadi.
2. Teori estimasi: memperkirakan parameter populasi melalui estimasi titik atau interval.
3. Teori Pengujian Hipotesis: Dasar untuk menentukan apakah klaim statistik atau hipotesis diterima atau ditolak berdasarkan data sampel.
4. Inferensi statistik: Proses penarikan kesimpulan tentang suatu populasi berdasarkan data sampel. Inferensi meliputi estimasi parameter, pengujian hipotesis, dan analisis prediktif.
5. Distribusi Statistik: Distribusi probabilitas digunakan untuk menggambarkan data atau fenomena tertentu. Contohnya adalah distribusi normal, distribusi binomial, dan distribusi Poisson.

Statistika matematika membantu dalam berbagai aspek, antara lain:
1. **Desain Eksperimen:** Mengoptimalkan cara pengumpulan data untuk meminimalkan bias dan meningkatkan efisiensi.
2. **Model Probabilistik:** Membantu memahami fenomena acak dalam berbagai konteks seperti risiko keuangan, epidemiologi, dan fisika kuantum.
3. **Validasi Hipotesis:** Membantu mengevaluasi klaim ilmiah atau bisnis berdasarkan data yang tersedia.
4. **Pengembangan Algoritma:** Digunakan dalam pembelajaran mesin, pengolahan sinyal, dan analisis data besar.

1.3 Ruang Sampel dan Titik Sampel

1.3.1 Pengertian Ruang Sampel dan Titik Sampel

Ruang sampel adalah himpunan semua kemungkinan hasil suatu percobaan atau kejadian acak yang diamati. Ruang sampel suatu percobaan dapat direpresentasikan dalam berbagai cara tergantung pada kompleksitas percobaan. Pada umumnya ruang sampel direpresentasikan dalam bentuk himpunan, diagram, atau tabel. Ruang sampel ditandai dengan simbol S.

Contoh 1.

1. Jika kita melempar sebuah dadu enam sisi, ruang sampelnya adalah himpunan angka yang dapat muncul, yaitu:
   $S = \{1, 2, 3, 4, 5, 6\}$

   Di sini, ruang sampelnya terdiri dari enam elemen yang masing-masing menunjukkan sisi yang mungkin muncul saat dadu dilempar.
2. Dalam percobaan melempar sebuah koin, hasil yang mungkin adalah "Muncul angka" atau "Muncul gambar". Ruang sampelnya adalah:
   $S = \{Angka, Gambar\}$

   Ini adalah contoh ruang sampel diskret yang hanya memiliki dua elemen.
3. Dalam percobaan melempar 2 koin, ruang sampelnya adalah $S = \{AA, AG, GA, GG\}$, di mana A menunjukkan angka dan G menunjukkan gambar.

Selanjutnya titik sampel adalah anggota ruang sampel atau kemungkinan terjadinya kejadian. Banyaknya anggota dalam ruang sampel dilambangkan dengan n(*S*). Titik sampel adalah elemen individu dari ruang sampel. Titik sampel mewakili hasil spesifik yang mungkin terjadi dalam suatu percobaan.

Contoh 2.

1. Dalam pelemparan sebuah dadu, salah satu titik sampelnya adalah angka 5.
2. Dalam pelemparan dua koin, salah satu titik sampelnya adalah pasangan (Angka, Gambar).
3. Dalam pengundian kartu, salah satu titik sampelnya adalah kartu “As Sekop”.

Titik sampel sangat penting dalam menentukan kejadian (*event*), yaitu kumpulan satu atau lebih titik sampel dari ruang sampel. Ruang sampel dan titik sampel sangat penting dalam analisis probabilitas karena menjadi dasar untuk menghitung probabilitas suatu kejadian.

1.3.2 Jenis-Jenis Ruang Sampel

Untuk menyatakan hasil percobaan sebagai bilangan sederhana, kita menggunakan apa yang disebut variabel acak. Variabel acak dapat didefinisikan sebagai deskripsi numerik dari suatu hasil eksperimen. Angka-angka tersebut dapat bersifat diskrit (hasil perhitungan) atau kontinu (hasil yang diukur), sehingga variabel acak dapat diklasifikasikan menjadi variabel acak diskrit dan variabel acak kontinu. Oleh karena itu, ruang sampel dapat dibagi menjadi:

1. **Ruang Sampel Diskrit:** Ruang sampel dengan jumlah elemen yang dapat dihitung. Ciri-ciri ruang sampel ini hasilnya terbatas atau berhingga (*finite*) atau tak hingga tetapi dapat dihitung (*countable infinity*) dan sering digunakan dalam percobaan dengan hasil tertentu.
   Contohnya :
   a. Hasil lemparan koin atau dadu
   b. Jumlah pelanggan yang datang ke toko dalam sehari
   c. Jumlah anak dalam sebuah keluarga
   d. Memilih huruf tanpa pengulangan huruf dari kata "statistik"
2. **Ruang Sampel Kontinu:** Ruang sampel dengan elemen yang tak terhingga banyaknya dalam interval tertentu. Nilainya tidak dapat dihitung satu per satu karena melibatkan bilangan real. Contohnya :
   a. Tinggi badan mahasiswa
   b. Suhu pada suatu wilayah pada kurun waktu tertentu
   c. Waktu antara dua kejadian dalam suatu proses
   d. Waktu yang diperlukan untuk menyelesaikan sebuah pekerjaan.

Dalam ruang sampel diskrit, setiap elemen memiliki probabilitas tertentu sedangkan dalam ruang kontinu, probabilitas ditentukan melalui fungsi densitas probabilitas (PDF) atau biasa disebut fungsi kepadatan peluang.

1.4 Kejadian Dalam Statistika

Kejadian adalah sebuah himpunan bagian dari ruang sampel yang berisi satu atau lebih hasil tertentu ari suatu percobaan. Jika A suatu kejadian, maka A adalah suatu kejadian dalam ruang sampel S, maka $A \cup B, A \cap B, A^c$ ketiganya merupakan anggota S. Operasi pada kejadian meliputi gabungan, irisan dan komplemen. Secara umum, kejadian ini sering diklasifikasikan ke dalam berbagai jenis berdasarkan karakteristiknya. Berikut adalah jenis-jenis kejadian dalam statistika :

1. **Kejadian Pasti** *(Certain Event)***, Kejadian pasti adalah kejadian yang pasti terjadi dalam suatu percobaan. Probabilitasnya adalah 1, contohnya:**
   a. **Dalam pelemparan sebuah koin, kejadian "mendapatkan sisi koin (gambar atau angka)" adalah kejadian pasti.**
   b. Dalam pelemparan sebuah dadu, kejadian "mendapatkan angka antara 1 dan 6" adalah kejadian pasti.
2. **Kejadian Mustahil** *(Impossible Event),* Kejadian mustahil adalah kejadian yang tidak mungkin terjadi. Probabilitasnya adalah 0, **contohnya:**
   a. Dalam pelemparan sebuah dadu, kejadian"mendapatkan angka 7" adalah kejadian mustahil.
   b. Dalam pelemparan sebuah koin, kejadian "mendapatkan angka 3" adalah kejadian mustahil.
3. **Kejadian Sederhana** *(Simple Event)***,** Kejadian sederhana adalah kejadian yang hanya memiliki satu kemungkinan hasil dari suatu percobaan, contohnya:
   a. Mendapatkan angka 2 pada pelemparan dadu.
   b. Dalam pelemparan sebuah koin, kejadian "mendapatkan gambar" adalah kejadian sederhana.
4. **Kejadian Majemuk** *(Compound Event)***,** Kejadian majemuk adalah kombinasi dari dua atau lebih kejadian sederhana, contohnya:
   a. Dalam pelemparan dua dadu, kejadian "jumlah mata dadu adalah 7" adalah kejadian majemuk.
   b. **Dalam pengundian kartu, kejadian "mengambil kartu As atau kartu King" adalah kejadian majemuk.**

5. **Kejadian Saling Lepas (*Mutually Exclusive Event*),** kejadian dikatakan saling lepas (*disjoint*) jika tidak memiliki elemen yang sama atau kejadian yang tidak mungkin terjadi secara bersamaan, contohnya:
   a. Dalam pelemparan sebuah dadu, kejadian "mendapatkan angka genap" dan "mendapatkan angka ganjil" adalah saling lepas.
   b. Dalam pengundian kartu, kejadian "mengambil kartu merah" dan "mengambil kartu hitam" adalah saling lepas.
6. **Kejadian Tidak Saling Lepas *(Non-Mutually Exclusive Event)***, kejadian tidak saling lepas jika dua atau lebih kejadian yang mungkin terjadi secara bersamaan, contohnya:
   a. Dalam pelemparan dua dadu, kejadian "jumlah mata dadu genap" dan "jumlah mata dadu lebih dari 7" dapat terjadi bersamaan.
   b. Dalam pelemparan dua dadu, kejadian "jumlah mata dadu genap" dan "jumlah mata dadu lebih dari 7" dapat terjadi bersamaan.
7. Kejadian Bersyarat (*Conditional Event*), **Kejadian bersyarat adalah kejadian yang terjadi dengan syarat bahwa kejadian lain telah terjadi sebelumnya, contohnya :**
   a. **Dalam pengundian kartu, kejadian "mengambil kartu Queen" dengan syarat kartu yang diambil pertama adalah As.**
   b. Dalam populasi siswa, kejadian "memilih siswa perempuan" dengan syarat siswa tersebut berasal dari kelas tertentu.
8. **Kejadian Independen (*Independent Event*),** Kejadian independen adalah dua kejadian yang tidak saling memengaruhi satu sama lain.
   Contoh
   a. Dalam pelemparan dua koin, hasil pelemparan koin pertama tidak memengaruhi hasil pelemparan koin kedua.
   b. Dalam pelemparan sebuah dadu dan pengundian sebuah kartu, hasil dadu tidak memengaruhi kartu yang diambil.

9. **Kejadian Tidak Independen *(Dependent Event)*,** Kejadian tidak independen adalah dua kejadian yang saling memengaruhi satu sama lain, contohnya:
   a. Dalam pengundian kartu tanpa pengembalian, kejadian "mengambil kartu As" dan "mengambil kartu King" saling mempengaruhi.
   b. Dalam memilih bola dari sebuah kantong tanpa pengembalian, warna bola pertama mempengaruhi peluang bola kedua.

Definisi 1.

Subset A dari ruang sampel S disebut kejadian jika termasuk dalam himpunan F sebagai himpunan S yang memenuhi tiga aturan berikut:

a. $S \in F$

b. jika $A \in F$ maka $A^c \in F$; dan

c. jika $A_j \in F$ untuk $j \geq 1$, maka $\bigcup_{j=1}^{\infty} \in F$

Contoh 3.

Misalnya, jika sebuah dadu dilempar, tentukan ruang sampelnya dan nyatakan kejadian A bahwa jumlah dadu yang dilempar adalah 5.

Penyelesaian:

Ruang sampel sepasang dadu ini adalah

$$S = \{(x,y)|x,y = 1,2,3,4,5,6\}$$

dan

$$A = \{(1,4),(4,1),(2,3),(3,2)\}$$

Contoh 4.

Jika $S = \{1,2,3,\dots,9\}$

$A = \{1,3,5,7\}$

$B = \{6,7,8,9\}$

$C = \{2,4,8\}$

Tentukan ruang kejadian dari :

a. $(A^C \cap B)$

b. $(A^C \cap B) \cap C$

c. $B^C \cup C$

Penyelesaian;

a. Jika $S = \{1,2,3, \dots ,9\}$, $B = \{6,7,8,9\}$ dan $A = \{1,3,5,7\}$ jadi

$A^C = \{2,4,6,8,9\}$ jadi :

$(A^C \cap B) = \{2,4,6,8,9\} \cap \{6,7,8,9\}$

$= \{6,8,9\}$

b. Jika $S = \{1,2,3, \dots ,9\}$ dan $(A^C \cap B) = \{6,8,9\}$ maka

$(A^C \cap B) \cap C = \{6,8,9\} \cap \{2,4,8\}$

$= \{8\}$

c. Jika $S = \{1,2,3, \dots ,9\}$, $B = \{6,7,8,9\}$ maka $B^C = \{1,2,3,4,5\}$

jadi :

$B^C \cup C = \{1,2,3,4,5\} \cup \{2,4,8\}$

$= \{1,2,3,4,5,8\}$

1.5 Menghitung Probabilitas dengan Menggunakan Ruang Sampel

Setelah kita mengetahui ruang sampel, kita bisa mulai menghitung probabilitas suatu kejadian. Probabilitas suatu kejadian didefinisikan sebagai perbandingan antara jumlah elemen dalam ruang sampel yang menguntungkan dengan jumlah total elemen dalam ruang sampel. Menghitung probabilitas dengan menggunakan ruang sampel adalah hal yang sangat penting dalam statistika karena memungkinkan kita untuk memahami seberapa besar kemungkinan terjadinya suatu kejadian dalam suatu percobaan acak.

Berikut diberikan beberapa contoh sederhana dalam menghitung Probabilitas dengan menggunakan ruang sampel.

1. Percobaan melempar sebuah uang logam, jika kita ingin menghitung probabilitas munculnya sisi gambar.
   Penyelesaian :
   kita lihat bahwa dalam ruang sampel $S = \{A, G\}$ hanya ada satu sisi gambar, sehingga probabilitasnya dapat dihitung :
   $$P(muncul\ G)) = \frac{1}{2}$$
2. Percobaan melempar dadu, jika kita ingin menghitung probabilitas munculnya angka 5,
   Penyelesaian :
   kita lihat bahwa dalam ruang sampel $S = \{1,2,3,4,5,6\}$ hanya ada satu angka 5, sehingga probabilitasnya dapat dihitung :
   $$\boldsymbol{P(muncul\ angka\ 5) = \frac{1}{6} = 0,167}$$
   Artinya, probabilitas munculnya angka 5 secara acak pada pelemparan sebuah dadu adalah 0,167.
3. Misalkan di sebuah sekolah, terdapat 200 siswa yang mengikuti ujian matematika, dan 150 siswa di antaranya berhasil lulus. Kita ingin mengetahui probabilitas bahwa seorang siswa yang dipilih secara acak dari sekolah tersebut lulus ujian.
   Penyelesaian :

Ruang sampel dalam hal ini adalah seluruh jumlah siswa yang mengikuti ujian, yaitu 200 siswa.

$$\boldsymbol{S} = \{Siswa\ 1, Siswa\ 2, Siswa\ 3, \dots, Siswa\ 200\}$$

Probabilitas seorang siswa lulus ujian adalah perbandingan antara jumlah siswa yang lulus dengan total jumlah siswa yang mengikuti ujian. Berdasarkan data, 150 siswa lulus dan 200 siswa mengikuti ujian. Maka, probabilitasnya adalah:

$$P(Lulus) = \frac{150}{200} = 0{,}75$$

Artinya, probabilitas bahwa seorang siswa yang dipilih secara acak lulus ujian adalah 75%.

1.6 Kesimpulan

Ruang sampel adalah konsep fundamental dalam statistika matematika yang menjadi dasar dari probabilitas dan analisis statistik. Pemahaman tentang ruang sampel dan hubungan antar kejadian sangat penting untuk mengembangkan teori dan metode statistik yang lebih kompleks.

DAFTAR PUSTAKA

Casella, G., & Berger, R. L. (2002). *Statistical Inference*. Cengage Learning.

Herrhyanto, N., & Gantini, T. (2011). Pengantar Statistika Matematis. Yrama Widia.

Ross, S. M. (2014). *Introduction to Probability and Statistics for Engineers and Scientists*. Academic Press.

Wackerly, D., Mendenhall, W., & Scheaffer, R. (2008). *Mathematical Statistics with Applications*. Cengage Learning.

Wattimanela, H. J. (2024). Pengantar Statistika Matematika. Penerbit Buku Pendidikan Deepublish.

Widodo, A., & Andawaningtyas, K. (2017). Pengantar Statistika. Universitas Brawijaya Press.

BAB 2
VARIABEL ACAK: DEFINISI, SIFAT, DAN JENIS

Oleh Muh. Khaedir Lutfi

2.1 Definisi Variabel Acak

Variabel acak adalah salah satu konsep fundamental dalam statistik dan probabilitas yang memungkinkan kita merepresentasikan fenomena acak dalam bentuk kuantitatif. Dalam konteks ini, variabel acak memetakan setiap elemen dalam ruang sampel ke bilangan real. Sebagai contoh, dalam pelemparan dadu, hasil setiap pelemparan (angka 1 hingga 6) dapat direpresentasikan sebagai variabel acak (Ross, 2007).

Definisi formal variabel acak melibatkan ruang sampel , yang mencakup semua kemungkinan hasil dari eksperimen acak. Variabel acak didefinisikan sebagai fungsi yang memetakan setiap elemen ke bilangan real . Pendekatan ini memungkinkan penghitungan probabilitas hasil tertentu dengan alat matematis yang tepat (Walpole, 2007).

Dalam praktik, variabel acak dapat berupa diskrit atau kontinu, bergantung pada jenis data yang dihasilkan dari eksperimen. Sebagai contoh, jumlah kepala dalam tiga kali pelemparan koin adalah variabel acak diskrit, sementara waktu yang diperlukan untuk menyelesaikan suatu tugas dapat dianggap sebagai variabel acak kontinu. Pemahaman mengenai perbedaan ini penting dalam menentukan distribusi probabilitas yang sesuai (Dekking *et al.*, 2006).

2.2 Eksperimen Acak

Eksperimen acak adalah proses atau aktivitas yang hasilnya tidak dapat diprediksi secara pasti sebelumnya, tetapi memiliki kumpulan hasil yang diketahui. Contoh klasik dari eksperimen acak adalah pelemparan koin atau dadu. Dalam pelemparan koin,

terdapat dua kemungkinan hasil, yaitu "kepala" atau "ekor." Demikian pula, dalam pelemparan dadu, hasil yang mungkin adalah angka 1 hingga 6. Kedua eksperimen ini menghasilkan ruang sampel yang terdefinisi dengan baik (Ross, 2007).

Eksperimen acak memiliki ciri utama yaitu ketidakpastian dalam hasil individu, namun dengan probabilitas yang dapat dihitung. Misalnya, dalam pelemparan dadu yang fair (adil), setiap sisi dadu memiliki peluang sama sebesar . Hal ini menjadikan eksperimen acak sebagai dasar dari banyak analisis probabilitas (Walpole, 2007).

Selain itu, eksperimen acak juga menjadi elemen penting dalam pengambilan sampel survei. Menurut Kish (1965), metode pengambilan sampel acak digunakan untuk memastikan bahwa setiap elemen populasi memiliki peluang yang sama untuk dipilih, sehingga hasil yang diperoleh lebih mewakili populasi secara keseluruhan. Dalam bidang penelitian, hal ini menjadi dasar bagi desain eksperimental yang valid.

Dalam implementasi yang lebih kompleks, eksperimen acak juga digunakan dalam simulasi komputer untuk memodelkan sistem yang rumit. Sebagai contoh, Papoulis dan Pillai (2002) menjelaskan bahwa Monte Carlo Simulation adalah metode yang menggunakan eksperimen acak untuk memprediksi hasil dari sistem yang melibatkan ketidakpastian. Dengan cara ini, eksperimen acak menjadi alat yang sangat penting dalam riset dan pengembangan.

2.3 Sifat Utama Variabel Acak

Variabel acak adalah fungsi yang menghubungkan setiap hasil dari suatu percobaan acak ke bilangan real. Konsep ini penting dalam probabilitas karena memungkinkan kita untuk mengukur, memprediksi, dan menganalisis hasil yang tidak pasti. Variabel acak dapat digunakan dalam berbagai bidang, seperti statistik, fisika, biologi, ekonomi, dan teknik.

2.3.1 Domain Variabel Acak

Setiap variabel acak didefinisikan atas ruang sampel (Ω), yaitu himpunan semua kemungkinan hasil dari percobaan acak.

Misalnya, dalam pelemparan koin dua kali, ruang sampelnya adalah $\Omega = \{HH, HT, TH, TT\}$. Variabel acak akan memetakan hasil-hasil ini ke nilai tertentu, seperti jumlah "kepala" yang muncul.

2.3.2 Jenis Variabel Acak

1. **Variabel Acak Diskrit**
   Variabel ini hanya dapat mengambil nilai dari himpunan yang dapat dihitung, seperti bilangan bulat. Contohnya adalah jumlah pelanggan yang datang ke restoran dalam satu jam.
2. **Variabel Acak Kontinu**
   Variabel ini dapat mengambil nilai dari interval pada garis bilangan real (R). Misalnya, tinggi badan seseorang atau waktu yang dibutuhkan untuk menyelesaikan sebuah tugas.

2.3.3 Distribusi Probabilitas

1. **Fungsi Massa Probabilitas (PMF)**
   Untuk variabel acak diskrit, probabilitas setiap nilai x diberikan oleh
   $P(X = x)$
   Sifat penting PMF adalah:
   $0 \leq P(X = x) \leq 1$ dan $\sum_x P(X = x) = 1$
   Contoh: Dalam pelemparan koin,
   $P(X = 0) = 0.25$
   $P(X = 1) = 0.5$
   $P(X = 2) = 0.25$
2. **Fungsi Kepadatan Probabilitas (PDF)**
   Untuk variabel acak kontinu, probabilitas dihitung dengan integrasi dari fungsi PDF $f(x)$ pada interval tertentu. Probabilitas suatu nilai spesifik $P(X = x)$ adalah nol, sehingga probabilitas dihitung untuk interval, misalnya:
   $$P(a \leq X \leq b) = \int_a^b f(x)dx$$
   Sifat penting PDF adalah:
   $f(x) \geq 0$ dan $\int_{-\infty}^{\infty} f(x)dx = 1$

2.3.4 Fungsi Distribusi Kumulatif (CDF)

CDF adalah fungsi yang memberikan probabilitas bahwa variabel acak X kurang dari atau sama dengan nilai tertentu x, yaitu:

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

Sifat-sifat CDF meliputi:

1. Fungsi $F_X(x)$ bersifat non-decreasing.
2. $\lim_{x \to -\infty} F_X(x) = 0$ dan $\lim_{x \to \infty} F_X(x) = 1$

2.3.5 Nilai Ekspektasi (Rata-Rata Teoretis)

Ekspektasi $E[X]$ adalah rata-rata teoretis dari nilai variabel acak, dihitung dengan:

1. Untuk variabel diskrit:

   $$E[X] = \sum_i x_i P(X = x_i)$$

2. Untuk variabel kontinu:

   $$E[X] = \int_{-\infty}^{\infty} x\, f(x) dx$$

   Ekspektasi memberikan gambaran pusat distribusi probabilitas.

2.3.6 Varians (Ukuran Penyebaran)

Varians mengukur sejauh mana nilai variabel acak menyimpang dari rata-ratanya. Varians dihitung sebagai:

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

Varians dapat ditulis ulang sebagai:

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2$$

2.3.7 Linearitas Ekspektasi

Ekspektasi bersifat linear, yaitu:

$$E[aX + b] = aE[X] + b$$

dengan a dan b sebagai konstanta.

2.3.8 Sifat Additif Variabel Acak

Jika $Z = X + Y$, maka ekspektasi bersifat aditif:

$$E[Z] = E[X] + E[Y]$$

Jika X dan Y independen, varians juga bersifat aditif:

$$\text{Var}(Z) = \text{Var}(X) + \text{Var}(Y)$$

2.3.9 Independensi

Dua variabel acak X dan Y dikatakan independen jika:

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$$

Sifat ini menyederhanakan perhitungan probabilitas gabungan.

2.3.10 Aplikasi Variabel Acak

Sifat variabel acak digunakan dalam berbagai aplikasi, seperti:

1. Menentukan hasil eksperimen ilmiah, misalnya probabilitas keberhasilan suatu metode.
2. Analisis risiko di bidang keuangan, seperti memprediksi keuntungan atau kerugian.
3. Model prediktif dalam data sains, termasuk analisis regresi dan machine learning.

2.4 Jenis Variabel Acak

2.4.1 Variabel Acak Diskrit

Variabel acak diskrit hanya dapat mengambil nilai dari himpunan yang terhingga atau dapat dihitung. Contoh klasik adalah jumlah kepala dalam 10 pelemparan koin, di mana hasilnya adalah nilai diskrit seperti 0, 1, hingga 10 (Ross, 2007).

Variabel diskrit sering dimodelkan menggunakan distribusi probabilitas seperti binomial atau Poisson. Distribusi binomial digunakan untuk eksperimen dengan dua kemungkinan hasil (sukses atau gagal), sedangkan distribusi Poisson memodelkan kejadian-kejadian dalam interval tertentu, seperti jumlah pelanggan yang datang ke toko dalam satu jam (Walpole, 2007).

2.4.2 Variabel Acak Kontinu

Variabel acak kontinu mengambil nilai dari interval bilangan real. Contohnya adalah tinggi badan atau waktu yang diukur dengan presisi tinggi. Variabel ini memiliki fungsi densitas probabilitas (PDF) yang mencerminkan peluang relatif dari setiap nilai dalam interval (Papoulis and Pillai, 2002).

Distribusi normal adalah salah satu contoh distribusi kontinu yang paling umum. Sebagai contoh, tinggi badan manusia sering dimodelkan dengan distribusi normal karena banyak data mengikuti pola berbentuk lonceng (Dekking *et al.*, 2005).

2.4.3 Variabel Acak Campuran

Variabel acak campuran menggabungkan elemen diskrit dan kontinu. Misalnya, dalam model stokastik untuk waktu tunggu bus, terdapat kemungkinan diskrit bahwa bus telah tiba (probabilitas diskrit), tetapi jika belum, waktu tunggu selanjutnya dideskripsikan oleh distribusi kontinu seperti eksponensial (Ross, 2007).

Variabel acak campuran banyak digunakan dalam simulasi antrian atau perencanaan logistik di mana kejadian tertentu bergantung pada kombinasi probabilitas diskrit dan kontinu (Walpole, 2007).

2.5 Pendalaman: Contoh Penerapan Sifat Variabel Acak

Untuk mendalami, mari kita fokus pada contoh penerapan variabel acak diskrit dan kontinu dalam skenario nyata. Bagian ini juga akan menjelaskan bagaimana sifat seperti distribusi probabilitas, ekspektasi, dan varians diterapkan.

2.5.1 Contoh Variabel Acak Diskrit

Skenario: Suatu toko memiliki peluang sukses dalam menjual produk kepada pelanggan sebesar 40%. Dalam sehari, terdapat 5 pelanggan yang datang ke toko tersebut. Berapa probabilitas bahwa toko berhasil menjual produk ke 3 pelanggan?

1. **Identifikasi Variabel Acak:**

   Definisikan variabel acak X sebagai jumlah pelanggan yang membeli produk. X adalah variabel acak diskrit dengan kemungkinan nilai $X = 0, 1, 2, 3, 4, 5$.
2. **Distribusi Probabilitas:**

   Distribusi yang relevan adalah **Binomial**, karena setiap pelanggan membeli (sukses) atau tidak membeli (gagal), dan hasil setiap pelanggan independen. Distribusi Binomial memiliki rumus:

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

   di mana:

   a. $n = 5$ (jumlah pelanggan)

   b. $k = 3$ (jumlah sukses yang diharapkan)

   c. $p = 0.4$ (peluang sukses)

3. **Perhitungan:**

   Substitusikan nilai:

   $$P(X = 3) = \binom{5}{3} (0.4)^3 (1-0.4)^2$$

   $$P(X = 3) = 10 \cdot (0.4)^3 \cdot (0.6)^2 = 10 \cdot 0.064 \cdot 0.36 = 0.2304$$

   Jadi, probabilitas bahwa toko berhasil menjual ke 3 pelanggan adalah $\mathbf{23.04\%}$.
4. **Ekspektasi dan Varians:**

   Ekspektasi $(E[X])$ dan varians $(\mathrm{Var}(X))$ dalam distribusi Binomial dihitung dengan:

   $E[X] = n \cdot p = 5 \cdot 0.4 = 2$

   $\mathrm{Var}(X) = n \cdot p \cdot (1-p) = 5 \cdot 0.4 \cdot 0.6 = 1.2$

   Artinya, rata-rata jumlah pelanggan yang membeli produk adalah 2, dengan penyebaran (varians) sebesar 1.2.

2.5.2 Contoh Variabel Acak Kontinu

Skenario: Sebuah perusahaan logistik mencatat waktu pengiriman barang. Waktu pengiriman rata-rata adalah 2 jam, dan distribusi waktu mengikuti distribusi eksponensial. Berapa peluang bahwa pengiriman selesai dalam waktu kurang dari 1 jam?

1. **Identifikasi Variabel Acak:**
   Definisikan variabel acak T sebagai waktu pengiriman (dalam jam). Karena T adalah waktu kontinu, distribusinya menggunakan fungsi kepadatan probabilitas (PDF).
2. **Distribusi Probabilitas:**
   Distribusi Eksponensial memiliki PDF:
   $f(t) = \lambda e^{-\lambda t}, \quad t \geq 0$
   di mana $\lambda = \frac{1}{\mu}$, dan μ adalah rata-rata. Dalam kasus ini, $\mu = 2$, sehingga $\lambda = \frac{1}{2} = 0.5$
3. **CDF untuk Peluang:**
   Peluang dihitung menggunakan fungsi distribusi kumulatif (CDF):
   $F(t) = P(T \leq t) = \int_0^t \lambda e^{-\lambda t} dt$
   Substitusikan nilai $\lambda = 0.5$:
   $F(t) = 1 - e^{-0.5t}$
4. **Perhitungan:**
   Untuk $t = 1$:
   $F(1) = 1 - e^{-0.5 \cdot 1} = 1 - e^{-0.5} \approx 1 - 0.6065 = 0.3935$
   Jadi, peluang pengiriman selesai dalam waktu kurang dari 1 jam adalah **39.35%**.
5. **Ekspektasi dan Varians:**
   Ekspektasi $(E[T])$ dan varians $(\mathrm{Var}(T))$ dalam distribusi Eksponensial dihitung dengan:
   $E[T] = \frac{1}{\lambda} = 2$
   $\mathrm{Var}(T) = \frac{1}{\lambda^2} = 4$

Artinya, rata-rata waktu pengiriman adalah 2 jam, dengan penyebaran waktu sebesar 4 jam^2.

Tabel 2.1. Perbandingan Aplikasi: Variabel Diskrit vs. Kontinu

Aspek	**Diskrit**	**Kontinu**
Contoh	Jumlah pelanggan yang membeli produk.	Waktu pengiriman barang.
Probabilitas	Dihitung langsung untuk nilai tertentu $(P(X = x))$.	Dihitung dengan integrasi PDF $(P(a \leq X \leq b))$.
Distribusi	Binomial, Poisson.	Eksponensial, Normal.
Ekspektasi $(E[X])$	Rata-rata tertimbang nilai X.	Rata-rata nilai kontinu pada interval tertentu.
Varians $(\mathrm{Var}(X))$	Pengukuran penyebaran nilai diskrit.	Pengukuran penyebaran nilai kontinu.

Pemahaman sifat variabel acak membantu kita memodelkan berbagai situasi nyata, baik yang melibatkan jumlah hasil (diskrit) maupun nilai yang bersifat kontinu. Dalam praktik, variabel acak diskrit sering digunakan untuk menghitung jumlah kejadian tertentu, sedangkan variabel acak kontinu lebih cocok untuk mengukur hasil yang tidak terbatas seperti waktu atau jarak. Dengan menggunakan distribusi yang sesuai, kita dapat menghitung probabilitas, ekspektasi, dan varians untuk mendukung pengambilan keputusan berdasarkan data probabilistik.

DAFTAR PUSTAKA

Dekking, F.M. et al. (2005) A Modern Introduction to Probability and Statistics. Springer.

Dekking, F.M. et al. (2006) A Modern Introduction to Probability and Statistics: Understanding Why and How. Springer London (Springer Texts in Statistics). Available at: https://books.google.co.id/books?id=TEcmHJX67coC.

Kish, L. (1965) Survey Sampling. Wiley (A Wiley Interscience Publication). Available at: https://books.google.co.id/books?id=DqsMvwEACAAJ.

Papoulis, A. and Pillai, S.U. (2002) Probability, Random Variables, and Stochastic Processes. McGraw-Hill.

Ross, S.M. (2007) Introduction to Probability Models. Elsevier Science (Introduction to Probability Models). Available at: https://books.google.co.id/books?id=12Pk5zZFirEC.

Walpole, R.E. (2007) Probability & Statistics for Engineers & Scientists. Pearson Prentice Hall (Probability \& Statistics for Engineers \& Scientists). Available at: https://books.google.co.id/books?id=N_weAQAAIAAJ.

BAB 3
DISTRIBUSI PELUANG: CONTOH DAN APLIKASI DALAM BERBAGAI BIDANG

Oleh Hanna Hilyati Aulia

3.1 Pendahuluan

Distribusi Peluang adalah konsep dasar dalam teori probabilitas yang digunakan untuk menggambarkan kemungkinan hasil dari suatu percobaan acak. Distribusi ini menunjukkan bagaimana peluang tersebar di antara berbagai hasil yang mungkin. Ada berbagai jenis distribusi peluang yang digunakan tergantung pada jenis percobaan dan sifat data yang digunakan.

Distribusi peluang dapat dibedakan menjadi dua kategori utama: diskrit dan kontinu. Perbedaan utama antara keduanya terletak pada jenis variabel yang dipertimbangkan. Variabel diskrit adalah variabel yang hanya bisa mengambil nilai tertentu (terbatas atau hitungable), seperti jumlah kejadian atau hasil percobaan yang terhitung. Variabel kontinu adalah variabel yang dapat mengambil nilai dalam suatu rentang kontinu, seperti waktu, panjang, atau suhu. Berikut adalah penjelasan mengenai distribusi peluang dalam kedua kategori tersebut, beserta contoh dan aplikasinya.

3.2 Distribusi Peluang Diskrit

Distribusi peluang diskrit menggambarkan kejadian yang hasilnya berupa nilai-nilai yang dapat dihitung, biasanya berupa bilangan bulat.

Sifat Utama:

1. Probabilitas setiap nilai P(X=x) adalah antara 0 dan 1.
2. Jumlah total probabilitas semua nilai adalah 1 atau $\sum P(X = x) = 1$

Fungsi Massa Peluang atau *probability mass function* (PMF):

Fungsi massa peluang $P(X = x)$ menggambarkan probabilitas variabel acak diskrit *X* dengan mengambil nilai tertentu x.

Nilai rata-rata atau nilai yang diharapkan (expected value) dari x diberikan sebagai

$$E(x) = \mu = \sum xp(x)$$

di mana elemen-elemen dijumlahkan atas semua nilai variabel acak x. Kemudian, variansi dari distribusi peluang diskrit ini adalah

$$\sigma^2 = E(x - \mu) = \sum (x - \mu)^2 p(x)$$

Beberapa jenis distribusi diskrit antara lain sebagai berikut: distribusi binomial, distribusi binomial negatif, distribusi poisson, distribusi geometrik dan distribusi hipergeometrik.

3.2.1 Distribusi Binomial

Distribusi binomial adalah distribusi peluang diskrit dari banyaknya kemunculan suatu peristiwa dalam barisan n percobaan bebas dari suatu percobaan acak dengan peluang terjadinya peristiwa tersebut sebesar p dalam setiap percobaan.

Eksperimen binomial adalah eksperimen yang memiliki lima ciri berikut:

1. Percobaan terdiri dari n percobaan yang identik.
2. Setiap percobaan menghasilkan salah satu dari dua hasil: sukses atau gagal.
3. Peluang keberhasilan dalam satu percobaan sama dengan p dan tetap sama dari percobaan ke percobaan. Peluang kegagalan sama dengan $q = (1 - p)$.
4. Uji cobanya bersifat independen.
5. Untuk suatu variabel x, banyaknya keberhasilan diamati selama n percobaan, untuk $x = 0, 1, 2, \ldots, N$.

Variabel Acak X adalah banyaknya kemunculan X sejumlah n percobaan bebas. Parameter n adalah jumlah percobaan dan p adalah peluang sukses pada setiap percobaan. Fungsi massa probabilitas nya adalah:

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

$x = 0,1,2,\dots,n$

Rata-rata/Mean distribusi binomial

$\mu = np$

Variansi distribusi binomial

$\sigma^2 = np(1-p)$

Beberapa contoh aplikasi distribusi binomial dalam bidang pendidikan yakni menilai kemungkinan jumlah siswa yang lulus ujian dalam suatu kelas. Pada Bidang kesehatan: Menghitung berapa banyak pasien yang sembuh setelah diberi pengobatan tertentu.

Contoh Soal:

Seorang siswa mengerjakan ujian yang terdiri dari 10 soal pilihan ganda, dengan masing-masing soal memiliki 4 pilihan jawaban (A, B, C, D), dan hanya satu yang benar. Jika siswa menebak secara acak semua jawaban, peluang keberhasilan untuk setiap soal adalah p=0.25. Distribusi binomial dapat digunakan untuk menghitung probabilitas siswa menjawab sejumlah soal dengan benar. Berapa probabilitas siswa menjawab tepat 5 soal dengan benar?

Penyelesaian

Diketahui

$n = \ jumlah\ soal = 10$

$p = probabilitas\ menjawab\ benar = 0{,}25$

$(1-p) = probabilitas\ menjawab\ soal\ salah = 1 - 0{,}25$
$= 0{,}75$

$x = \ jumlah\ soal\ yang\ dijawab\ benar = 5$

maka dengan rumus diperoleh

$p_5(5;10;0{,}25) = C_5^{10} 0{,}25^5 (0{,}75)^{10-5} = 0{,}0584.$

Probabilitas bahwa siswa menjawab tepat 5 soal dengan benar adalah $0{,}0584$ atau sekitar 5,84%.

3.2.2 Distribusi Binomial Negatif

Distribusi binomial negatif adalah distribusi probabilitas diskret yang digunakan untuk memodelkan jumlah percobaan yang diperlukan dalam sebuah proses Bernoulli untuk mencapai sejumlah keberhasilan tertentu, yang dilambangkan dengan j. Distribusi ini sangat berguna ketika kita ingin mengetahui berapa banyak percobaan yang dibutuhkan untuk mencapai jumlah keberhasilan yang diinginkan.

Variabel Acak X menyatakan jumlah total percobaan (x) yang diperlukan untuk mencapai j keberhasilan. Parameter n adalah jumlah percobaan dan p adalah peluang sukses pada setiap percobaan. Fungsi massa probabilitas nya adalah :

$$p_x(x-1; j-1, p) = C_{j-1}^{x-1} p^j (1-p)^{x-j}$$

$$x = 0,1,2, \dots, n$$

Rata-rata/Mean distribusi binomial negatif

$$\mu = j/p$$

Variansi distribusi binomial negatif

$$\sigma^2 = j(1-p)/p^2$$

Beberapa contoh aplikasi distribusi binomial negatif untuk kontrol kualitas, yaitu menghitung jumlah barang yang harus diperiksa sebelum menemukan sejumlah barang cacat tertentu. Dalam biologi/genetika untuk memodelkan jumlah percobaan yang dibutuhkan untuk mengamati sejumlah mutasi atau kejadian tertentu pada populasi.dalam keuangan untuk memprediksi jumlah transaksi yang diperlukan untuk mencapai target jumlah transaksi yang menguntungkan.

Contoh Soal:

Dalam sebuah permainan, seorang pemain memiliki peluang 30% untuk menang dalam satu kali percobaan. Berapa rata-rata dan variansi jumlah percobaan yang diperlukan agar pemain menang sebanyak 4 kali?

Penyelesaian
Diketahui
$p = peluang\ menang = 0{,}3$
$(1-p) = peluang\ kalah = 1 - 0{,}3 = 0{,}7$
$j = \ jumlah\ menang\ yang\ diinginkan = 4$
maka dengan rumus diperoleh
Rata-rata $\mu = j/p = 4/0{,}3 = 13{,}3$
dan

Variansi $\sigma^2 = j(1-p)/p^2 = 4.0{,}7/0{,}3^2 = 2{,}8/0{,}09 = 31{,}11.$

3.2.3 Distribusi Multinomial

Distribusi multinomial adalah perluasan dari distribusi binomial. Distribusi ini digunakan untuk memodelkan probabilitas hasil dari serangkaian percobaan di mana setiap percobaan memiliki lebih dari dua kemungkinan hasil (kategori). Dalam setiap percobaan, probabilitas masing-masing kategori tetap konstan, dan jumlah percobaan total adalah tetap.

Variabel Acak X menyatakan jumlah total kejadian (x_i) dalam kategori ke -iyang diperlukan dalam n percobaan. Parameter k adalah jumlah kategori dan p_i adalah peluang sukses pada kategori ke-i. Dengan $i = 1,2,\ldots,k$ dan $\sum_{i=1}^{k} x_i = n$.
Fungsi massa probabilitas nya adalah

$$p(x_1,x_2,\ldots,x_k) = \frac{n!}{x_1!x_2!\ldots,x_k!} {p_1}^{x_1}{p_2}^{x_2}\ldots{p_k}^{x_k}$$

$x_i = 0,1,2,\ldots,n$ dan $\sum_{i=1}^{k} x_i = n$
Rata-rata/Mean distribusi binomial negatif
$\mu = np_i$
Variansi distribusi binomial negatif
$\sigma^2 = np_i(1-p_i)$
Distribusi multinomial sering digunakan dalam situasi di mana terdapat lebih dari dua kemungkinan hasil untuk setiap percobaan. Beberapa aplikasi umumnya antara lain

1. **Survei dan Polling**, Menghitung distribusi tanggapan dari responden terhadap beberapa pilihan dalam survei.
2. **Pemasaran**, Menganalisis pangsa pasar untuk beberapa merek dalam suatu kategori produk.
3. **Genetika**, Memodelkan distribusi frekuensi genotipe pada populasi.
4. **Permainan Dadu atau Kartu**, Menghitung probabilitas hasil dalam permainan dengan banyak hasil.

Contoh Soal:
Sebuah survei dilakukan terhadap 100 orang untuk memilih preferensi mereka terhadap tiga jenis minuman: teh, kopi, dan jus. Probabilitas seseorang memilih teh adalah 0.4, kopi 0.35, dan jus 0.25. Hitung probabilitas bahwa:

- 40 orang memilih teh.
- 35 orang memilih kopi.
- 25 orang memilih jus.

Penyelesaian
Diketahui

$n = total\ responden = 100$

$p_i = \text{probabilitas kategori}$

$x_i = \text{jumlah masing} - \text{masing kategori}$

$p_1 = 0{,}4; p_2 = 0{,}35; p_3 = 0{,}25$

$x_1 = 40; x_2 = 35; x_3 = 25$

maka dengan rumus diperoleh

$$p(40{,}35{,}25) = \frac{100!}{40!\,35!\,25!} 0{,}4.0{,}35.0{,}25 = 0{,}0443.$$

Probabilitas bahwa 40 orang memilih teh, 35 orang memilih kopi, 25 orang memilih jus adalah $0{,}0443$ atau sekitar $4{,}43\%$.

3.2.4 Distribusi Geometrik

Distribusi geometrik adalah distribusi probabilitas yang menggambarkan jumlah percobaan yang diperlukan untuk mencapai keberhasilan pertama dalam sebuah proses Bernoulli (percobaan yang hanya memiliki dua kemungkinan: sukses atau

gagal). Distribusi ini cocok digunakan ketika kita ingin mengetahui berapa banyak percobaan yang diperlukan untuk memperoleh keberhasilan pertama.

Variabel Acak x menyatakan jumlah percobaan diperlukan untuk memeperoleh keberhasilan pertama. Parameter p adalah peluang sukses pada setiap percobaan, dan $1-p$ adalah peluang kegagalan setiap percobaan Fungsi massa probabilitas nya adalah :

$p_x(x;p) = p(1-p)^{x-1}$

$x = 0,1,2,\dots,n$

Rata-rata/Mean distribusi geometrik

$\mu = 1/p$

Variansi distribusi geometrik

$\sigma^2 = (1-p)/p^2$

Distribusi geometrik sering digunakan dalam berbagai bidang untuk memodelkan situasi di mana kita mengulang percobaan sampai keberhasilan pertama tercapai. Beberapa aplikasi umum antara lain: menghitung berapa banyak interaksi dengan pelanggan yang diperlukan untuk memperoleh satu penjualan, menghitung berapa banyak transaksi yang perlu dilakukan untuk mendapatkan keuntungan pertama dan memodelkan berapa banyak pasien yang perlu diuji untuk mendapatkan pasien dengan respons positif terhadap pengobatan.

Contoh Soal:

Seorang pemain game memiliki peluang 0.2 untuk menang dalam satu kali percobaan. Hitung probabilitas bahwa pemain tersebut akan menang untuk pertama kalinya pada percobaan ke-4!

Penyelesaian

Diketahui

$p = peluang\ menang = 0,2$

$(1-p) = peluang\ kalah = 1 - 0,2 = 0,8$

$x = \ percobaan\ pertama\ menang = 4$

maka dengan rumus diperoleh

$p_4(4;0,2) = 0,2(0,8)^{4-1} = 0,1024.$

Probabilitas bahwa pemain akan menang untuk pertama kalinya pada percobaan ke-4 adalah 0,1024 atau sekitar 10,24%.

3.2.5 Distribusi Hipergeometrik

Distribusi hipergeometrik adalah distribusi probabilitas yang digunakan untuk menggambarkan jumlah keberhasilan dalam pengambilan sampel tanpa pengembalian dari suatu populasi terbatas. Berbeda dengan distribusi binomial, pada distribusi hipergeometrik, setiap elemen yang dipilih tidak dikembalikan ke populasi, yang berarti ada perubahan probabilitas pada setiap percobaan.

Variabel Acak X menyatakan jumlah kemunculan suatu peristiwa x dalam sampel berukuran n (diambil tanpa pengembalian), dari populasi berukuran N yang berisi k kemungkinan terjadinya tertentu.
Fungsi massa probabilitas nya adalah:

$$p_x(x;N,n,k) = \frac{C_x^k \times C_{n-x}^{N-k}}{C_n^N}$$

Rata-rata/Mean distribusi binomial

$$\mu = nk/N$$

Variansi distribusi binomial

$$\sigma^2 = nk(N-k)(N-n)/N^2(N-1)$$

Distribusi hipergeometrik ini digunakan dalam situasi di mana pengambilan sampel dilakukan tanpa pengembalian. Beberapa aplikasi umum seperti memeriksa sejumlah produk dari sebuah batch untuk menentukan jumlah barang cacat atau rusak, memilih kandidat secara acak dari kelompok terbatas dengan kriteria tertentu dan menghitung jumlah sistem yang mengalami kegagalan dalam pemilihan dari sistem yang ada.

Contoh Soal:

Di dalam sebuah kotak terdapat 20 bola, 8 bola merah dan 12 bola biru. Jika kita memilih 5 bola secara acak tanpa pengembalian, berapa probabilitas bahwa 3 bola yang terpilih adalah bola merah?

Penyelesaian

Jumlah total bola $N = 20$

Jumlah bola merah $K = 8$:

Jumlah bola yang dipilih $n = 5$

Jumlah bola merah yang dipilih $x = 3$

$$p_3(3;20,5,8) = \frac{C_3^8 \times C_{5-3}^{20-8}}{C_5^{20}} = \frac{56 \times 66}{15504} = 0{,}238$$

Probabilitas memilih 3 bola merah dari 5 bola yang dipilih adalah sekitar 0,238 atau 23,8%.

3.2.6 Distribusi Poisson

Distribusi Poisson adalah distribusi probabilitas yang digunakan untuk menggambarkan jumlah kejadian suatu peristiwa dalam interval waktu atau ruang tertentu, dengan asumsi kejadian tersebut independen dan terjadi dengan rata-rata yang tetap.

Suatu proses didefinisikan sebagai proses Poisson jika peristiwa terjadi dalam waktu/area/ruang memenuhi tiga asumsi:

1. Banyaknya peristiwa yang terjadi dalam selang waktu yang saling lepas adalah bebas.
2. Peluang terjadinya satu kejadian dalam selang waktu yang kecil adalah proporsional dengan panjang intervalnya.
3. Kemungkinan terjadinya lebih dari satu kejadian dalam selang waktu yang kecil dapat diabaikan.

Variabel Acak X menyatakan banyaknya kemunculan suatu peristiwa (hasil Proses Bernoulli) x dalam selang waktu tertentu. Parameter λ, juga dikenal sebagai parameter bentuk, menunjukkan angka rata-rata kejadian per satuan selang waktu atau jumlah kejadian yang diharapkan.

Fungsi massa probabilitas nya adalah:

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

$$x = 0,1,2, \dots ; \ \lambda > 0$$

Rata-rata/Mean distribusi binomial

$E(X) = \lambda$

Variansi distribusi binomial

$Var(X) = \lambda$

Distribusi Poisson sering digunakan untuk memodelkan kejadian yang jarang atau acak dalam interval waktu atau ruang tertentu. Beberapa aplikasi umum:

1. **Telekomunikasi**: Menghitung jumlah panggilan telepon yang diterima oleh pusat panggilan dalam satu jam.
2. **Kesehatan**: Menghitung jumlah pasien yang datang ke rumah sakit dalam satu hari.
3. **Antrian**: Menghitung jumlah kendaraan yang melewati suatu titik pengamatan dalam satu jam.

Contoh Soal:

Rata-rata pelanggan yang datang ke sebuah toko adalah 3 pelanggan per jam. Hitung probabilitas bahwa dalam satu jam, toko tersebut akan kedatangan 5 pelanggan!

Penyelesaian

Diketahui

$\lambda = rata - ratabpelanggan\ per\ jam = 3$

$x = jumlah\ pelanggan\ yang\ datang\ dalam\ satu\ jam = 5$

maka dengan rumus diperoleh

$$p_5(5;3) = 3^5 \frac{e^{-3}}{5!} = 243 \frac{0{,}0498}{120} = 0{,}1008$$

Probabilitas bahwa dalam satu jam toko akan kedatangan 5 pelanggan adalah sekitar 0,1008 atau 10,08%.

3.3 Distribusi Peluang Kontinu

Distribusi peluang kontinu adalah distribusi probabilitas yang digunakan untuk menggambarkan variabel acak kontinu, yaitu variabel yang dapat mengambil nilai dalam rentang atau interval tertentu, termasuk nilai-nilai di antara dua angka. Probabilitas dalam distribusi kontinu tidak dihitung untuk nilai tunggal

melainkan untuk interval nilai, dengan fungsi kepadatan probabilitas (*Probability Density Function*) atau PDF sebagai representasinya. PDF dinotasikan sebagai $f(x)$. Total area di bawah kurva PDF adalah 1, yang mewakili probabilitas total. Probabilitas atau (*Cummulative density function*) yang disingkat CDF dirumuskan sebagai berikut

$$P(a \leq x \leq b) = \int_a^b f(x)\,dx$$

Beberapa jenis distribusi kontinu antara lain distribusi uniform, distribusi eksponensial, distribusi Gamma, Distribusi Normal, Distribusi Lognormal, distribusi Weibull dan distribusi beta. Distribusi lain yang sering digunakan ada distribusi F, t dan Chi-Square. Namun pada buku ini akan dibahas beberapa distribusi yang sering digunakan pada analisis statsitika dasar.

3.3.1 Distribusi Uniform

Distribusi uniform kontinu adalah distribusi probabilitas yang menggambarkan kejadian di mana semua hasil dalam suatu interval memiliki kemungkinan yang sama untuk terjadi. Distribusi ini disebut "kontinu" karena mencakup semua nilai real dalam interval tertentu.

Variabel Acak: X yang mempunyai peluang sama pada setiap subinterval dengan panjang yang sama dalam dukungannya.

Parameter: α dan β, dimana α dan β masing-masing adalah batas minimum dan maksimum dukungan. Secara umum, kita mengatakan bahwa X adalah variabel acak seragam pada interval (α, β) jika fungsi kepadatan probabilitasnya diberikan oleh

$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha}, & \alpha \leq x \leq \beta \\ 0. \quad untuk\ x\ lainnya \end{cases}$$

Rata-rata dan variansinya adalah

$$E(x) = \frac{\beta + \alpha}{2}$$

$$Var(x) = \frac{(\beta - \alpha)^2}{12}$$

Contoh Soal

Suatu waktu tunggu bus di halte mengikuti distribusi uniform kontinu pada interval 0 hingga 15 menit.

1. Tentukan fungsi densitas probabilitasnya.
2. Berapa probabilitas bahwa waktu tunggu bus kurang dari 5 menit?
3. Berapa probabilitas bahwa waktu tunggu bus berada antara 5 hingga 12 menit?

Penyelesaian

Waktu tunggu mengikuti distribusi uniform kontinu pada interval $[0,15]$, fungsi densitas probabilitasnya adalah

$$f(x) = \begin{cases} \frac{1}{15-0} = \frac{1}{15}, & 0 \le x \le 15 \\ 0. \quad untuk\ x\ lainnya \end{cases}$$

Probabilitas untuk interval tunggu kurang dari 5 menit atau $[0,5]$ adalah

$$P(0 \le x \le 5) = \int_0^5 \frac{1}{15} dx = \frac{1}{15} x|_0^5 = \frac{1}{15}(5-0) = \frac{1}{3}$$

Probabilitas untuk interval tunggu antara 5 hingga 10 menit atau $[5,10]$ adalah

$$P(5 \le x \le 12) = \int_5^{12} \frac{1}{15} dx = \frac{1}{15} x|_5^{12} = \frac{1}{15}(12-5) = \frac{7}{15}$$

Jadi, probabilitas waktu tunggu bis kurang dari 5 menit adalah $1/3$ atau 0,3333 atau sekitar $33{,}33\%$ dan probabilitas waktu tunggu bis antara 5 hingga 12 menit adalah $7/15$ atau 0,4667 atau sekitar $46{,}67\%$.

3.3.2 Distribusi Eksponensial

Distribusi eksponensial adalah distribusi probabilitas kontinu yang digunakan untuk memodelkan waktu antar kejadian dalam proses Poisson. Distribusi ini sering digunakan untuk

menggambarkan waktu tunggu atau waktu hingga kejadian tertentu terjadi, seperti waktu kerusakan mesin, waktu tunggu pelanggan, atau waktu layanan.

Variabel Acak: X adalah variabel acak kontinu yang bisa menunjukkan waktu tunggu aau durasi dan parameter λ adalah tingkat kerjadian atau rata-rata waktu antar kejadian, $\lambda = 1/\mu$. Fungsi kepadatan probabilitasnya diberikan oleh

$f(x) = \lambda e^{-\lambda x}, x \geq 0$

Rata-rata dan variansinya adalah

$$E(x) = \frac{1}{\lambda}$$

$$Standar\ deviasi = \frac{1}{\lambda}$$

Contoh Soal

Sebuah mesin ATM melayani pelanggan dengan rata-rata waktu 4 menit per layanan. Berapa probabilitas bahwa layanan pelanggan selesai dalam waktu kurang dari 3 menit?

Penyelesaian

Rata-rata $\mu = 4$ maka $\lambda = 1/\mu = 1/4 = 0{,}25$, sehingga

CDF dsitribusi eksponensialnya untuk waktu kurang dari 3 menit atau $x \leq 3$ adalah

$$P(X \leq 3) = \int_0^3 \lambda e^{-\lambda x}\, dx = \left[-e^{-\lambda x}\right]_0^3 = 1 - e^{-\lambda x} = 1 - e^{-0{,}25 \times 3}$$
$$= 1 - 0.4724 = 0.5276$$

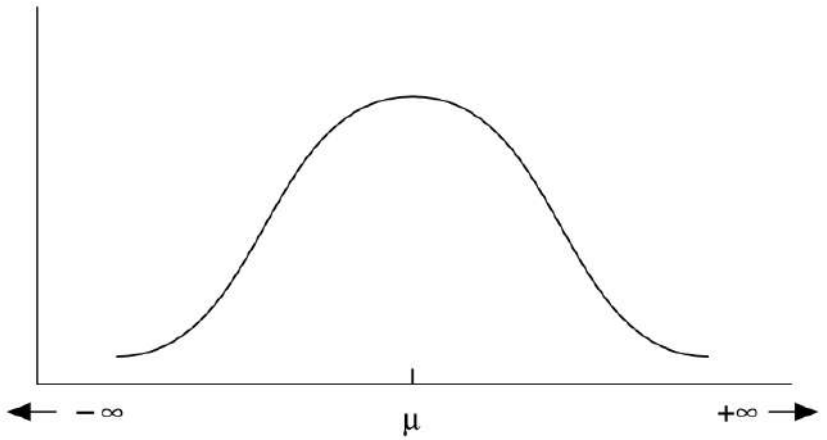
Probabilitas bahwa layanan selesai dalam waktu kurang dari 3 menit adalah 52.76%

3.3.3 Distribusi Normal

Distribusi normal adalah salah satu distribusi probabilitas kontinu yang paling sering digunakan. Distribusi ini memiliki bentuk simetris menyerupai lonceng dan sering disebut sebagai distribusi Gaussian. Distribusi ini ditentukan oleh dua parameter: mean (μ) dan standar deviasi (σ).

Distribusi normal adalah fungsi distribusi probabilitas kontinu yang paling sering digunakan. Jika mean sama dengan nol

dan varians sama dengan 1, maka distribusi tersebut disebut distribusi normal baku. Pdf distribusi normal standar ditunjukkan pada Gambar 3.1. Terlihat simetris terhadap mean dan bentuk khasnya dikenal sebagai kurva berbentuk lonceng. Garis simetri dan bentuknya akan berubah bergantung pada nilai mean dan variansnya masing-masing.



Gambar 3.1. Kurva Distribusi Normal

Fungsi kepadatan peluang distribusi normal adalah

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

Fungsi Distribusi Kumulatif (CDF) dari Distribusi Normal diberikan oleh:

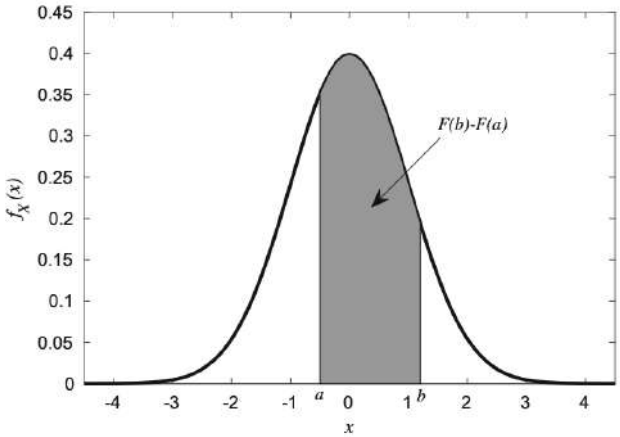
$$F(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} e^{-(x-\mu)^2/2\sigma^2} dx\,, -\infty < x < \infty$$

Beriku ini adalah fungsi kepadatan peluang untuk distribusi normal standar mean $\mu = 0$ dan varians $\sigma^2 = 1$

$$f(z) = \frac{1}{\sqrt{2\pi\sigma}} e^{-z^2/2}, -\infty < z < \infty$$

Untuk mencari probabilitas suatu variabel acak yang mempunyai distribusi normal standar seperti berikut akan mempunyai nilai antara a dan b, kita menggunakan persamaan

$P\,(a < Z \leq b) = F\,(b) - F\,(a)$ seperti ditunjukkan pada Gambar 3.2.



Gambar 3.2. Kurva $P\,(a < Z \leq b) = F\,(b) - F\,(a)$

Jika variabel acak X mengikuti distribusi normal dengan mean μ dan varians σ^2, maka variabel acak Z yang diberikan di bawah ini mengikuti distribusi normal dengan mean 0 dan varians 1, yaitu distribusi normal standar.

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

Jadi, ketika X mengikuti distribusi normal dengan mean μ dan varians σ^2, probabilitas X berada di antara p dan q diberikan oleh,

$$P(p < X \leq q) = P\left(\frac{p-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{q-\mu}{\sigma}\right)$$
$$= P\left(\frac{p-\mu}{\sigma} < Z < \frac{q-\mu}{\sigma}\right) = F\left(\frac{q-\mu}{\sigma}\right) - F\left(\frac{p-\mu}{\sigma}\right)$$

Probabilitas ini ditunjukkan oleh area yang diarsir pada Gambar 3.2 jika $b = \frac{q-\mu}{\sigma}$ dan $a = \frac{p-\mu}{\sigma}$. Sebuah contoh akan ditampilkan nanti tentang cara menggunakan tabel normal standar untuk menghitung nilainya.

Distribusi normal digunakan secara luas di berbagai bidang, untuk mengukur skor IQ atau tinggi badan dan lebih banyak digunakan dalam analsis uji hipotesis dan nilainya telah dibuat dalam suatu tabel distribusi normal standar.

Contoh Soal

Berat badan manusia dewasa berdistribusi normal dengan rata-rata $\mu = 70$kg dan standar deviasi $\sigma = 5$kg. Hitung probabilitas bahwa seseorang memiliki berat badan antara 65 kg dan 75 kg.

Penyelesaian

Dalam mengerjakan soal ini, kita akan menstandarisasi perhitungan dengan

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

untuk $x_1 = 65$ dan $x_2 = 75$

$$Z_1 = \frac{65 - 70}{5} = -1$$

$$Z_2 = \frac{75 - 70}{5} = 1$$

Selanjutnya kita hitung CDF untuk distribusi normal standar, bisa dengan mengintegralkan fungsi pdf saat $Z_1 = -1$ dan saat $Z_2 = 1$ atau mengguankan tabel distribusi normal standar.

$$P(65 < X \leq 75) = P(-1 < Z \leq 1) = P(Z \leq 1) - P(Z \leq -1)$$
$$= 0{,}8413 - 0{,}1587 = 0{,}6826$$

DAFTAR PUSTAKA

Maity, R. (2018). Probability distributions and their applications. In Springer Transactions in Civil and Environmental Engineering (pp. 93–94). https://doi.org/10.1007/978-981-10-8779-0_4

Ross, S. M. (2010). Introduction to probability models (Tenth). Academic Press. https://faculty.ksu.edu.sa/sites/default/files/introduction-to-probability-model-s.ross-math-cs.blog_.ir_.pdf

BAB 4
METODE PEMUSATAN DAN PENYEBARAN DATA : NILAI HARAPAN DAN VARIANS

Oleh Yusup Junaedi

4.1 Nilai Harapan (Nilai Ekspektasi)

Nilai harapan atau nilai ekspektasi sering disebut sebagai mean. Jika X merupakan variabel random melalui distribusi peluang $f(x)$, nilai harapan dari X dapat didefinisikan sebagai berikut.

Definisi 4.1.

$E(x) = \mu = \sum_x x \cdot f(x)$ apabila X merupakan variabel random diskrit.

$E(x) = \mu = \int_{-\infty}^{\infty} x \cdot f(x)dx$, X merupakan variabel random kontinu.

Contoh 4.1.

Tentukan nilai ekspektasi jumlah matematikawan pada sebuah panitia beranggotakan 3 orang yang ditentukan secara acak dari kelompok yang terdiri atas 4 matematikawan dan 3 fisikawan.

Penyelesaian :

Apabila X menentukan banyaknya matematikawan dalam kepanitiaan. Variabel random X ialah variabel random diskrit, sehingga nilai-nilai X sebagai berikut:

$x = 0$, apabila tidak ada matematikawan terpilih

$x = 1$, apabila hanya 1 orang matematikawan terpilih

$x = 2$, apabila terdapat 2 orang matematikawan terpilih

$x = 3$, apabila terdapat 3 orang terpilih adalah matematikawan.

Berdasarkan nilai x, diperoleh distribusi peluang X sebagai berikut:

$$f(x) = \frac{\binom{4}{x}\binom{3}{3-x}}{\binom{7}{3}}, x = 0,1,2,3.$$

Melalui distribusi peluang, maka

$$f(0) = \frac{1}{35}, f\,(1) = \frac{12}{35}, f\,(2) = \frac{18}{35}\ dan\ f\,(3) = \frac{14}{35}.$$

Sehingga:

$$\mu = E(X)$$
$$= (0)\left(\frac{1}{35}\right) + (1)\left(\frac{12}{35}\right) + (2)\left(\frac{18}{35}\right) + (3)\left(\frac{14}{35}\right)$$
$$= \frac{12}{7} = 1{,}7$$

Dengan demikian, apabila kepanitiaan terdiri dari 3 orang yang dipilih secara acak berulang kali dari 4 matematikawan dan 3 fisikawan, rata-rata anggota panitia tersebut akan terdiri dari 1,7 matematikawan.

Contoh 4.2.

Jika X adalah variabel random yang merepresentasikan usia (umur) untuk jam suatu jenis lampu neon. Peluang dari fungsi padatnya dinyatakan sebagai berikut.

$$f(X) = \begin{cases} \frac{20.000}{x^3}, x > 100 \\ 0, untuk\ x\ lainnya \end{cases}$$

Tentukan nilai harapan dari umur neon tersebut.

Penyelesaian:

Variabel random X ialah variabel random kontinu.

Melalui definisi yang telah disajikan di atas diperoleh

$$\mu = E(X) = \int_{100}^{\infty} x\frac{20.000}{x^3}dx = \int_{100}^{\infty} x\frac{20.000}{x^2}dx = 200$$

Dengan demikian, lampu neon tersebut memiliki umur rata-rata yang diharapkan sekitar 200 jam.

Selanjutnya, misalkan terdapat variabel random baru $g(X)$, yang nilainya tergantung pada X. Artinya, setiap nilai $g(X)$ dapat dihitung jika nilai X telah ada. Sebagai contoh, jika $g(X)$ sama dengan X^2atau $3X-1$, sehingga X bernilai 2 dan $g(X)$ bernilai $g(2)$. Apabila X variabel acak

$P[g(X)=0]=P(X-0)=f(0)$

$P[g(X)=1]=P(X=-1)+P(X=1)=f(-1)+f(1)$

$P[g(X)=4]=P(X=2)=f(2)$

Sehingga distribusi peluang $g(X)$ dinyatakan sebagai berikut:

$g(X)$	0	1	4
$P[g(X)=g(x)]$	$f(0)$	$f(-1)+f(1)$	$f(2)$

Berdasarkan definisi dari nilai ekspektasi variabel random, dihasilkan:

$$\mu_{g(x)} = E[g(X)]$$
$$= (-1)^2 f(-1) + (0)^2 f(0) + (1)^2 f(1) + (2)^2 f(2)$$
$$= \sum g(x) f(X)$$

Teorema 4.1.

Misalkan X suatu peubah acak dengan distribusi peluang $f(x)$. Rataan atau nilai harapan peubah acak $g(X)$ adalah:

$$\mu_{g(x)} = E[g(X) = \sum g(x) f(X)$$

Bila X diskrit, dan

$$\mu_{g(x)} = E[g(X) = \int_{-\infty}^{\infty} g(X) f(X) dx$$

Bila X kontinu.

Contoh 4.3.

Jumlah kendaraan H yang datang ke tempat pencucian mobil yang setiap harinya antara pukul 13.00 hingga 14.00 mengikuti distribusi peluang sebagai berikut.

x	4	5	6	7	8	9
$P(X=x)$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{6}$

Penyelesaian :

Berdasarkan teorema 4.1, nilai ekspektasi penerimaan karyawan adalah sebagai berikut

$$E[g(X)] = E(2X-1)$$

$$= \sum_{4}^{9} (2x-1)(f(x))$$

$$= (7)\left(\frac{1}{12}\right) + (9)\left(\frac{1}{12}\right) + (11)\left(\frac{1}{4}\right) + (13)\left(\frac{1}{4}\right) + (15)\left(\frac{1}{6}\right) + (17)\left(\frac{1}{6}\right)$$

$$= Rp.\,12{,}67$$

Contoh 4.4.

Apabila Z merupakan suatu variabel random melalui fungsi densitas probabilitas:

$$f(x,y) = \begin{cases} \frac{x^2}{3}, -1 < x < 2 \\ 0, untuk\ x\ lain \end{cases}$$

Tentukan nilai ekspektasi dari $g(X) = 4X + 3$!

Penyelesaian:

$$E[g(X)] = E(4X+3) = \int_{-1}^{2} (4x+3) \cdot \frac{x^2}{3} dx$$

$$= \frac{1}{3} \int_{-1}^{2} (4x^3 + 3x^2)dx$$

$$= \frac{1}{3}[x^4 + x^3]_{-1}^{2} = 8$$

Teorema 4.2

Misalkan X dan Y merupakan variabel acak diskrit, maka:

a. Untuk sembarang konstanta a, maka:
   $E(a) = a$ dan $E(aX) = aE(X)$
b. $E(X + Y) = E(X) + E(Y)$
c. Jika terdapat konstanta a, b, c maka:
   $E(aX + bY + c) = aE(X) + bE(Y) + c$

Contoh 4.5.

Seorang pengusaha kecantikan baru-baru ini meluncurkan 2 jenis krim pelembap yang dirancang khusus bagi perempuan dewasa perempuan remaja. Namun, mengingat banyaknya berbagai jenis kecantikan serupa yang menyebarluas dipasaran, penjualan produk kencantikan ini tidak dapat diperkirakan dengan pasti. Berdasarkan hasil penjualan sebelumnya, estimasi penjualan produk kecantikan berupa pelembap untuk perempuan dewasa adalah sebagai berikut:

Jumlah produk kecantikan terjual	1.000	3.000	5.000	10.000
Peluang $(f(x)$	0,1	0,3	0,4	0,2

Namun prediksi penjualan produk kecantikan untuk perempuan remaja yakni:

Jumlah produk kecantikan terjual	300	500	750
Peluang $(f(x)$	0,4	0,5	0,1

Apabila X mengungkapkan jumlah produk kecantikan untuk perempuan dewasa yang terjual, dan Y menyatakan jumlah produk kecantikan bagi perempuan remaja yang terjual.

1. Berapakah nilai harapan dari jumlah penjualan produk kecantikan bagi perempuan dewasa dan remaja?
2. Produsen kosmetik menetapkan keuntungan sebesar 2.000 dolar dalam setiap produk kecantikan dewasa dan 3.500 dolar untuk setiap botol kecantikan remaja. Berapakah nilai harapan keuntungan dalam masing-masing produk?

Berapakah nilai ekspektasi total keuntungan kedua jenis krim tersebut?

Penyelesaian:

1. Nilai harapan untuk kecantikan perempuan dewasa

$$E(X) = \sum xf(x)$$
$$= (1.000)(0,1) + (3.000)(0,3) + (5.000)(0,4) + (10.000)(0,2)$$
$$= 100 + 900 + 2.000 + 2.000 = 5.000$$

Nilai harapan untuk kecantikan perempuan remaja

$$E(Y) = \sum yf(y)$$
$$= (300)(0,4) + (500)(0,5) + (750)(0,1)$$
$$= 120 + 250 + 75 = 445$$

2. Keuntungan untuk perempuan dewasa (D) ialah $\$2.000$, maka $D = 2.000X$

$$E(D) = E(2.000X) = 2.000E(X) = 2.000(5.000) = \$10.000.000$$

Keuntungan produk kecantikan remaja (R) sebesar $\$3.500$, maka $R + 3.500Y$

Jadi

$$E(R) = E(3.500Y) = 3.500E(Y) = 3.500(445) = \$1.557.500$$

3. Jumlah keuntungan (T) antara produk kecantikan perempuan dewasa dan remaja.
   $T = D + R = 2.000X + 3.500Y$
   Jadi
   $E(T) = E(D + R)$
   $= E(D) + E(R)$
   $= \$10.000.000 + \$1.557.500$
   $= \$11.557.500,-$

4.2 Variansi

Nilai ekspektasi dalam variabel random X ialah ukuran pada statistik yang menunjukkan titik pusat dari distribusi probabilitas. Namun, nilai ekspektasi saja tidak cukup untuk menggambarkan bentuk keseluruhan dari suatu distribusi. Untuk memahami bentuk distribusi secara lebih lengkap, diperlukan informasi mengenai variabilitasnya (Harsono, 2024). Salah satu ukuran variabilitas pada statistik ialah varians. Varians dari variabel random X, atau varians dari distribusi probabilitas X, dilambangkan $Var(X)$ dan notasinya σ_x^2 atau σ^2.

Definisi 4.2.
Misalkan X merupakan variabel acak dengan distribusi probabilitas $f(x)$ dan nilai harapan μ. Variansi dari X adalah:

$$\sigma^2 = E[(X-\mu)^2] = \sum_x (x-\pi)^2 \cdot f(x).$$

Jika X diskrit.

$\sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty}(x-\pi)^2 \cdot f(x)dx$, jika X kontinu.

Nilai $x - y$ yang ditentukan definisi 4.2. menggambarkan deviasi suatu pengamatan dari rata-ratanya. Karena deviasi ini dikuadratkan dan kemudian dirata-ratakan, nilai σ^2 akan lebih kecil

untuk kumpulan nilai x yang mendekati μ dibandingkan dengan kumpulan nilai yang berjauhan dengan μ (Walpole & Myers, 1995).

Contoh 4.6.
Misal peubah random X menyatakan jumlah motor yang dikendarai untuk berkantor yang digunakan setiap harinya. Distribusi peluang perusahaan J yaitu:

X	1	2	3
$f(x)$	0,3	0,4	0,3

Sedangkan perusahaan K yaitu:

X	0	1	2	3	4
$f(x)$	0,2	0,1	0,3	0,3	0,1

Tentukan variansi distribusi peluang perusahaan J lebih dari variansi perusahaan K.

Penyelesaian:
Untuk perusahaan J, diperoleh:

$$\mu = E(X) = (1)(0,3) + (2)(0,4) + (3)(0,3) = 2,0$$

$$\sigma^2 = \sum_{x=1}^{4} (x-2)^2 f(x)$$

$$= (1-2)^2(0,3) + (2-2)^2(0,4) + (3-2)^2(0,3) = 0,6$$

Untuk perusahaan K, diperoleh:

$$\mu = E(X) = (0)(0,2) + (1)(0,1) + (2)(0,3) + (4)(0,1) = 2,0$$

$$\sigma^2 = \sum_{x=0}^{4} (x-2)^2 f(x)$$

$$= (0-2)^2(0,2) + (1-2)^2(0,1) + (2-2)^2(0,3) + (3-2)^2(0,3) + (4-2)^2(0,1) = 1,6$$

Berdasarkan perolehan di atas, terlihat bahwa variansi jumlah motor yang digunakan oleh perusahaan J dibandingkan perusahaan K.

Teorema 4.3
Variansi variabel random X adalah $\sigma^2 = E(x^2) - \mu^2$

Pembuktian:
Permasalahan diskrit dapat dijabarkan:

$$\sigma^2 = \sum_x (x-\mu)^2 \cdot f(x)$$

$$\sigma^2 = \sum_x (x^2 - 2\mu x + \mu^2) \cdot f(x)$$

$$\sigma^2 = \sum_x x^2 \cdot f(x) - 2\mu \sum_x x \cdot f(x) + \mu^2 \sum_x x \cdot f(x)$$

Karena
$\mu \sum_x x \cdot f(x)$ dan $\sum_x x \cdot f(x) = 1$ untuk setiap distribusi probabilitas diskrit, sehingga diperoleh:

$$\sigma^2 = \sum_x x^2 \cdot f(x) - 2\mu^2 + \mu^2$$

Sehingga

$$\sigma^2 = \sum_x x^2 \cdot f(x) - \mu^2$$

$$= E(x^2) - \mu^2$$

Untuk variabel random kontinu, tahapan pembuktiannya sama namun penjumlahannya diubah oleh integral.

Contoh 4.7.

Apabila peubah acak X menentukan jumlah bagian yang cacat dalam suatu mesin jika 3 *sparepart* disampling dari rantai produksi serta diuji. Berikut merupakan distribusi peluang X

X	0	1	2	3
$f(x)$	0,51	0,38	0,10	0,01

Tentukan σ^2 dengan Teorema 4.3.

Penyelesaian :

Berdasarkan teorema 4.3.

$= E(x^2) - \mu^2$

Langkah pertama, mencari μ

$$\mu = E(X) = (0)(0{,}51) + (1)(0{,}38) + (2)(0{,}10) + (3)(0{,}01)$$
$$= 0{,}61$$

Selanjutnya

$E(X^2) = (0)(0{,}51) + (1)(0{,}38) + (4)(0{,}10) + (9)(0{,}01) = 0{,}87$

Sehingga

$\sigma^2 = 0{,}87 - (0{,}61)^2 = 0{,}4979.$

Contoh 4.8.

Apabila X merupakan suatu variabel random dengan fungsi densitas probabilitas:

$$f(x,y) = \begin{cases} \frac{x^2}{3}, -1 < x < 2 \\ 0, untuk\ x\ lain \end{cases}$$

Tentukakn nilai varians $g(X) = 4X + 3$!

Penyelesaian:

$$E[g(x)]^2 = E(4X+3)^2 = \int_{-1}^{2} (4X+3)^2 \cdot \frac{x^2}{3} dx$$

$$= \frac{1}{3}\int_{-1}^{2} 16x^2 + 24x + 9 \cdot \frac{x^2}{3} dx$$

$$= \frac{1}{3}\int_{-1}^{2} (16x^4 + 24x^3 + 9x^2) dx$$

$$= \frac{1}{3}\left[\frac{16}{5}x^5 + 6x^4 + 3x^3\right]_{-1}^{2}$$

$$= 74{,}2$$

4.3 Latihan Soal

1. Sebuah koin seimbang dilempar sebanyak 2 kali. Tentukan nilai ekspektasi untuk kejadian di mana lemparan kesatu memperoleh sisi belakang, dan lemparan kedua menghasilkan sisi depan.
2. Jasa cuci mobil akan dibayarkan sesuai dengan jumlah mobil yang dicuci. Apabila dalam sehari memperoleh penerimaan (ribuan rupiah) yakni 7, 9, 11, 13, 15 atau 17 dengan probabilitas $\frac{1}{12}, \frac{1}{12}, \frac{1}{4}, \frac{1}{4}, \frac{1}{6}, \frac{1}{6}$. Tentukan nilai ekspektasi untuk penghasilan jasa cuci mobil tersebut dalam setiap harinya.
3. Apabila $0{,}1{,}2\ atau\ 3$ sering terjadinya mati listrik dalam daerah tertentu selama satu bula dengan masing-masing peluang $0{,}4; 0{,}3; 0{,}2\ dan\ 0{,}1$. Tentukan rata-rata dan variansi variabel random X untuk menentukan jumlah mati listrik didaerah tersebut.

DAFTAR PUSTAKA

Harsono, I., Wonar, K., Desviona, N., Utama, R. C., Masruroh, M., Damayanti, I. D., ... & Matsaany, B. (2024). *Buku Ajar Metode Statistika 1*. PT. Sonpedia Publishing Indonesia.

Walpole, R. E., & Myers, R. H. (1995). Ilmu peluang dan Statistika untuk Insinyur dan Ilmuwan. *Bandung: Itb*, 149-150.

BAB 5
TRANSFORMASI VARIABEL ACAK: MENYEDERHANAKAN DAN MEMAHAMI DISTRIBUSI

Oleh Agus winarji

5.1 Pendahuluan

Transformasi variabel acak merupakan metode statistik dan probabilitas yang digunakan untuk mengubah satu variabel acak menjadi variabel acak lainnya, dengan tujuan menyederhanakan analisis atau memahami distribusi baru yang dihasilkan dari transformasi tersebut. Konsep ini penting dalam berbagai aplikasi seperti analisis data, inferensi statistik, *machine learning*, simulasi dan lainnya.

Dalam analisis data, transformasi variabel acak digunakan untuk mengatasi masalah seperti distribusi data yang tidak normal atau skala data yang tidak sesuai untuk analisis tertentu. Misalnya, transformasi logaritmik dapat digunakan untuk mengurangi data yang sangat miring, sehingga menghasilkan distribusi yang lebih mudah dianalisis (Kutner, 2005).

Dalam inferensi statistik, transformasi variabel acak memungkinkan perhitungan yang lebih efisien, terutama ketika distribusi asli sulit ditangani. Sebagai contoh, metode transformasi sering digunakan dalam penghitungan probabilitas atau penurunan estimasi parameter yang melibatkan distribusi kompleks (Hogg, 2019).

Selain itu, transformasi variabel acak juga sangat berguna dalam simulasi. Banyak algoritma simulasi, seperti metode *Monte Carlo*, mengandalkan kemampuan untuk menghasilkan sampel dari distribusi tertentu melalui transformasi variabel acak (Robert, 2004). Dengan memetakan sampel dari distribusi yang lebih mudah dihasilkan, seperti distribusi uniform, ke distribusi target, kita dapat

melakukan simulasi fenomena acak dengan tingkat akurasi yang tinggi.

5.2 Transformasi Peubah Acak

5.2.1 Transformasi Peubah Acak Diskrit

Transformasi peubah acak diskrit adalah proses untuk mengubah suatu peubah acak diskrit X menjadi peubah acak lain Y melalui fungsi tertentu. Proses ini dilakukan dengan mendefinisikan hubungan antara X dan Y menggunakan fungsi transformasi $g(x)$, sehingga:

$$Y = g(x)$$

Jika (X) adalah variabel acak diskrit dengan fungsi probabilitas $(p_X(x))$, maka fungsi probabilitas $(p_Y(y))$ untuk $Y = g(X)$ diberikan oleh:

$p_Y(y) = P[Y = y] = P[g(X) = y] = P[X = g^{-1}(y)] = p_x(g^{-1}(y))$

Sehingga $p_Y(y) = p_x\left(g^{-1}(y)\right)$. (Hogg *et al.*, 2019; Herrhyanto, 2009)

Contoh 1

Misalkan fungsi probabilitas variabel acak X adalah:

$p_X(x) = 2^{-x}, x = 1, 2, 3, \ldots$

Tentukan fungsi probabilitas variabel acak $Y = X^4 + 1$!

Penyelesaian

Relasi X dan Y diberikan dengan $y = x^4 + 1$ dan dapat dilihat pada tabel 5.1 berikut:

Tabel 5.1. Relasi nilai X dan Y

X	1	2	3	…
$Y = X^4 + 1$	2	17	82	…

Nilai-nilai yang mungkin dari $Y \in \{2, 17, 82, \ldots\}$

Invers dari $x = (y-1)^{\frac{1}{4}}$

Jadi fungsi probabilitas dari variabel acak Y adalah:

$$p_Y(y) = 2^{-(y-1)^{\frac{1}{4}}}, y = 2, 17, 82, \dots$$

Contoh 2

Misalkan ada sebuah dadu enam sisi standar. Variabel acak X melambangkan hasil lemparan dadu, dengan $X \in \{1,2,3,4,5,6\}$. Probabilitas setiap nilai adalah sama:

$$p_X(x) = \frac{1}{6}, x \in \{1,2,3,4,5,6\}$$

Tentukan distribusi variabel acak Y yang didefinisikan sebagai kuadrat dari hasil lemparan dadu, yaitu:

$Y = X^2$

Penyelesaian

1. Menentukan Ruang Sampel Y

   Karena $Y = X^2$ dan X hanya dapat mengambil nilai dari himpunan $\{1,2,3,4,5,6\}$, maka nilai Y adalah:

   $Y \in \{1^2, 2^2, 3^2, 4^2, 5^2, 6^2\} = \{1,4,9,16,25,36\}$

2. Menghitung Probabilitas $P_Y(y)$

   Karena transformasi $Y = X^2$ dan X adalah fungsi satu-satu (setiap nilai X unik menghasilkan nilai Y), maka:

   $p_Y(y) = p(x), y = x^2$

   Sehingga, probabilitas masing-masing nilai y adalah:

   $p_Y(1) = p_X(1) = \frac{1}{6}$

   $p(4) = p(2) = \frac{1}{6}$

   $p_Y(9) = p(3) = \frac{1}{6}$

   $p_Y(16) = p_X(4) = \frac{1}{6}$

   $p_Y(25) = p_X(5) = \frac{1}{6}$

   $p_Y(36) = p_X(6) = \frac{1}{6}$

3. Distribusi Probabilitas

Distribusi probabilitas untuk Y dapat dirangkum dalam tabel 5.2 berikut:

Tabel 5.2. Tabel distribusi probabilitas variabel acak Y

y	1	4	9	16	25	36
$p_Y(y)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Distribusi variabel acak hasil transformasi $Y = X^2$ memiliki nilai probabilitas yang sama untuk semua y, karena setiap nilai X memiliki peluang yang sama. Contoh ini menunjukkan bagaimana transformasi variabel acak diskrit dapat dilakukan untuk memahami karakteristik distribusi baru.

Contoh 3

Misalkan kita memiliki 3 koin yang dilempar secara bersamaan. Setiap koin memiliki dua kemungkinan hasil, yaitu Gambar (G) atau Angka (A), dengan peluang masing-masing P(G)=0.5 dan P(A)= 0.5. Jika Variabel acak X merupakan jumlah total koin yang muncul Gambar (G) dalam tiga lemparan, tentukan distribusi probabilitas dari transformasi variabel acak Y berdasarkan fungsi $Y = 2X + 1$!

Penyelesaian

1. Menentukan Distribusi Probabilitas untuk variabel acak X

   Variabel X melambangkan jumlah koin yang muncul gambar (G), yang dapat bernilai:

   $X \in \{0,1,2,3\}$

   Peluang setiap nilai X dihitung menggunakan distribusi binomial:

   $$p_X(x) = \binom{3}{x}.(0{,}5)^x.(0{,}5)^{3-x}$$

   Untuk $x = 0$

   $$p_X(0) = \binom{3}{3}.(0{,}5)^0.(0{,}5)^{3-0} = (1).(1).(0{,}125) = 0{,}125$$

Untuk $x = 1$

$p_X(1) = \binom{3}{1}.(0{,}5)^1.(0{,}5)^{3-1} = (3).(0{,}5).(0{,}25) = 0{,}375$

Untuk $x = 2$

$p_X(2) = \binom{3}{2}.(0{,}5)^2.(0{,}5)^{3-2} = (3).(0{,}25).(0{,}5) = 0{,}375$

Untuk $x = 3$

$p_X(3) = \binom{3}{3}.(0{,}5)^3.(0{,}5)^{3-3} = (1).(0{,}125).(1) = 0{,}125$

Tabel distribusi probabilitas variabel acak X dapat dirangkum pada tabel 5.3 berikut:

Tabel 5.3. Distribusi probabilitas variabel acak $\boldsymbol{X}$

x	0	1	2	4
$p_X(x)$	0,125	0,375	0,375	0,125

2. Menentukan distribusi probabilitas untuk variabel acak Y

   Dengan fungsi transformasi $Y = 2X + 1$, nilai Y untuk setiap X dapat dihitung:

   $Y = 2X + 1 \rightarrow Y \in \{1,3,5,7\}$

   Relasi nilai X dan Y dapat dilihat pada tabel 5.4 berikut:

Tabel 5.4. Relasi nilai $\boldsymbol{X}$ dan $\boldsymbol{Y}$

X	0	1	2	3
$Y = 2X + 1$	1	3	5	7

Karena fungsi transformasi $Y = 2X + 1$ adalah fungsi satu-satu (unik untuk setiap X), peluang untuk Y langsung diturunkan dari peluang X:

$p_Y(y) = P_X(x)$, di mana $y = 2x + 1$

$p_Y(1) = p_X(0) = 0{,}125$

$p_Y(3) = p_X(1) = 0{,}375$

$p_Y(5) = p_X(2) = 0{,}375$

$p_Y(7) = p_X(3) = 0{,}125$

Tabel distribusi probabilitas variabel acak Y dapat dirangkum pada tabel 5.5 berikut:

Tabel 5.5. Distribusi probabilitas variabel acak Y

y	1	3	5	7
$p_Y(y)$	0,125	0,375	0,375	0,125

5.2.2 Transformasi Peubah Acak Kontinu

Transformasi peubah acak kontinu adalah proses mengubah suatu peubah acak kontinu X menjadi peubah acak kontinu lain Y menggunakan fungsi tertentu $g(X)$. Transformasi ini melibatkan penghitungan fungsi densitas probabilitas (PDF) atau fungsi distribusi kumulatif (CDF) dari Y berdasarkan distribusi X dan hubungan antara keduanya.

Dalam transformasi peubah acak kontinu, nilai Y yang diperoleh melalui transformasi juga bersifat kontinu, dan distribusi Y bergantung pada sifat fungsi transformasi $g(X)$.

Jika X adalah variabel acak kontinu dengan fungsi densitas probabilitas $(f_X(x))$, maka densitas probabilitas $f_Y(y)$ untuk $Y = g(X)$ dapat diperoleh dengan cara:

1. Jika $g(X)$ monoton (naik atau turun) maka $f_Y(y)$ dihitung dengan

   $$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|$$

   di mana $\frac{dx}{dy}$ adalah turunan dari invers fungsi $g(x)$, dan x adalah nilai X yang sesuai dengan $Y = y$. (Wasserman, 2014; Wackerly *et al.*, 2009)

2. Jika $g(X)$ tidak monoton, perhatikan setiap interval monoton tersebut. Lakukan penjumlahan terhadap kontribusi PDF dari setiap interval. (Hogg *et al.*, 2019)

Contoh 1

Misalkan X adalah peubah acak kontinu dengan fungsi densitas probabilitas (PDF):

$$f_X(x) = \begin{cases} 2x, 0 \leq x \leq 1 \\ 0, lainnya \end{cases}$$

Kita ingin mentransformasi X menjadi $Y = 3X - 1$. Tentukan fungsi densitas probabilitas $f_Y(y)$!

Penyelesaian

Fungsi transformasi relasi antara X dan Y diberikan oleh:

$$y = g(x) = 3x - 1$$

Invers dari $x = \frac{y+1}{3}$ dan turunannya $\frac{dx}{dy} = \frac{1}{3}$

Maka fungsi densitas probabilitas Y adalah:

$$\begin{aligned} f_Y(y) &= f_X(x)\left|\frac{dx}{dy}\right| \\ &= 2\left(\frac{y+1}{3}\right)\left(\frac{1}{3}\right), 0 \leq \frac{y+1}{3} \leq 1 \\ &= \frac{2(y+1)}{9}, -1 \leq y \leq 2 \end{aligned}$$

Jadi fungsi densitas probabilitas variabel acak Y adalah:

$$f_Y(y) = \begin{cases} \dfrac{2(y+1)}{9}, -1 \leq y \leq 2 \\ 0, lainnya \end{cases}$$

Distribusi Y adalah distribusi berbentuk segitiga, dengan fungsi densitas probabilitas (PDF) meningkat (*increasing*) secara linier dalam interval $[1-, 2]$. Transformasi linier $Y = 3X - 1$ mengubah bentuk PDF asal $f_X(x)$ menjadi distribusi baru $f_Y(x)$ yang tetap kontinu, tetapi dengan skala dan batas yang berbeda.

Contoh 2

Misalkan X adalah peubah acak kontinu dengan fungsi densitas probabilitas:

$$f_X(x) = \begin{cases} 1, -1 \leq x \leq 1 \\ 0, lainnya \end{cases}$$

Kita ingin melakukan transformasi X menjadi $Y = X^2$. Tentukan fungsi densitas probabilitas $f_Y(y)$!

Penyelesaian

Fungsi $g(X) = X^2$ tidak monoton karena nilainya sama untuk x positif dan negatif $(g(-x) = g(x))$. Dalam hal ini, $g(x)$ monoton naik untuk $x \geq 0$ dan monoton turun untuk $x \leq 0$.

Karena X berada di interval $[-1,1]$, maka $Y = X^2$ berada di interval $[0,1]$

Fungsi transformasi relasi antara X dan Y diberikan oleh:

$y = x^2$

Invers dari $x = \pm\sqrt{y}$, $x_1 = \sqrt{y}$ dan $x_2 = -\sqrt{y}$

Turunannya adalah:

$$\frac{dx_1}{dy} = \frac{1}{2\sqrt{y}}$$

$$\frac{dx_2}{dy} = -\frac{1}{2\sqrt{y}}$$

Maka fungsi densitas probabilitas dari variabel acak Y adalah:

$$\begin{aligned} f_Y(y) &= f_X(x_1)\left|\frac{dx_1}{dy}\right| + f_X(x_2)\left|\frac{dx_2}{dy}\right| \\ &= 1.\left|\frac{1}{2\sqrt{y}}\right| + 1.\left|-\frac{1}{2\sqrt{y}}\right| \\ &= \frac{1}{2\sqrt{y}} + \frac{1}{2\sqrt{y}} \\ &= \frac{1}{\sqrt{y}} \end{aligned}$$

Jadi fungsi densitas probabilitas dari variabel acak Y adalah:

$$f_Y(y) = \begin{cases} \dfrac{1}{\sqrt{y}}, 0 \leq y \leq 1 \\ 0, lainnya \end{cases}$$

Pada contoh ini, kita melihat bahwa ketika $g(x)$ tidak monoton, kontribusi dari setiap interval monoton perlu dijumlahkan untuk mendapatkan PDF peubah acak baru. Metode ini berlaku untuk berbagai transformasi yang lebih kompleks.

5.3 Penerapan Transformasi Variabel Acak

Beberapa contoh penerapan praktis transformasi variabel acak dalam bidang statistik, probabilitas dan ilmu data, diantaranya:

1. Pengubahan distribusi

   Transformasi variabel acak digunakan untuk mengubah bentuk distribusi agar lebih sesuai untuk analisis statistik, misalnya mengubah data menjadi distribusi normal atau mengurangi *skewness* (distribusi data condong/ miring).

   Contoh:

   - Transformasi logaritmik $(Y = \log X)$ untuk mengurangi skewness pada data (Kutner, 2005; Osborne, 2002).
   - Transformasi kuadrat $(Y = X^2)$ untuk data yang mengandung nilai negative (Hair, 2014).

2. Penyesuaian skala atau lokasi

   Transformasi linier seperti $Y = aX + b$ digunakan untuk menyesuaikan skala atau lokasi data. Contoh: normalisasi data menjadi rata-rata nol dan variansi satu (*Z-Score normalization*):

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

   di mana μ adalah rata-rata dan σ adalah standar deviasi (Montgomery, 2014).

3. Perhitungan probabilitas non-standar

   Transformasi variabel acak digunakan untuk menghitung probabilitas yang sulit dianalisis secara langsung. Contoh: jika X adalah waktu kegagalan mesin dengan distribusi eksponensial, transformasi $Y = \ln X$ dapat membantu menghitung probabilitas waktu kegagalan tertentu (Montgomery, 2014).

4. Analisis hidrologi atau risiko

   Di bidang hidrologi, transformasi digunakan untuk menganalisis data yang mengikuti distribusi ekstrem. Contoh**:** Transformasi $Y = -\log(X)$ digunakan untuk mengubah data menjadi distribusi *Weibull* atau *Gumbel* untuk analisis banjir atau curah hujan ekstrem. Transformasi ini sangat berguna untuk mengidentifikasi pola-pola ekstrem dalam data dan

memodelkan kejadian langka untuk perencanaan mitigasi risiko (Haan, 2002).

Penerapan transformasi variabel acak kontinu sangat luas dalam berbagai bidang. Pemilihan transformasi yang tepat memungkinkan analisis lebih mendalam dan penyederhanaan model matematika yang digunakan.

Latihan

1. Misalkan kita memiliki 4 koin yang dilempar secara bersamaan. Setiap koin memiliki dua kemungkinan hasil, yaitu Gambar (G) atau Angka (A), dengan peluang masing-masing P(G)=0.5 dan P(A)= 0.5. Jika Variabel acak X merupakan jumlah total koin yang muncul angka (A) dalam empat lemparan, tentukan distribusi probabilitas dari transformasi variabel acak Y berdasarkan fungsi $Y = 2X + 1$!
2. Diketahui variabel acak X dengan fungsi probabilitas:

   $$P_X(x) = \frac{1}{3}, x = 1,2,3$$

   Tentukan fungsi probabilitas variabel acak Y yang ditransformasikan oleh:
   a. $Y = 2x + 1$
   b. $Y = X^2$
3. Diketahui variabel acak X dengan fungsi densitas probabilitas:

   $$f_X(x) = \begin{cases} 2(1-x), 0 \leq x \leq 1 \\ 0, lainnya \end{cases}$$

   Tentukan fungsi densitas probabilitas variabel acak Y yang ditransformasikan oleh:
   a. $Y = 2X - 1$
   b. $Y = 1 - 2X$
   c. $Y = X^2$

DAFTAR PUSTAKA

Haan, C. T. 2002. *Statistical Methods in Hydrology* (2nd ed.). Iowa State Press

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. 2014. *Multivariate Data Analysis* (7th ed.). Pearson.

Herrhyanto, Nar., Gantini, Tuti. 2009. *Pengantar Statistika Matematis*. Bandung: CV. Yrama Widya

Hogg, Robert V., McKean, Joseph W., Craig, Allen T. 2019. *Introduction to Mathematical Statistics Eighth Edition.* Pearson Education.

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. 2005. *Applied Linear Statistical Models* (5th ed.). McGraw-Hill

Montgomery, D. C., & Runger, G. C. 2014. *Applied Statistics and Probability for Engineers*. John Wiley & Sons

Osborne, J. W. 2002. Notes on the use of data transformations. *Practical Assessment, Research, and Evaluation*, 8(6), 1–8

Robert, C., & Casella, G. 2004. *Monte Carlo Statistical Methods* (2nd ed.). Springer.

Wackerly, Dennis D., Mendenhall, William., Scheaffer, Richard L. 2009. *Mathematical Statistics with Aplications*. Cengage Lerning.

Wasserman, Larry. 2014. *All of Statistics A Concise Course in Statistical Inference*. Springer

BAB 6
METODE PENDUGAAN PARAMETER: ESTIMASI KARAKTERISTIK POPULASI

Oleh Ridha Yuniara

Dalam statistik, estimasi parameter adalah proses untuk memperkirakan nilai parameter tertentu dari suatu populasi, berdasarkan data yang diperoleh dari sampel. Estimasi ini sangat penting dalam statistika inferensial, di mana kita membuat kesimpulan tentang populasi berdasarkan informasi terbatas dari sampel yang diambil. Contoh parameter populasi meliputi:

1. Rata-rata (*mean*) populasi, yang biasa dilambangkan dengan μ
2. Proporsi populasi, dilambangkan dengan *p*
3. Variansi populasi, dilambangkan dengan σ^2

Karena seringkali kita tidak bisa mengukur seluruh populasi, kita mengambil sampel dari populasi tersebut dan menggunakan data dari sampel untuk memperkirakan parameter populasi.

Adapun tujuan utama estimasi parameter adalah untuk:

1. **Memberikan gambaran** yang representatif dari parameter populasi menggunakan data sampel.
2. **Mengurangi ketidakpastian** tentang parameter yang sebenarnya dengan memberikan perkiraan atau interval.
3. **Membantu pengambilan keputusan** dalam situasi di mana data populasi tidak sepenuhnya dapat dijangkau.

Selanjutnya, ada dua metode utama dalam pendugaan parameter, yaitu estimasi titik (*point estimation)* dan estimasi interval (*interval estimation*).

6.1 Estimasi Titik (*Point Estimation*)

Estimasi titik adalah suatu teknik dalam statistika untuk memperkirakan nilai suatu parameter populasi berdasarkan data yang diperoleh dari sampel. Estimasi titik memberikan satu nilai tunggal yang dianggap sebagai perkiraan terbaik untuk parameter yang tidak diketahui dari populasi (Cahyono, 2018). Dengan kata lain, estimasi titik menggunakan informasi dari sampel untuk memberi perkiraan atau estimasi terhadap parameter populasi.

Tujuan utama dari estimasi titik adalah untuk memberikan satu nilai yang paling mungkin atau paling representatif sebagai perkiraan parameter populasi. Sebagai contoh, kita mungkin ingin mengestimasi rata-rata tinggi badan populasi, tetapi hanya memiliki data dari sebagian kecil populasi (sampel), sehingga kita menggunakan rata-rata sampel sebagai estimasi titik untuk rata-rata populasi.

Beberapa parameter populasi yang sering ditaksir dengan estimasi titik antara lain:

1. **Rata-rata Populasi (Mean, μ)**: Rata-rata populasi yang tidak diketahui dapat ditaksir dengan rata-rata sampel.
2. **Proporsi Populasi (Proportion, *p*)**: Proporsi populasi yang tidak diketahui dapat ditaksir dengan proporsi sampel.
3. **Variansi Populasi (Variance, σ^2)**: Variansi populasi yang tidak diketahui dapat ditaksir dengan variansi sampel.
4. **Deviasi Standar Populasi (Standard Deviation, σ)**: Deviasi standar populasi yang tidak diketahui dapat ditaksir dengan deviasi standar sampel.

Berikut adalah beberapa rumus dasar untuk menghitung estimasi titik dari berbagai parameter populasi.

1. **Estimasi Titik untuk Rata-rata Populasi**: Jika kita ingin mengestimasi rata-rata populasi μ, kita menggunakan rata-rata sampel $\overline{x}$ sebagai estimasi titik.

   $$\hat{\mu} = \overline{x}$$

   Di mana:

   - $\hat{\mu}$ adalah estimasi titik untuk rata-rata populasi.

- $\overline{x}$ adalah rata-rata sampel, yang dihitung sebagai:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- Di mana $x_1, x_2, ..., x_n$ adalah nilai data sampel, dan n adalah ukuran sampel.

2. **Estimasi Titik untuk Proporsi Populasi**: Untuk proporsi populasi *p*, estimasi titik menggunakan proporsi sampel $\hat{p}$, yang dihitung sebagai:

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

Di mana:

- $\hat{p}$ adalah estimasi titik untuk proporsi populasi.
- x adalah jumlah kejadian yang sukses dalam sampel (misalnya jumlah individu yang memiliki suatu sifat atau karakteristik).
- n adalah ukuran sampel.

3. **Estimasi Titik untuk Variansi Populasi**: Variansi populasi σ^2dapat ditaksir dengan variansi sampel s^2, yang dihitung sebagai:

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

Di mana s^2 adalah estimasi titik untuk variansi populasi, x_i adalah nilai-nilai dalam sampel, dan $\overline{x}$ adalah rata-rata sampel.

4. **Estimasi Titik untuk Deviasi Standar Populasi**: Deviasi standar populasi σ dapat ditaksir dengan deviasi standar sampel *s*, yang dihitung sebagai:

$$s = \sqrt{s^2}$$

Di mana s adalah estimasi titik untuk deviasi standar populasi, dan s^2 adalah estimasi titik untuk variansi populasi.

Estimasi titik yang baik harus memiliki beberapa karakteristik berikut:

1. **Tidak Bias (*Unbiasedness*)**: Estimasi titik yang tidak bias adalah estimasi yang, rata-rata, akan memberikan nilai yang benar-benar mendekati parameter populasi yang sebenarnya. Artinya, tidak ada kecenderungan sistematis dalam kesalahan estimasi. Sebagai contoh, rata-rata sampel ($\bar{x}$) adalah estimasi tidak bias untuk rata-rata populasi μ.

2. **Efisiensi (*Efficiency*)**: Estimasi titik yang efisien adalah estimasi yang memiliki variansi terkecil di antara estimasi yang tidak bias. Estimasi yang lebih efisien menghasilkan perkiraan yang lebih konsisten dan lebih tepat.

3. **Konsistensi (*Consistency*)**: Estimasi yang konsisten berarti bahwa dengan bertambahnya ukuran sampel, estimasi titik akan semakin mendekati nilai parameter populasi yang sebenarnya. Artinya, semakin besar ukuran sampel, semakin akurat estimasi yang dihasilkan.

4. **Simplicity**: Estimasi titik yang baik adalah estimasi yang mudah dihitung dan dipahami. Meskipun estimasi yang lebih kompleks bisa lebih akurat, dalam banyak kasus, kesederhanaan adalah keuntungan.

Untuk menghitung estimasi titik, kita bisa menggunakan beberapa metode, antara lain:

1. Metode Maximum Likelihood
2. estimasi parameter dengan cara memaksimalkan kemungkinan data Estimation (MLE)

MLE adalah metode yang digunakan untuk menghitung sampel yang diamati. MLE sering kali menghasilkan estimasi yang tidak bias dan efisien.

Prinsip Dasar MLE

Dalam MLE, kita ingin menemukan nilai parameter θ (misalnya, rata-rata atau variansi dalam distribusi normal) yang memaksimalkan kemungkinan (*likelihood*) dari data yang sudah ada. Dengan kata lain, kita mencari parameter yang membuat data yang kita amati paling "terlihat" atau paling "mungkin" terjadi berdasarkan model yang kita pilih.

Langkah-langkah MLE

1. Menulis Fungsi Likelihood
   Fungsi likelihood $L(\theta)$ adalah probabilitas data yang diamati $X = (x_1, x_2, \ldots, x_n)$ diberikan parameter θ. Jika data independen, maka fungsi likelihood adalah hasil perkalian dari fungsi kepadatan probabilitas (PDF) atau massa probabilitas (PMF) untuk setiap observasi x_i:
   $$L(\theta|X)=\prod_{i=1}^{n} P(x_i|\theta)$$
2. Menulis Log-Likelihood
   Karena seringkali lebih mudah untuk bekerja dengan log-likelihood, kita mengambil logaritma natural dari fungsi likelihood:
   $$\ln L(\theta|X)=\sum_{i=1}^{n} \ln P(x_i|\theta)$$
   Log-likelihood ini mengubah produk menjadi jumlah, yang lebih mudah dihitung dan dioptimalkan.
3. Maksimalkan Log-Likelihood
   Setelah kita menulis fungsi log-likelihood, langkah selanjutnya adalah mencari nilai θ yang memaksimalkan fungsi tersebut. Secara matematis, ini dilakukan dengan mengambil turunan pertama dari log-likelihood terhadap θ, lalu menyamakannya dengan nol untuk menemukan titik maksimum.
   $$\frac{d}{d\theta} \ln L(\theta|X) = 0$$
4. Penyelesaian untuk θ
   Setelah mendapatkan turunan pertama, kita selesaikan persamaan tersebut untuk θ. Nilai θ yang memaksimalkan log-likelihood ini adalah estimasi maksimum likelihood untuk parameter yang tidak diketahui.

Contoh Penerapan MLE

Misalkan kita memiliki data yang diambil dari distribusi **Normal** dengan parameter yang tidak diketahui μ(rata-rata) dan σ^2 (varians), dan kita ingin memperkirakan nilai-nilai parameter tersebut.

1. **Menulis Fungsi Likelihood**

   Fungsi distribusi normal untuk satu data xix_ixi adalah:

   $$f(x_i|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\, exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

   Maka, fungsi likelihood untuk data $X = (x_1, x_2, \ldots, x_n)$ adalah:

   $$L(\mu,\sigma^2|X) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}}\, exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

2. **Menulis Log-Likelihood**

   Ambil logaritma dari fungsi likelihood:

   $$\ln L(\mu,\sigma^2|X) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

3. **Maksimalkan Log-Likelihood**

   Untuk memaksimalkan log-likelihood terhadap μ\muμ dan σ2\sigma^2σ2, kita ambil turunan pertama dari log-likelihood:

   - Turunan terhadap μ:

     $$\frac{\partial}{\partial\mu}\ln L(\mu,\sigma^2|X) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)$$

     Setarakan dengan nol untuk mencari estimasi μ:

     $$\sum_{i=1}^{n}(x_i-\mu) = 0$$

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- Turunan terhadap σ^2:

$$\frac{\partial}{\partial \sigma^2} \ln L(\mu, \sigma^2 | X) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2$$

Setarakan dengan nol untuk mencari estimasi σ^2:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2$$

Keunggulan MLE

1. **Konsistensi:** Estimator MLE cenderung mendekati nilai parameter yang sebenarnya seiring bertambahnya jumlah data.
2. **Efisiensi:** Dalam banyak kasus, estimator MLE adalah estimator yang tidak bias dan efisien, yang berarti bahwa ia memiliki varians terkecil di antara estimator yang tidak bias lainnya.

Kelebihan dan Kekurangan Estimasi Titik

Kelebihan:

1. Mudah dihitung dan mudah dimengerti.
2. Memberikan satu nilai yang dapat digunakan sebagai perkiraan parameter populasi.

Kekurangan:

1. Tidak memberikan gambaran mengenai ketepatan atau keakuratan estimasi tersebut. Estimasi titik tidak memberi tahu kita seberapa jauh estimasi tersebut mungkin berbeda dari parameter populasi yang sebenarnya.
2. Estimasi titik sangat tergantung pada data sampel yang diambil. Jika sampel yang diambil tidak representatif atau terlalu kecil, estimasi titik bisa sangat bias atau tidak akurat.

Contoh Kasus Estimasi Titik

Misalkan kita melakukan survei terhadap 200 orang untuk mengetahui rata-rata pengeluaran bulanan mereka. Hasil survei menunjukkan bahwa rata-rata pengeluaran sampel adalah 3 juta rupiah, dan deviasi standar sampel adalah 500 ribu rupiah. Estimasi titik untuk rata-rata pengeluaran bulanan populasi adalah $\hat{\mu} = 3.000.000$ rupiah.

Jika kita ingin mengestimasi proporsi orang yang memilih produk A di suatu pasar, dan dari 500 sampel, 300 orang memilih produk A, maka estimasi titik untuk proporsi populasi adalah:

$$\hat{p} = \frac{300}{500} = 0.60$$

Ini berarti bahwa estimasi proporsi orang yang memilih produk A di pasar adalah 60%.

6.2 Estimasi interval

Estimasi interval memberikan rentang nilai untuk parameter populasi dengan tingkat kepercayaan tertentu. Rentang nilai ini menggambarkan ketidakpastian kita terhadap nilai parameter yang sebenarnya. Misalnya, jika kita mengestimasi rata-rata suatu populasi, estimasi interval akan memberikan dua angka yang menyatakan batas bawah dan batas atas rata-rata tersebut, dengan tingkat kepercayaan tertentu. Sehingga, terdapat nilai tertinggi (max) dan nilai terendah (min) dalam estimasi tersebut

Sebagai contoh:

1. **Estimasi Titik**: Rata-rata sampel $\overline{x} = 100$.
2. **Estimasi Interval**: Rata-rata populasi diperkirakan berada dalam interval [95, 105] dengan tingkat kepercayaan 95%.

Adapun tingkat kepercayaan **(*Confidence Level*)** adalah probabilitas bahwa interval estimasi yang dihitung akan mencakup nilai parameter populasi yang sebenarnya. Misalnya: Dengan tingkat kepercayaan 95%, kita berharap bahwa 95%

dari interval estimasi yang dibuat dengan cara yang sama akan mencakup parameter populasi yang sebenarnya, sedangkan 5% akan gagal melakukannya.

Tingkat kepercayaan ini seringkali digambarkan dalam bentuk persentase, seperti 90%, 95%, atau 99%. Semakin tinggi tingkat kepercayaan, semakin lebar interval yang dihasilkan.
Selanjutnya, estimasi interval umumnya memiliki bentuk:
$(\hat{\theta} - E, \hat{\theta} + E)$

Dimana:

1. $\hat{\theta}$ adalah estimasi titik (seperti rata-rata sampel, proporsi sampel, dll.).
2. E adalah **margin of error**, yang menunjukkan seberapa jauh estimasi titik dapat menyimpang dari nilai parameter populasi yang sebenarnya.

Contoh: Jika estimasi titik $\hat{\mu} - 50$ dengan margin of error $E - 2$,

maka interval estimasi untuk rata-rata populasi adalah $[[50 - 2, 50 + 2] = [48, 52]$.

Estimasi interval yang paling umum digunakan adalah untuk rata-rata populasi dan proporsi populasi. Di bawah ini, akan dijelaskan cara menghitung interval estimasi untuk kedua parameter ini.

Estimasi Interval untuk Rata-rata Populasi

Jika kita ingin mengestimasi rata-rata populasi (μ\muμ) berdasarkan sampel, kita dapat menggunakan rumus estimasi interval untuk rata-rata dengan **distribusi normal** atau **distribusi t**, tergantung pada apakah variansi populasi diketahui dan ukuran sampel.

1. Interval Kepercayaan untuk Rata-rata Populasi dengan Variansi Diketahui (Z-Interval)
   Jika variansi populasi (σ^2) diketahui dan ukuran sampel besar (n >30), kita menggunakan distribusi normal untuk menghitung interval kepercayaan:

   $$\overline{x} \pm Z_{\alpha/2}\, x\, \frac{\sigma}{\sqrt{n}}$$

   Di mana:
   - $\overline{x}$ adalah rata-rata sampel.
   - $Z_{\alpha/2}$ adalah nilai kritikal dari distribusi normal yang sesuai dengan tingkat kepercayaan (α) yang diinginkan.
   - σ adalah deviasi standar populasi.
   - n adalah ukuran sampel.

   Contoh: Misalkan rata-rata sampel adalah $\overline{x} = 50$, deviasi standar populasi σ=10, dan ukuran sampel n = 100. Untuk tingkat kepercayaan 95% ($Z_{\alpha/2} = 1.96$), interval kepercayaan untuk rata-rata populasi adalah:

   $$50 \pm 1.96\, x\, \frac{10}{\sqrt{100}} = 50 \pm 1.96\, x\, 1 = 50 \pm 1.96$$

   Interval estimasi adalah [48.04, 51.96].

2. Interval Kepercayaan untuk Rata-rata Populasi dengan Variansi Tidak Diketahui (T-Interval)
   Jika variansi populasi tidak diketahui dan ukuran sampel kecil (n≤30n \leq 30n≤30), kita menggunakan distribusi t untuk menghitung interval kepercayaan:

   $$\overline{x} \pm t_{\alpha/2,n-1}\, x\, \frac{s}{\sqrt{n}}$$

   Di mana:
   - $\overline{x}$ adalah rata-rata sampel.

- $t_{\alpha/2,n-1}$ adalah nilai kritikal dari distribusi t dengan $n-1$ derajat kebebasan.
- s adalah deviasi standar sampel.
- n adalah ukuran sampel.

Contoh: Misalkan rata-rata sampel adalah $\bar{x} = 50$, deviasi standar sampel $s = 10$, dan ukuran sampel $n = 25$. Untuk tingkat kepercayaan 95%, kita mencari nilai t dari tabel t dengan derajat kebebasan $df = n - 1 = 24$, yang menghasilkan $t_{\alpha/2,24} = 2.064$. Maka interval kepercayaan untuk rata-rata populasi adalah:

$$50 \pm 2.064 \; x \; \frac{10}{\sqrt{25}} - 50 \pm 2.064 \; x \; 2 - 50 \pm 4.128$$

Interval estimasi adalah [45.872, 54.128].

Estimasi interval untuk proporsi populasi dapat dihitung dengan rumus berikut:

$$\hat{p} \pm Z_{\alpha/2} \; x \; \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Di mana:

- $\hat{p}$ adalah proporsi sampel.
- $Z_{\alpha/2}$ adalah nilai kritikal dari distribusi normal yang sesuai dengan tingkat kepercayaan yang diinginkan.
- n adalah ukuran sampel.

Contoh: Misalkan dalam sebuah survei, 60 dari 100 orang yang disurvei memilih produk A. Maka proporsi sampel

adalah $\hat{p} = \frac{60}{100} = 0.60$. Untuk tingkat kepercayaan 95%, nilai $Z_{\alpha/2} = 1.96$. Maka interval kepercayaan untuk proporsi populasi adalah:

$$0.60 \pm 1.96\ x\ \sqrt{\frac{0.60(1-0.60)}{100}} - 0.60 \pm 1.96\ x\ \sqrt{\frac{0.24}{100}}$$
$$= 0.60 \pm 1.96\ x\ 0.049$$

Interval estimasi adalah [0.502, 0.698].

Adapun *Margin of error* (*E*) adalah selisih antara estimasi titik dan batas interval estimasi. Margin of error memberikan gambaran tentang ketepatan estimasi. Semakin kecil margin of error, semakin akurat estimasi tersebut. Margin of error bergantung pada:

1. **Ukuran sampel**: Semakin besar ukuran sampel, semakin kecil margin of error.
2. **Tingkat kepercayaan**: Semakin tinggi tingkat kepercayaan, semakin lebar interval estimasi dan semakin besar margin of error.

Selanjutnya, meskipun estimasi interval memberikan rentang nilai untuk parameter populasi, ada dua jenis kesalahan yang dapat terjadi:

1. **Kesalahan Tipe I**: Menyatakan bahwa parameter berada di luar interval, padahal sebenarnya tidak.
2. **Kesalahan Tipe II**: Menyatakan bahwa parameter berada dalam interval, padahal sebenarnya tidak.

Kelebihan dan Kekurangan Estimasi Interval

Kelebihan:

1. Memberikan rentang nilai yang mencakup ketidakpastian estimasi.

2. Memberikan informasi yang lebih lengkap tentang parameter populasi dibandingkan dengan estimasi titik.
3. Dilengkapi dengan tingkat kepercayaan, yang memungkinkan penilaian tentang akurasi estimasi.

Kekurangan:

1. Interval yang lebih lebar mengindikasikan ketidakpastian yang lebih besar.
2. Membutuhkan asumsi tentang distribusi data (normalitas, distribusi t, dll.), yang kadang tidak dapat dipenuhi oleh data yang ada.

DAFTAR PUSTAKA

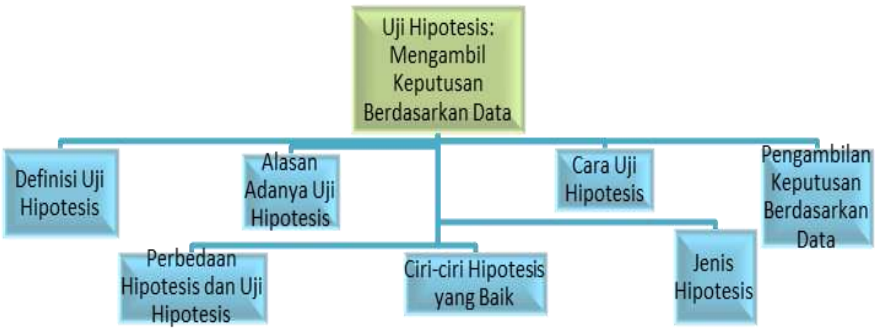
Anwar M, dkk. 2024. *Statistika II.* Batam: Cendikia Mulia Mandiri.

Cahyono. 2018. *Pengaruh Media Sosial terhadap Perubahan Sosial Masyarakat Indonesia.* Bandung: PT Angkasa Pura

BAB 7
UJI HIPOTESIS: MENGAMBIL KEPUTUSAN BERDASARKAN DATA

Oleh Mia Audina Musyadad

7.1 Peta Konsep Uji Hipotesis: Mengambil Keputusan Berdasarkan Data



Gambar 7.1. Peta Konsep Uji Hipotesis: Mengambil Keputusan Berdasarkan Data

7.2 Definisi Uji Hipotesis

Penelitian dilakukan karena ada tujuan tertentu untuk pengambilan sebuah keputusan. Di dalam sebuah penelitian ada rumusan masalah yang ditentukan kebenarannya. Namun sebelum adanya pengambilan keputusan, harus adanya sebuah hipotesis. Menurut (Sugiyono, 2016), Hipotesis digunakan sebagai jawaban sementara terhadap rumusan masalah penelitian. Sedangkan menurut (Nurdin, 2019) hipotesis merupakan satu kesimpulan sementara yang belum akhir; yang merupakan konstruk peneliti terhadap masalah penelitian, serta mendeskripsikan hubungan antara dua atau lebih variabel. Sehingga hipotesis adalah dugaan sementara berdasarkan teori dan penelitian sebelumnya di dalam penelitian yang bertujuan

untuk menjawab rumusan masalah. Pada sebuah hipotesis harus melakukan pengujian.

Menurut (Arifin, 2017), Uji hipotesis dilakukan untuk menyari kebanaran dari prasangka awal atau dugaan sementara. Sehingga definisi uji hipotesis adalah pengujian yang dilakukan untuk menyatakan suatu kebenaran dari praduga sebelumnya. Uji hipotesis adalah teknik statistik untuk menguji kebenaran hipotesis (dugaan sementara) sebuah sampel yang berdasar dari populasi. Proses pembuktiannya berdasarkan variabel yang ada dan hipotesis yang diajukan. Hipotesis terdiri dari Ho yang artinya tidak ada perbedaan atau tidak ada pengaruh dan Ha yang artinya terdapat perbedaan atau terdapat pengaruh.

7.3 Perbedaan Hipotesis dan Uji Hipotesis

Hipotesis dan uji hipotesis adalah dua hal yang berkaitan erat, namun keduanya memiliki peran dan fungsi yang berbeda. Hipotesis adalah dugaan sementara yang mengajukan berdasarkan rumusan masalah yang ada dan hubungan dengan variabel yang ada. Fungsi hipotesis adalah sebagai dugaan awal atau asumsi sementara dari sebuah penelitian. Sedangkan uji hipotesis adalah teknik statistik untuk menguji hipotesis yang sudah dibuat yaitu menguji Ho dan Ha berdasarkan data empiris. Fungsi uji hipotesis adalah untuk menguji keabsahan dugaan awala atau asumsi sementara berdasarkan data empiris.

7.4 Jenis Hipotesis

Berikut jenis hipotesis pada penelitian:

1. Hipotesis deskriptif
   Hipotesis deskriptif adalah dugaan awal suatu sampel dari sebuah populasi yang dapat menunjukkan hubungan antar variabel. Ada juga yang mengatakan bahwa hipotesis pencitraan merupakan perkiraan masalah pencitraan yang berkaitan dengan suatu variabel. Misalkan masalah yang berhubungan langsung dengan suatu variabel. Contohnya saja akan dilakukan kajian terhadap kinerja pasar produk beras serta analisis margin dan efektivitas pemasaran. Oleh

karena itu, hipotesis deskriptif dapat dibuat dari proyek penelitian, yaitu margin yang diperoleh pedagang lebih tinggi atau lebih rendah. Pasar yang tercipta berkaitan dengan peluang pemasaran yang menguntungkan atau tidak.

2. Hipotesis Komparatif
   Hipotesis komparatif adalah sebuah pernyataan sementara yang berasal dari rumusan masalah untuk menujukkan sebuah perbadingan. Ada dua jenis hipotesis komparatif, yaitu komparasi independent (tidak berhubungan) dan komparasi dependen (berhubungan).
   Ciri khas dari hipotesis komparatif adalah menyatakan perbandingan antara dua sampel atau lebih. Contoh hipotesis komparatif yang berdasarkan dari sebuah rumusan masalah Pendidikan yaitu Apakah terdapat perbedaan perilaku yang signifikan antara anak berkebutuhan khusus di Sekolah Luar Biasa Negeri (SLBN) dengan Sekolah Luar Biasa Swasta (SLBS)?
   Berdasarkan rumusan masalah tersebut, maka kita dapat membuat sebuah hipotesis Ho dan Ha, yaitu:
   Ho: Tidak terdapat perbedaan perilaku yang signifikan antara anak berkebutuhan khusus di Sekolah Luar Biasa Negeri (SLBN) dengan Sekolah Luar Biasa Swasta (SLBS)
   Ha : Terdapat perbedaan perilaku yang signifikan antara anak berkebutuhan khusus di Sekolah Luar Biasa Negeri (SLBN) dengan Sekolah Luar Biasa Swasta (SLBS)
   Kesimpulannya jika sebuah penelitian membandingkan beberapa hal yang berbeda, maka hipotesisnya akan mengarah ke hipotesis komparatid, yang artinya kana membandingan perbedaan hal tersebut.

3. Hipotesis Asosiatif
   Hipotesis asosiatif adalah hipotesis yang memiliki kesamaan data. Misalkan sama-sama data ordinal, data interval, data rasio ataupun sama-sama data nominal. Hipotesis asosiatif mengkhususkan pada kedua variabel sama-sama berubah.

Selanjutnya hipotesis asosiatif dibagi menjadi tiga jenis, yaitu:

a. Hipotesis hubungan simetris, yaitu hubungan yang mengerucutkan hubungan kebersamaan antara variabel, bukan hubungan sebab akibat.
   Contoh: Terdapat hubungan antara kemandirian belajar siswa dengan model pembelajaran yang digunakan.
b. Hipotesis hubungan sebab akibat, hubungan yang saling mempengaruhi sebab akibat antara dua variabel atau lebih.
   Contoh: pekerjaan yang sesuai job berpengaruh positif dengan kebahagiaan pegawai, keluarga yang harmonis berpengaruh positif terhdapa tumbuh kembang anak.
c. Hipotesis Interaktif adalah hipotesis yang menyatakan hubungan timbal balik.
   Contoh: terdapat hubungan yang saling mempengaruhi antara penjualan dengan pemasaran, terdapat hubungan yang saling mempengaruhi antara kecerdasan anak dengan kecerdasan ibu, terdapat hubungan yang saling mempengaruhi antara kenaikan bahan pokok dengan ketersediaan barang.

4. Hipotesis Kausal
   Sementara itu, suatu hipotesis yang menyatakan hubungan sebab akibat dari dua variabel atau lebih disebut Hipotesis sementara.

7.5 Ciri-ciri Hipotesis yang Benar

Berikut ciri hipotesis yang benar, di antaranya:

1. Menyatakan hubungan
   Hipotesis yang benar adalah hipotesis yang menyatakan hubungan dari semua variabel yang akan diuji, baik hanya satu variable, dua atau lebih.
2. Sesuai dengan fakta
   Maksud dari harus sesuai fakta adalah hipotesis yang baik disesuaikan dengan fakta di lapangan, jangan sampai

sebuah hipotesis hanya sebuah angan-angan atau gurauan semata.

3. Sesuai dengan ilmu
   Hipotesis harus sesuai dengan ilmu adalah hipotesis berasal dari teori-teori dan jika ada penelitian sebelumnya, harus dapat menggambarkan baik dari teori maupun hasil penelitian sebelumnya.
4. Dapat diuji
   Hipotesis dapat diuji adalah hipotesis yang dibuat harus dapat dilakukan sebuah pengujian sebelum menjawab rumusan masalah.
5. Harus dapat menerangkan fakta
   Maksud hipotesis harus dapat menerangkan fakta adalah ketika kita melakukan penelitian berarti ada fakta yang menyatakan dua permasalahan.
6. Harus dapat menjawab rumusan masalah
   Maksud hipotesis harus dapat menjawab rumusan masalah adalah hipotesis yang baik setelah diuji harus dapat menghasilkan sebuah Keputusan yang mana Keputusan ini untuk menjawab rumusan masalah yang ada.

7.6 Alasan Adanya Uji Hipotesis

Dalam sebuah penelitian harus adanya uji hipotesis, uji hipotesis ini memiliki tujuan:

1. Membuktikan kebenaran hipotesis
   Hipotesis adalah asumsi awal yang harus dibuktikan kebenarannya. Uji hipotesis dilakukan untuk mengetahui asumsi sementara yang sudah dirumuskan berdasarkan dari rumusan masalah.
2. Memvalidasi temuan
   Uji hipotesis dapat digunakan untuk memvalidasi temuan yang didapatkan dari analisis data.
3. Membuat prediksi
   Hipotesis harus dapat diuji agar dapat membuat prediksi tentang hasil percobaan atau penyelidikan.

4. Memajukan pengetahuan
   Hipotesis dapat membantu ilmuwan untuk memajukan pengetahuan karena dapat diuji secara bebas dari nilai dan pendapat peneliti.

7.7 Cara Uji Hipotesis

Uji hipotesis adalah teknik statistika untuk pengambilan keputusan yang didasarkan pada analisis. Berikut adalah beberapa langkah yang dapat dilakukan untuk melakukan uji hipotesis:

1. Menentukan Ho (hipotesis nol) dan Ha (hipotesis alternatif).
   Hipotesis nol mengungkapkan bahwa tidak ada perbedaan, tidak ada peningkatan atau tidak ada korelasi yang signifikan antara dua variabel, sedangkan hipotesis alternatif menyatakan terdapat hubungan atau perbedaan.
2. Menentukan tingkat signifikansi (α).
   Tingkat signifikansi ada beberapa macam. Tingkat signifikansi yang biasa digunakan dalam penelitian adalah tingkat signifikansi 0,05 (5%) dan 0,01 (1%). Tingkat signifikansi yang sesuai dipilih berdasarkan wilayah studi dan keseimbangan yang diinginkan antara kesalahan tipe I dan tipe II. Tingkat signifikansi adalah probabilitas suatu peristiwa terjadi secara kebetulan. Semakin rendah tingkat kekritisannya, semakin rendah pula risikonya. Tingkat signifikansi mencerminkan seberapa yakin peneliti bahwa hasil yang diperoleh bukanlah hasil yang tidak diharapkan. Misalnya, tingkat signifikansi p=0,05 berarti terdapat kemungkinan 95% bahwa hasil yang diamati adalah hasil dari hubungan yang sebenarnya antara kelompok yang dibandingkan biasanya penelitian ini dilakukan untuk penelitian yang tidak berhubungan langsung dengan nyawa. Selanjutnya jika tingkat signifikansi p=0,01 berarti terdapat kemungkinan 99% bahwa hasil yang diamati adalah hasil dari hubungan yang sebenarnya antara kelompok yang dibandingkan biasanya penelitian ini dilakukan untuk penelitian yang berhubungan langsung dengan nyawa. Diambil nilai signifikansi yang sangat kecil agar meminimalisir tingkat kesalahan dalam sebuah penelitian.

3. Memilih metode statistik.

   Dalam memilih metode statistik ini adalah bagian paling penting. Peneliti harus memahami kriteria-kriteria tertentu untuk memilih metode statistik dalam sebuah penelitian. Misalnya jika variabel penelitian sebanyak dua variabel atau tiga variabel maka metode statistik yang digunakan juga berbeda. Kemudian memperhatikan *normality and homogenity* sebuah data. Jika data berdistribusi normal dan homogen, maka menggunakan metode statitika parametrik. Kemudian jika data tidak memenuhi satu kriteria baik tidak berdistribusi normal, ataupun tidak homogen atau bahkan keduanya, maka memilih metode statistik non-parametrik.
4. Mengumpulkan data.

   Data dalam penelitian dapat diklasifikasikan berdasarkan jenis, sumber dan skala pengukurannya:

   a. Jenis Data penelitian dapat diurutkan menjadi data kualitatif dan data kuantitatif. Data kualitatif terdiri dari data nominal dan ordinal, sedangkan data kuantitatif terdiri dari data spasial dan rasio.
   b. Sumber data. Data penelitian dapat diklasifikasikan menjadi data primer dan data sekunder. Data primer adalah data yang diperoleh langsung dari peneliti, sedangkan data sekunder adalah data yang diperoleh dari sumber yang tersedia.
   c. Skala pengukuran. Data survei dapat diklasifikasikan menjadi data nominal, volume, spasial dan rasio. Data nominal dan ordinal merupakan jenis data kualitatif, sedangkan data spasial dan statistik merupakan jenis data kuantitatif.

   Beberapa contoh data penelitian adalah:

   a. Artikel (teks, kata)
   b. Lembar kerja
   c. Buku catatan laboratorium, buku catatan lapangan, buku harian
   d. Wawancara
   e. Nilai tes sebelum dan sesudah pengajian

5. Melakukan uji statistik

   Uji statistik dilakukan sesuai metode yang dipilih, variabel dan karakteristik data yang digunakan. Dalam melakukan uji statistik deperlukan ketelitian agar hasil sesuai dengan data yang sudah diambil saat penelitian.
6. Menghitung statistik uji.

   Perhitungan statistik uji bisa dilakukan dengan manual dengan rumus-rumus yang sesuai dengan teori yang sudah ada. Ataupun bisa menggunakan bantuan SPSS.
7. Menafsirkan hasil.

   Menafsirkan hasil dalam sebuah penelitian harus disesuaikan dengan perhitungan statistiknya. Contoh misalkan perhitungan menggunakan manual, maka yang dibandingkan adalah nilai hitung dibandingkan dengan nilai tabel. Namun jika menggunakan software SPSS membandingkan antara nilai signifikansi dengan nilai alpha.
8. Membuat kesimpulan.

   Pada uji hipotesis yang terakhir adalah membuat Kesimpulan. Berdasarkan uji yang sudah dilakukan, maka dipertimbangkan antara Ho dengan Ha dengan melihat ketentuan yang ada. Misalkan jika penghitungan manual, yang dibandingkan adalah nilai hitung dibandingkan dengan nilai tabel. Sedangkan jika menggunakan bantuan *Sofware SPSS* membandingkan antara nilai signifikan dengan nilai *alpha* yang telah ditentukan oleh peneliti.

7.8 Jenis Uji Hipotesis

Untuk pertanyaan penelitian yang berbeda, sumber statistik dapat menyediakan jenis uji hipotesis yang berbeda. Berikut jenis uji hipotesis berikut:

1. Uji T

   Uji T digunakan ketika dalam statitika penelitian terdapat dua variabel yang berdistribusi normal dan memiliki data yang homogen. Artinya ketika data berdistribusi normal dan homogen, maka data tersebut disebut data parametrik. Karena hanya menggunakan dua variabel, maka pengujian

menggunakan uji T. Berikut contoh hipotesis dalam penelitian yang menggunakan uji T:

Ho : Tidak terdapat perbedaan yang signifikan antara kinerja karyawan Gen Milenial dengan Gen Z

Ha : Terdapat perbedaan yang signifikan antara kinerja karyawan Gen Milenial dengan Gen Z

Selanjutnya berikut contoh hasil Uji T menggunakan bantuan software SPSS dengan nilai $\alpha = 5\%$:

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5,028	1,065		4,722	0,000
	Kepuasan	0,885	0,057	0,812	15,545	**0,000**

Gambar 7.2. Contoh Hasil T menggunakan SPSS

Berdasarkan Gambar 7.2 nilai Sig (0,00) < α (0,05) dengan demikian, dapat diartikan bahwa Ho ditolak dan Ha diterima. Sehingga kesimpulannya adalah **Terdapat perbedaan yang signifikan antara kinerja karyawan Gen Milenial dengan Gen Z.**

2. Uji Wilcoxon dan Uji Mann-Whitney

   Keduanya tidak kalah dengan uji lainnya yaitu uji Z dan uji T. Uji Mann-Whitney berguna untuk membandingkan bidang antar kelompok independen. Misalnya, dapat mengandalkan uji Wilcoxon untuk data berpasangan, yang keduanya berguna jika asumsi distribusi normal tidak terpenuhi dalam penelitian. Jika hasil uji Mann-Whitney memperlihatkan nilai x kurang dari atau sama dengan tingkat signifikansi, maka hipotesis nol ditolak. Hal ini berarti bahwa perbedaan antara median populasi signifikan

secara statistik. Uji *Mann-Whitney* adalah uji statistik non-parametrik yang digunakan untuk membandingkan dua sampel atau kelompok. Uji ini digunakan ketika data tidak memenuhi asumsi normalitas. Hipotesis nol (H0) dari uji *Mann-Whitney* adalah bahwa kedua populasi itu sama, sedangkan hipotesis alternatif (H1) adalah bahwa kedua populasi tidak sama. Berikut contoh hipotesis dalam penelitian yang menggunakan uji Mann-Whitney:

Ho : Tidak terdapat perbedaan cara berpakaian mahasiswa ekonomi dengan mahasiswa hukum di Kampus X

Ha : Terdapat perbedaan cara berpakaian mahasiswa ekonomi dengan mahasiswa hukum di Kampus X

Selanjutnya berikut contoh hasil Uji Hipotesis Mann-Whitney menggunakan bantuan software SPSS dengan nilai $\alpha = 5\%$:

Test Statistics[a]

	KKK_TINGGI
Mann-Whitney U	14.500
Wilcoxon W	29.500
Z	-2.024
Asymp. Sig. (2-tailed)	.043
Exact Sig. [2*(1-tailed Sig.)]	.042[b]

a. Grouping Variable: KELAS

b. Not corrected for ties.

Gambar 7.3. Contoh Hasil Uji Mann-Whitney menggunakan SPSS

Berdasarkan Gambar 7.3 nilai Sig (0,043) < α (0,05) dengan demikian, dapat disimpulkan bahwa Ho ditolak dan Ha diterima. Sehingga kesimpulannya adalah **Terdapat perbedaan cara berpakaian mahasiswa ekonomi dengan mahasiswa hukum di Kampus X.**

3. ANOVA

ANOVA adalah singkatan dari *Analysis of Variance*. ANOVA digunakan untuk membandingkan rata-rata tiga (atau lebih) kelompok sedangkan ANOVA adalah kombinasi analisis regresi dan ANOVA. Jika hasil uji Anova menunjukkan nilai x kurang dari atau sama dengan tingkat signifikansi, maka hipotesis nol ditolak dan Ha diterima. Berikut contoh hipotesis dalam penelitian yang menggunakan uji Anova:

Ho : Tidak terdapat perbedaan siginfikan hasil penjualan yang menggunakan media sosial antara *Facebook, Intagram dan Tiktok*

Ha : Terdapat perbedaan signifikan hasil penjualan yang menggunakan media sosial antara *Facebook, Intagram dan Tiktok*

Berikut contoh hasil Uji Hipotesis Anova menggunakan bantuan software SPSS dengan nilai $\alpha = 5\%$:

ANOVA

GAIN

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.907	2	.953	18.130	.000
Within Groups	6.258	119	.053		
Total	8.165	121			

Gambar 7.4. Contoh Hasil Uji Anova menggunakan SPSS

Berdasarkan Gambar 7.4 nilai Sig (0,00) < α (0,05) dengan demikian, dapat diartikan bahwa Ho ditolak dan Ha diterima. Sehingga kesimpulannya adalah **Terdapat perbedaan signifikan hasil penjualan yang menggunakan media social antara *Facebook, Intagram dan Tiktok***.

4. Uji Kruskall-Wallis

Uji Kruskal-Wallis adalah uji statistik nonparametrik yang digunakan untuk membandingkan tiga atau lebih kelompok data sampel. Uji ini juga disebut Uji H atau ANOVA satu arah pada peringkat. Jika hasil uji Kruskal-Wallis menunjukkan

nilai x kurang dari atau sama dengan tingkat signifikansi, maka hipotesis nol ditolak dan Ha diterima. Hipotesis nol tersebut adalah bahwa median populasi semuanya sama. Uji Kruskal-Wallis adalah metode nonparametrik untuk menguji apakah sampel berasal dari distribusi yang sama. Uji ini biasanya digunakan untuk menentukan apakah median dari sedikitnya tiga kelompok tidak sama. Uji Kruskal-Wallis memiliki beberapa keunggulan, yaitu:

a. Tidak membuat asumsi apa pun tentang sifat distribusi yang mendasarinya, kecuali kontinuitas
b. Membuat lebih sedikit asumsi tentang data daripada padanan parametriknya

Berikut contoh hipotesis dalam penelitian yang menggunakan uji Kruskall-Wallis:

Ho : Tidak terdapat perbedaan hasil panen tanaman yang berada di dalam pot, yang langsung di tanah dan di polibag.

Ha : Terdapat perbedaan hasil panen tanaman yang berada di dalam pot, yang langsung di tanah dan di polibag

Berikut contoh hasil Uji Hipotesis Kruskall-Wallis menggunakan bantuan software SPSS dengan nilai $\alpha = 5\%$:

Test Statistics[b,c]

			GAIN
Chi-Square			27.312
df			2
Asymp. Sig.			.000
Monte Carlo Sig.	Sig.		.000[a]
	95% Confidence Interval	Lower Bound	.000
		Upper Bound	.000

a. Based on 10000 sampled tables with starting seed 2000000.

b. Kruskal Wallis Test

c. Grouping Variable: 1.2.3

Gambar 7.5. Contoh Hasil Uji Kruskall-Wallis menggunakan SPSS

Berdasarkan Gambar 7.5 nilai Sig (0,00) < α (0,05) dengan demikian, dapat disimpulkan bahwa Ho ditolak dan Ha

diterima. Sehingga kesimpulannya adalah Terdapat perbedaan hasil panen tanaman yang berada di dalam pot, yang langsung di tanah dan di polibag.

7.9 Pengambilan Keputusan Berdasarkan Data

Pengujian hipotesis merupakan alat statistik yang sering digunakan dalam penelitian untuk mengetahui apakah terdapat cukup bukti untuk mendukung atau menolak suatu hipotesis. Hal ini melibatkan pembentukan hipotesis tentang parameter populasi, pengumpulan data, dan analisis data untuk menentukan kemungkinan bahwa hipotesis tersebut benar. Ini adalah bagian penting dari metode ilmiah dan digunakan di berbagai bidang. Proses pengujian hipotesis biasanya melibatkan dua hipotesis: hipotesis nol (H_0) dan hipotesis alternatif (H_a) atau hipotesis satu (H_1). Hipotesis nol merupakan pernyataan tidak adanya perbedaan atau hubungan yang signifikan antara dua variabel, sedangkan hipotesis alternatif menunjukkan adanya hubungan atau perbedaan. Peneliti mengumpulkan data dan melakukan analisis statistik untuk menentukan apakah hipotesis nol dapat ditolak dan mendukung hipotesis alternatif. Uji hipotesis digunakan untuk mengambil keputusan berdasarkan data yang telah diuji. Penting untuk memahami asumsi dan batasan yang mendasari proses tersebut. Memilih uji statistik dan ukuran sampel yang tepat penting untuk memastikan hasilnya akurat dan dapat diandalkan. Hal ini memberikan alat yang ampuh bagi peneliti untuk menguji teori dan membuat keputusan berdasarkan bukti.

DAFTAR PUSTAKA

Arifin. (2017). SPSS 24 untuk Penelitian dan Skripsi. Jakarta: Gramedia.

Nurdin, I. (2019). *Metodologi Penelitian sosial.* Surabaya: Media Sahabat Cendikia.

Sugiyono. (2016). *Metode Penelitian Administrasi Dilengkapi Dengan Metode R&D* (23 ed.). Bandung: Alfabeta.

BAB 8
ANALISIS REGRESI HUBUNGAN ANTARA VARIABEL DALAM MODEL MATEMATIS

Oleh Fitri Anisa Kusumastuti

8.1 Pendahuluan

Analisis regresi merupakan metode statistika yang digunakan untuk mengidentifikasi dan memprediksi hubungan antara variabel independen (prediktor) dan variabel dependen (respons). Dengan pendekatan matematis, analisis regresi membantu menyederhanakan fenomena kompleks menjadi model linear yang dapat digunakan untuk inferensi atau prediksi (Gujarati et al., 2003).

Dalam berbagai disiplin ilmu, seperti ekonomi, pendidikan, dan kesehatan, analisis regresi sering kali menjadi alat utama untuk mengukur pengaruh variabel independen terhadap variabel dependen(Montgomery et al., 2021). Hal ini memungkinkan peneliti untuk memahami dan memprediksi fenomena yang diamati secara kuantitatif.

8.2 Konsep Dasar Analisis Regresi

8.2.1 Definisi dan Jenis-jenis Regresi

1. **Regresi Linear Sederhana**: Analisis yang melibatkan satu variabel independen dan satu variabel dependen, dirumuskan sebagai:
   $Y = \beta_0 + \beta_1 X + \epsilon$
   Y: Variabel dependen
   β_0: Intersep
   β_1: Koefisien regresi
   ϵ: Eror residual
2. **Regresi Linear Berganda**: Melibatkan lebih dari satu variabel independent, dirumuskan sebagai:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \epsilon,$$

8.2.2 Variabel dalam Analisis Regresi

1. **Variabel Independen (X)**: Variabel bebas yang memengaruhi atau digunakan untuk memprediksi variabel lainnya.
2. **Variabel Dependen (Y)**: Variabel yang dipengaruhi oleh variabel independen.

8.2.3 Asumsi Dasar

1. **Linearitas**: Hubungan antara variabel dependen dan independen harus linear.
2. **Independensi Error**: Residual (ϵ) harus independen.
3. **Homoskedastisitas**: Variansi residual adalah konstan.
4. **Normalitas Residual**: Residual harus terdistribusi normal.
5. **Tidak Ada Multikolinearitas**: Tidak ada hubungan kuat antar variabel independen dalam regresi berganda.

8.3 Regresi Sederhana

Garis regresi sederhana adalah garis lurus yang memperlihatkan hubungan antara dua variabel, yaitu X dan Y. Persamaan regresi digunakan untuk mendapatkan garis regresi pada data diagram pencar (scatter diagram).

8.3.1 Persamaan Garis Regresi Sederhana

$Y' = a + bX$

Keterangan:

- X: Nilai dari variabel bebas
- Y': Prediksi Y untuk nilai X tertentu

8.3.2 Contoh Soal

Seorang peneliti ingin mengetahui hubungan antara jam belajar (X) dengan nilai ujian (Y) pada siswa. Data yang diperoleh sebagai berikut:

Jam Belajar (X)	Nilai Ujian (Y)
1	55
2	60
3	70
4	75
5	80

Tentukan persamaan regresi linear sederhana yang menggambarkan hubungan antara jam belajar dan nilai ujian, serta prediksi nilai ujian jika seorang siswa belajar selama 6 jam!

Pembahasan:

Pada regresi linear sederhana, persamaan yang digunakan adalah:

$Y = a + bX$

di mana:

- Y adalah variabel dependen (nilai ujian),
- X adalah variabel independen (jam belajar),
- a adalah intercept (nilai Y saat $X = 0$),
- b adalah koefisien regresi (slope), yang menunjukkan seberapa besar perubahan Y per unit perubahan X.

Langkah pertama adalah menghitung koefisien b dan a.

1. **Menghitung b:**

   Koefisien regresi b dihitung dengan rumus:

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

   di mana n adalah jumlah data (dalam hal ini, 5).

   Dari data:

   - $\sum X = 1 + 2 + 3 + 4 + 5 = 15$
   - $\sum Y = 55 + 60 + 70 + 75 + 80 = 340$,
   - $\sum XY = (1)(55) + (2)(60) + (3)(70) + (4)(75) + (5)(80) = 55 + 120 + 210 + 300 + 400 = 1085$

   ,

- $\sum X^2 = (1^2) + (2^2) + (3^2) + (4^2) + (5^2) = 1 + 4 + 9 + 16 + 25 = 55$

.

Sekarang, hitung b:

$$b = \frac{5(1085) - (15)(340)}{5(55) - (15)^2} = \frac{5425 - 5100}{275 - 225} = \frac{325}{50} = 6$$

2. **Menghitung a:**

   Koefisien a dihitung dengan rumus:

   $$a = \frac{340 - 6.5(15)}{5} = \frac{340 - 97.5}{5} = \frac{242.5}{5} = 48.5$$

3. **Persamaan regresi:**

   $Y = 48.5 + 6.5X$

4. **Prediksi untuk X=6X = 6X=6:**

   $Y = 48.5 + 6.5(6) = 48.5 + 39 = 87.5$

   Jadi, jika seorang siswa belajar selama 6 jam, prediksi nilai ujian adalah **87.5**.

8.4 Regresi Linear Berganda

Apabila terdapat lebih dari dua variabel, yaitu satu variabel tidak bebas (Y) dan beberapa variabel bebas $(X_1, X_2, \dots, X_k)$, hubungan liniernya dinyatakan dalam:

$Y' = b_0 + b_1X_1 + b_2X_2 + \cdots + b_kX_k$

8.4.1 Metode Kuadrat Terkecil

Untuk menghitung $b_0, b_1, b_2, \dots, b_k$, digunakan metode kuadrat terkecil. Persamaan normalnya:

$b_0 n + b_1\Sigma X_1 + b_2\Sigma X_2 + \dots = \Sigma Y$

$b_1\Sigma X_1 + b_1\Sigma X_1^2 + b_2\Sigma X_1X_2 + \cdots = \Sigma X_1Y$

$b_2\Sigma X_2 + b_1\Sigma X_1X_2 + b_2\Sigma X_2^2 + \cdots = \Sigma X_2Y$

8.4.2 Contoh Soal

Seorang peneliti ingin mengetahui pengaruh jumlah jam belajar (X1) dan frekuensi latihan soal (X2) terhadap nilai ujian (Y) pada siswa. Data yang diperoleh adalah sebagai berikut:

Jam Belajar (X1)	Latihan Soal (X2)	Nilai Ujian (Y)
1	2	50
2	3	60
3	4	70
4	5	80
5	6	90

Tentukan persamaan regresi linear berganda yang menggambarkan hubungan antara jam belajar, latihan soal, dan nilai ujian!

Berikut adalah pembahasan lengkap untuk regresi linear berganda dengan cara manual. Untuk regresi linear berganda dengan dua variabel independen (misalnya X_1 dan X_2), persamaan umum regresi adalah:

$Y = a + b_1X_1 + b_2X_2$

Di mana:

- Y adalah variabel dependen (nilai ujian),
- X_1 adalah variabel independen pertama (jam belajar),
- X_2 adalah variabel independen kedua (latihan soal),
- a adalah intercept (konstanta),
- b_1 dan b_2 adalah koefisien regresi yang perlu kita tentukan.

Langkah-langkah untuk menghitung koefisien a, b_1, dan b_2 adalah sebagai berikut:

1. **Menyusun Sistem Persamaan**

   Diberikan data sebagai berikut:

Jam Belajar (X1)	Latihan Soal (X2)	Nilai Ujian (Y)
1	2	50
2	3	60
3	4	70

Jam Belajar (X1)	Latihan Soal (X2)	Nilai Ujian (Y)
4	5	80
5	6	90

Kita ingin menemukan persamaan $Y = a + b_1X_1 + b_2X_2$. Langkah pertama adalah menyusun matriks berdasarkan data ini untuk regresi linear berganda.

2. Menyusun Matriks X dan Y

Buatlah matriks X dan vektor Y sebagai berikut:

- Matriks X (termasuk kolom untuk konstanta):

$$X = \begin{bmatrix} 1 & X_1^1 & X_2^1 \\ 1 & X_1^2 & X_2^2 \\ 1 & X_1^3 & X_2^3 \\ 1 & X_1^4 & X_2^4 \\ 1 & X_1^5 & X_2^5 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 3 \\ 1 & 3 & 4 \\ 1 & 4 & 5 \\ 1 & 5 & 6 \end{bmatrix}$$

- Vektor Y (nilai ujian):

$$Y = \begin{bmatrix} 50 \\ 60 \\ 70 \\ 80 \\ 90 \end{bmatrix}$$

3. Menghitung Matriks Transpose X^T

Sekarang kita akan menghitung matriks transpos dari X (yaitu X^T):

$$X^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 5 & 6 \end{bmatrix}$$

4. Menghitung Matriks $X^T X$

Selanjutnya, kita hitung hasil perkalian $X^T X$:

$$X^T X = \begin{bmatrix} 5 & 15 & 20 \\ 15 & 55 & 70 \\ 20 & 70 & 90 \end{bmatrix}$$

5. **Menghitung Matriks $X^T Y$**
   Kemudian, kita hitung perkalian XTYX^T YXTY:

$$X^T Y = \begin{bmatrix} 350 \\ 1300 \\ 1600 \end{bmatrix}$$

6. **Menyelesaikan Sistem Persamaan Linear**
   Sekarang kita punya sistem persamaan yang berbentuk $(X^T X)\beta = X^T Y$, dengan β adalah vektor koefisien a, b_1, dan b_2. Kita dapat menyelesaikannya dengan cara invers matriks untuk mendapatkan β:
   $\beta = (X^T X)^{-1}(X^T Y)$
   a. Menghitung invers dari matriks $X^T X$:

      $(X^T X)^{-1} = [inv(5,15,20,55,70,90)]$
      Namun, karena perhitungan invers matriks memerlukan banyak langkah aljabar, sering kali kita menggunakan software seperti Excel, SPSS, atau Python untuk menghitung invers dan mendapatkan nilai β.
      Setelah menghitung invers matriks $(X^T X)^{-1}$, kita kalikan dengan matriks $X^T Y$ untuk mendapatkan koefisien-koefisien a, b_1, dan b_2.
      Misalnya, setelah dihitung menggunakan metode numerik atau perangkat lunak, kita mendapatkan hasil sebagai berikut:

$$\beta = \begin{bmatrix} a \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 40 \\ 6 \\ 5 \end{bmatrix}$$

7. **Persamaan Regresi**
   Dengan demikian, persamaan regresi linear berganda yang diperoleh adalah:
   $Y = 40 + 6X_1 + 5X_2$
8. **Interpretasi**
   - **Intercept $a = 40$:**
     Jika seorang siswa tidak belajar (jam belajar = 0) dan tidak melakukan latihan soal (latihan soal = 0), nilai ujian diperkirakan adalah 40.
   - **Koefisien $b_1 = 6$:**
     Setiap tambahan 1 jam belajar (X_1) akan meningkatkan nilai ujian sebanyak 6 poin, dengan asumsi frekuensi latihan soal (X_2) tetap.
   - **Koefisien $b_2 = 5$:**
     Setiap tambahan 1 latihan soal (X_2) akan meningkatkan nilai ujian sebanyak 5 poin, dengan asumsi jam belajar (X_1) tetap.
9. **Prediksi Nilai Ujian**
   Jika seorang siswa belajar selama 3 jam dan melakukan 4 latihan soal, maka prediksi nilai ujian Y dapat dihitung sebagai:
   $Y = 40 + 6(3) + 5(4) = 40 + 18 + 20 = 78$
   Jadi, nilai ujian yang diprediksi adalah **78**.

8.5 Pengujian Hipotesis dalam Regresi

8.5.1 Uji Signifikansi Parameter

Hipotesis untuk parameter β_1 biasanya dirumuskan sebagai:

1. $H_0: \beta_1 = 0$ (tidak ada pengaruh X terhadap Y).
2. $H_1: \beta_1 \neq 0$(ada pengaruh X terhadap Y).

Statistik uji:

$$t = \frac{\widehat{\beta_1}}{SE(\widehat{\beta_1})}$$

Daerah kritis ditentukan berdasarkan tingkat signifikansi (α) dan derajat kebebasan.

8.5.2 Uji Goodness-of-Fit

Koefisien determinasi (R^2) digunakan untuk mengevaluasi seberapa baik model menjelaskan variabilitas Y:

$$R^2 = \frac{\text{SSR}}{\text{SST}}$$

8.6 Studi Kasus

8.6.1 Pengaruh Waktu Belajar terhadap Nilai Ujian

Deskripsi Kasus

Sebuah studi dilakukan untuk mengetahui hubungan antara waktu belajar siswa (dalam jam) dan nilai ujian matematika. Sampel acak sebanyak 10 siswa dianalisis, dengan data sebagai berikut:

Tabel 8.1. Waktu belajar dan nilai ujian siswa

Waktu Belajar (jam)	Nilai Ujian
1	55
2	58
3	60
4	62
5	65
6	68
7	72
8	74
9	78
10	80

Langkah Penyelesaian

1. **Model Matematis**

   Model regresi sederhana:

   $Y = \beta_0 + \beta_1 X + \epsilon$

2. **Perhitungan Koefisien Regresi**
   Menggunakan metode *Least Squares* untuk menghitung koefisien:

   Dari data $\bar{X} = 5.5\bar{X} = 5.5,\ \bar{Y} = 67.2$

   $$\beta_1 = \frac{\sum(XiYi) - n\bar{X}\bar{Y}}{\sum(Xi^2) - nX^2}$$

   Setelah perhitungan diperoleh

   $\beta 1 = 2.5, \qquad \beta 0 = 53.7.$
3. **Model Akhir**
   Model regresi:

   $Y = 53.7 + 2.5X$
4. **Interpretasi**
   Setiap peningkatan 1 jam waktu belajar meningkatkan nilai ujian rata-rata sebesar 2.5 poin.
5. **Goodness-of-Fit**
   Nilai R^2 dihitung untuk mengetahui proporsi variansi nilai ujian yang dijelaskan oleh waktu belajar:

   $R^2 = 0.95$

   R^2 =0.95 menunjukkan bahwa 95% variansi nilai ujian dapat dijelaskan oleh waktu belajar.

8.6.2 Analisis Hubungan Pengeluaran Konsumsi dan Pendapatan Rumah Tangga

Deskripsi Kasus

Suatu survei dilakukan untuk memahami hubungan antara pendapatan bulanan (dalam juta rupiah) dan pengeluaran konsumsi rumah tangga. Sampel data:

Tabel 8.2. Pendapatan dan pengeluaran

Pendapatan (X)	Pengeluaran (Y)
2	3
3	4
4	5
5	6
6	7

Langkah Penyelesaian

1. **Model Matematis**
   Model regresi sederhana:
   $Y = \beta_0 + \beta_1 X + \epsilon$
2. **Perndugaan Parameter**
   Menggunakan metode *Least Squares,* diperoleh:
   $\beta_1 = 1, \beta_0 = 1$
   $Y = 1 + X$
3. **Uji Signifikansi Parameter**
   Dengan t-statistik untuk β_1:
   $t = \frac{\beta_1}{SE(\beta_1)} \approx 5.0 \quad (p < 0.05)$
   Kesimpulan: Parameter signifikan pada tingkat kepercayaan 95%
4. **Goodness-of-Fit**
   $R^2 = 0.95$
   Model menjelaskan 95% variabilitas pengeluaran konsumsi.

8.7 Latihan

Diketahui data dari variabel X dan Y sebagai berikut:

X	66	62	65	64	69	60	72	67	68	68
Y	69	63	67	65	68	64	68	66	70	68

Tugas:

1. Buat diagram pencarnya!
2. Tentukan persamaan garis regresi sederhana!
3. Jelaskan arti dari persamaan tersebut!

8.8 Kesimpulan

Bab ini membahas secara mendalam mengenai analisis regresi, sebuah metode statistik yang digunakan untuk memahami hubungan antara variabel independen dan variabel dependen dalam model matematis. Regresi linear sederhana memungkinkan analisis

hubungan dua variabel, sementara regresi linear berganda memodelkan hubungan antara satu variabel dependen dengan beberapa variabel independen.

Melalui pendugaan parameter dan pengujian hipotesis, analisis regresi memberikan kerangka kerja untuk:

1. **Membuat prediksi**: Model regresi dapat digunakan untuk meramalkan nilai variabel dependen berdasarkan nilai variabel independen.
2. **Memahami hubungan**: Koefisien regresi menunjukkan sejauh mana variabel independen memengaruhi variabel dependen.
3. **Mengukur kekuatan hubungan**: Koefisien korelasi dan koefisien determinasi (R^2) memberikan ukuran statistik untuk menilai kekuatan dan kontribusi hubungan antarvariabel.

Analisis regresi menjadi alat yang esensial dalam berbagai bidang, seperti ekonomi, pendidikan, teknik, dan kesehatan, karena kemampuannya mengintegrasikan data ke dalam model yang dapat diinterpretasikan dan diterapkan secara luas. Dengan memahami asumsi dan keterbatasannya, analisis regresi dapat digunakan secara efektif untuk menjawab pertanyaan penelitian dan mendukung pengambilan keputusan yang berbasis data.

Dengan demikian, regresi tidak hanya memberikan wawasan statistik tetapi juga berfungsi sebagai alat prediksi yang kuat untuk memahami pola dalam data dan hubungan antarvariabel dalam fenomena yang kompleks.

DAFTAR PUSTAKA

Gujarati, D. N., & Porter, D. C. (2009). *Basic Econometrics*. McGraw-Hill Education.

Kutner, M. H., Nachtsheim, C. J., & Neter, J. (2004). *Applied Linear Statistical Models*. McGraw-Hill.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to Linear Regression Analysis*. Wiley.

Walpole, R. E., Myers, R. H., & Myers, S. L. (2012). *Probability and Statistics for Engineers and Scientists*. Pearson.

BAB 9
ANALISIS VARIANSI : MEMAHAMI PERBEDAAN ANTAR KELOMPOK

Oleh Novela Wulandari

9.1 Pendahuluan

Analisis Variansi merupakan salah ilmu statistik yang digunakan untuk menguji hipotesis mengenai apakah ada perbedaan rata-rata atau membandingkan rata-rata dari tiga atau lebih kelompok sampel. Teknik ini membantu untuk menentukan apakah terdapat perbedaan signifikan antara kelompok, melihat secara rata-rata apakah ada perbedaan respons antara beberapa kelompok, akibat adanya berbagai jenis perlakuan. Analisis variansi dalam Bahasa Inggris adalah *analisis of variance* kemudian disingkat dengan ANOVA. Rancangan percobaan analisis varians dua arah dikenal dengan rancangan blok acak lengkap (*Completely Randomized Block Design*), sedangkan analisis varians satu arah dikenal dengan rancangan acak lengkap (*Completely Randomized Design*) (Asra, Rudiansyah ,2017).

Dalam melakukan suatu penelitian biasanya dilakukan pengujian hipotesis perbedaan rata-rata beberapa populasi atau kelompok. Bila hanya ada satu atau dua populasi maka pengujian rata-rata dapat dilakukan dengan uji-Z atau uji-T, seperti yang telah dibahas dalam sebelumnya. Tetapi, untuk perbandingan lebih dari dua rata-rata maka digunakan metode analisis varians sehingga teknik anova ini untuk mengatasi kelemahan uji-T yang mana analisis uji-T tidak bisa mengatasi lebih dari dua kelompok hal ini dikarenakan uji-T akan menyulitkan menentukan taraf signifikan sehingga hasil tidak maksimal dalam keakuratannya. Kelompok sampel perlakuan ini biasa sering digunakan seperti hasil suatu percobaan terkontrol (*controlled/laboratory experiment*) untuk melihat efek dari beberapa perlakuan (*treatment*) seperti beberapa jenis obat, beberapa jenis pupuk, beberapa jenis metode

pembelajaran atau beberapa metode pengukuran lainnya yang juga bisa berdasarkan hasil percobaan lapangan (*field experiment*) (Asra, Rudiansyah ,2017).

Analisis varians dengan statistik uji-F ini untuk mengetahui kehomogenan beberapa nilai tengah yang dijadikan kelompok sampel dalam penelitian. Hipotesis Ho menyatakan bahwa nilai tengah tersebut semuanya sama sehingga dapat diuji dengan menggunakan analisis ragam. Namun jika Ho ditolak maka Hi diterima artinya terdapat perbedaan nilai tengah. Hal ini berarti tidak semua nilai tengah tersebut adalah sama, kemungkinan sekurang-kurangnya terdapat dua nilai Tengah yang berbeda.

9.1.1 Komponen Variansi

Konsep Dasar ANOVA didasarkan pada pengukuran dua jenis variabilitas dalam data menjadi dua komponen utama:
ANOVA membagi total variansi dalam data menjadi dua komponen:

1. Variabilitas Antar Kelompok (*Between-Group Variability*): mengukur sejauh mana rata-rata kelompok berbeda (Variasi yang disebabkan oleh perbedaan perlakuan atau faktor). dihitung dengan membandingkan rata-rata setiap kelompok dengan rata-rata keseluruhan.
2. Variabilitas Dalam Kelompok (*Within-Group Variability*): mengukur variansi data dalam kelompok yang sama. (Variasi yang terjadi di dalam masing-masing kelompok akibat perbedaan individu).

9.1.2 Tujuan Utama

Analisis variansi adalah teknik statistik yang digunakan untuk membandingkan rata-rata dari dua atau lebih kelompok data. Tujuan utama analisis variansi adalah:

1. Mengidentifikasi perbedaan antar kelompok: Apakah rata-rata kelompok berbeda secara signifikan?
2. Mengukur variansi: Mengukur seberapa besar variansi (perbedaan) antara kelompok data.
3. Mengidentifikasi faktor yang mempengaruhi: Menentukan faktor mana yang mempengaruhi perbedaan antara kelompok data.

4. Mengukur variabilitas data: Seberapa besar variasi data di antara dan dalam kelompok?

Tujuan Lain
1. Menguji hipotesis: Menguji hipotesis tentang perbedaan rata-rata antara kelompok data.
2. Menganalisis interaksi: Menganalisis interaksi antara dua atau lebih faktor yang mempengaruhi variabel dependen.
3. Mengidentifikasi pola: Mengidentifikasi pola atau tren dalam data.
4. Membantu pengambilan keputusan: Memberikan informasi yang akurat untuk pengambilan keputusan.

9.1.3 Jenis Analisis Variansi
1. ANOVA satu arah (*One-Way ANOVA*) : Membandingkan rata-rata dari satu faktor independen.
2. ANOVA dua arah (*Two-Way ANOVA*) : Membandingkan rata-rata dari dua faktor independen, termasuk interaksi antar faktor.
3. ANOVA berulang (*Repeated Measures ANOVA*): Digunakan untuk data yang diukur berulang kali pada subjek yang sama.
4. ANOVA kovarian (ANCOVA).

Analisis varians dua arah juga terbagi menjadi dua bagian yaitu:
1. Analisis varians klasifikasi dua arah tanpa interaksi.
2. Analisis varians klasifikasi dua arah dengan interaksi.

9.2 Analisis Varians Satu Arah (*One Way Anova*)

Untuk menentukan jenis analisis variansi satu arah atau dua arah dapat ditentukan dengan melihat jumlah kelompok dan perlakuan yang akan diuji. Analisis varians merupakan analisis yang tepat untuk menguji perbedaan rata-rata dengan banyak kelompok sampel yang pilih secara acak. Hipotesis dalam menguji analisis varians ini menggunakan uji statistik Uji-F.

Secara umum, Anova satu jalur terdiri dari tiga kelompok atau lebih yang dilakukan perlakuan sehingga desain penelitiannya terdiri dari satu baris dan lebih dari 2 kolom. Eksperimen yang

digunakan dalam Anova 1 jalur menggunakan 2 kelas atau lebih yang dijadikan kelas treatment atau perlakuan, dan satu kelas lainnya sebagai kelas kontrol yang dijadikan kelas pembanding.

Misalkan ada k populasi yang bebas menyebar secara normal dengan nilai tengah$\mu_1, \mu_2, \mu_3, \dots, \mu_k$. dan ragam σ_2 sama, maka:

H_O: $\mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k$

H_I:sekurang-kurangnya dua nilai tengah tidak sama

$X_{ij} = \mu^1 + \varepsilon ij = \mu + a_i + \varepsilon ij$, dimana:

εij adalah simpangan pengamatan $ke - j$ dari nilai tengah populasi $ke - i$, dengan ketentuan bahwa:

$$\sum_{i=1}^{k} ai = \sum_{i=1}^{k} (\mu_{ij} - \mu) = 0$$

ai = Pengaruh populasi ke – i

9.2.1 Desain Anova Satu Arah

Tabel 9.1. Desain Uji Anova Dua Arah

Pengamatan	Populasi/Kelompok						
	1	2	…	i	…	k	
1	X_{11}	X_{21}	…	X_{i1}	…	X_{k1}	
2	X_{12}	X_{22}	…	X_{i2}	…	X_{k2}	
⋮	⋮	⋮	⋱	⋮	⋱	⋮	
j	X_{1i}	X_{2i}	…	X_{ij}	…	X_{jk}	
⋮	⋮	⋮	⋱	⋮	⋱	⋮	
n	X_{1n}	X_{2n}	…	X_{in}	…	X_{kn}	
Total	Y_1	Y_2	…	Y_i	…	Y_1	$\boldsymbol{Y}$
Nilai Tengah	$\bar{Y}_1$	$\bar{Y}_2$	…	$\bar{Y}_i$	…	$\bar{Y}_k$	$\bar{\boldsymbol{Y}}$

Keterangan:

X_{ij} = Pengamatan ke-i dari populasi ke-j

Y_i = Total semua pengamatan dalam contoh dari populasi ke-i

$\bar{X}_i$ = Rata-rata semua pengamatan dalam contoh dari populasi ke-i

Y = Total semua nk pengamatan

$\bar{Y}$ = Rata-rata semua nk pengamatan

n = banyak populasi/kelompok

k = banyak sampel pada setiap kelompok

Rumus yang digunakan pada analisis satu arah antara lain:

1. **Jumlah Kuadrat Total (JKT)**

   JKT = $\sum Y_{tot}^2 - \frac{(\sum Y_{tot})^2}{nk}$

2. **Jumlah Kuadrat Untuk Tengah Tiap Kolom (JKK)**

   JKK = $\sum_{i=1}^{k} \frac{(\sum Y_i)^2}{n} - \frac{(\sum Y_{tot})^2}{nk}$

3. **Jumlah Kuadrat Galat (JKG)**

   JKG = $\sum_{i=1}^{k} \left[\sum Y_i^2 - \frac{(\sum Y_i)^2}{n}\right] = \sum y_i^2$

   atau JKG = JKT – JKK

4. **Tabel ANOVA**

Tabel 9.2. Analisis Variansi Satu Arah

Sumber Keragaman	Jumlah Kuadrat	Derajat Bebas (db)	Kuadrat Tengah	F Hitung	F Tabel
Nilai Tengah Kolom	JKK	(k-1)	$S_1^2 = \frac{JKK}{k-1}$	$F = \frac{S_1^2}{S_2^2}$	Signifikansi α =0,05 α = 0,01
Galat	JKG	k(n-1)	$S_2^2 = \frac{JKG}{k(n-1)}$		

Keterangan:

k = banyak yang diuji (populasi/kelompok)

n = banyak pengujian (pengamatan)

menghitung F hitung dan F tabel

$$F\ tabel = \frac{V_{pembilang}}{V_{penyebut}} = \frac{S_1^2}{S_2^2}$$

Keterangan:

$V_{pembilang}$: Derajat Bebas Nilai Tengah Kolom (RKJ_{ant})

$V_{penyebut}$: penyebut Derajat Bebas Galat (RKJ_{dal})

5. Penarikan Kesimpulan

Yaitu dengan membandingkan F hitung dan F tabel
Jika F hitung > F tabel, maka Ho ditolak
Jika F hitung ≤ F tabel, maka Ho diterima
Jika Ho ditolak maka dapat dilakukan uji lanjut yaitu *post hoc*, dimana tujuannya yaitu untuk mengetahui perbedaan antar kelompok yang berbeda secara signifikan. Uji lanjut ini dapat lakukan dengan menggunakan rumus **Uji t-Dunnet** atau **uji lainnya**. Uji ini membandingkan dua kelompok secara bergantian. Banyaknya cara perbandingan kelompok yaitu kombinasi C_2^k (k adalah banyak kelompok)
Perhitungan Uji t-Dunnet

$$t_D = \frac{\bar{Y}_i - \bar{Y}_j}{\sqrt{RJK_{DAL}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}}$$

9.2.2 Langkah-langkah Analisis

1. Pengujian asumsi: Pastikan data memenuhi asumsi normalitas, homogenitas variansi dan independensi.
2. Pembentukan hipotesis: Tentukan hipotesis nol (Ho tidak ada perbedaan) dan hipotesis alternatif (Hi ada perbedaan).
3. Penghitungan statistik: Hitung statistik F dengan rumus F
4. Pengujian signifikansi: Bandingkan nilai F dengan nilai kritis dari distribusi F atau gunakan nilai p untuk menentukan signifikansi.
5. Interpretasi hasil: Jika nilai p < 0,05, tolak Ho dan terima Hi yang berarti ada perbedaan signifikan antara kelompok-kelompok.

9.2.3 Contoh aplikasi perhitungan analisis varians

Pengaruh metode pembelajaran yaitu metode penemuan terbimbing (A_1), metode berbasis proyek (A_2) dan metode pembelajaran berbasis teknologi (A_3) mahasiswa di fakultas X. data berpikir kritis sebagai variabel Y yang disajikan dalam bentuk tabel berikut ini:

X_1	X_2	X_3
8	7	4
8	6	6
6	7	3
7	6	5
9	7	6
6	7	6
8	8	5
7	9	6
9	8	7
7	9	5

Penyelesaian:

1. Hipotesis statistik

   H_o : Tidak terdapat perbedaan rata-rata kemampuan berpikir kritis metode pembelajaran dengan penemuan terbimbing, berbasis proyek,, dan berbasis teknologi

   Hi : Terdapat perbedaan rata-rata kemampuan berpikir kritis metode pembelajaran dengan penemuan terbimbing berbasis proyek, dan berbasis teknologi

2. Menyusun tabel persiapan

Tabel 9.3. Mencari nilai $\sum Y_1$, $\sum Y_2$, $\sum Y_3$, $\sum Y_n$, $\sum Y_1^2$, $\sum Y_2^2$, $\sum Y_3^2$, $\sum Y_n^2$

No.	Y_1	Y_1^2	Y_2	Y_2^2	Y_3	Y_3^2
1	8	64	7	49	4	16
2	8	64	6	36	6	36
3	6	36	7	49	3	9
4	7	49	6	36	5	25
5	9	81	7	49	6	36
6	6	36	7	49	6	36

7	8	64	8	64	5	25
8	7	49	9	81	6	36
9	9	81	8	64	7	49
10	7	49	9	81	5	25
Σ	75	573	74	558	53	293

Tabel 9.4. Menghitung nilai $\sum N_i \sum Y_i, \sum Y_i^2, \sum y_i^2, \bar{Y}_1$

Statistik	X_1	X_2	X_3	Jumlah
N_i	10	10	10	30
$\sum Y_i$	75	74	53	202
$\sum Y_i^2$	573	558	293	1424
$\sum y_i^2$	10.50	10,40	12,10	33
$\bar{Y}_1$	7,50	7,40	5,30	20,20

3. Menentukan Jumlah Kuadrat (JK) beberapa sumber varians
   - Jumlah Kuadrat Total (JKT)

     JKT = $\sum Y_{tot}^2 - \frac{\sum Y_{tot}}{nk}$

     JKT = $1424 - \frac{202^2}{30}$=63,87
   - Jumlah Kuadrat Untuk Tengah Tiap Kolom (JKK)

     JKK = $\sum_{i=1}^{k} \frac{(\sum Y_i)^2}{n} - \frac{\sum Y_{tot}}{nk}$

     JKK = $\frac{75^2}{10} + \frac{74^2}{10} + \frac{53^2}{10} - \frac{202^2}{30} = 30{,}87$
   - Jumlah Kuadrat Galat (JKG)

     JKG = $\sum_{i=1}^{k} \left[\sum Y_i^2 - \frac{(\sum Y_i)^2}{n}\right]$= $\sum y_i^2$ atau

     JKG = JKT-JKK = 33
   - Menentukan harga derajat bebas (db) yakni:

     a. $db_{tot} = 30 - 1 = 29$

     b. $db_{ant} = 3 - 1 = 2$

     c. $db_{dal} = 30 - 3 = 27$
   - Menentukan nilai rata-rata jumlah kuadrat:

$$S_1^2 = \frac{JKK}{k-1} = \frac{30.87}{2} = 15{,}44$$

$$RKJ_{dal}\ atau\ S_2^2 = \frac{JKG}{k(n-1)} = \frac{33}{27} = 1{,}22$$

Sehingga

$F_{hitung} = \frac{S_1^2}{S_2^2} = \frac{15{,}44}{1{,}22} = \mathbf{12{,}656}$

- Tabel ANOVA

Tabel 9.5. Tabel Anova

Sumber Keragaman	**Jumlah Kuadrat**	**Derajat Bebas (db)**	**Kuadrat Tengah**	**F Hitung**	**F Tabel**
Nilai Tengah Kolom	30,87	2	15,44	**12,656**	Signifikan α =0,05= 3,35 α = 0,01=5,49
Galat	33	27	1,22		

4. Kesimpulan
   - Dari hasil F_{hitung} dan F_{tabel} dapat diambil Kesimpulan bahwa **$F_{hitung} > F_{tabel}$** (12,656 >3,35 atau 12,656 > 5,49) sehingga Ho ditolak. Maka Kesimpulan penelitian ini adalah Terdapat perbedaan rata-rata kemampuan berpikir kritis metode pembelajaran dengan penemuan terbimbing berbasis proyek, dan berbasis teknologi
   - Ho ditolak maka dapat dilakukan uji lanjut untuk mengetahui perbedaan antar kelompok dengan menggunakan rumus uji t-Dunnet yaitu dengan membandingkan antara Y_1 dengan Y_2 , Y_1 dengan Y_3, dari Y_2 dengan Y_3, Langkah pertama adalah membuat hipotesis pada masing-masing kelompok yakni:

Dengan menggunakan uji lanjut t-Dunnet, maka:

a. Hipotesis

Ho = Rata-rata Kemampuan Berpikir Kritis kelompok yang menggunakan metode penemuan terbimbing lebih rendah atau sama dengan metode pembelajaran berbasis proyek.

Hi = Rata-rata Kemampuan Berpikir Kritis kelompok yang menggunakan metode penemuan terbimbing lebih rendah atau sama dengan metode pembelajaran berbasis proyek.

Perhitungan Uji t-Dunnet

$$t_D = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{RJK_{DAL}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$t_D = \frac{7,5 - 7,4}{\sqrt{1,22 \times \left(\frac{1}{30} + \frac{1}{30}\right)}} = \frac{0,10}{0,29} = 0,34$$

Kesimpulan:

Dari hasil $T_{D.hit}$ dan $T_{D.tabel}$. Diketahui nilai $T_{D.hit} > T_{D.tabel}$ 13:27:0,05 (0,34 < 1,99), maka Ho diterima. Arti kesimpulannya adalah bahwa tidak ada perbedaan rata-rata kemampuan berpikir kritis kelompok yang menggunakan metode penemuan terbimbing dengan metode pembelajaran berbasis proyek.

b. Hipotesis

Ho = rata-rata Kemampuan Berpikir Kritis kelompok yang menggunakan metode penemuan terbimbing lebih rendah atau sama dengan metode pembelajaran berbasis teknologi.

Hi= rata-rata Kemampuan Berpikir Kritis kelompok yang menggunakan metode penemuan terbimbing lebih tinggi atau sama dengan metode pembelajaran berbasis teknologi.

Perhitungan Uji t-Dunnet

$$t_D = \frac{\bar{Y}_1 - \bar{Y}_3}{\sqrt{RJK_{DAL}\left(\frac{1}{n_1} + \frac{1}{n_3}\right)}}$$

$$t_D = \frac{7,5 - 5,30}{\sqrt{1,22 \times \left(\frac{1}{30} + \frac{1}{30}\right)}} = \frac{2,20}{0,29} = 7,59$$

Kesimpulan:

Dari hasil $T_{D.hit}$ dan $T_{D.tabel}$. Diketahui nilai $T_{D.hit} > T_{D.tabel}$ 13:27:0,05 (7,59 > 1,99),

Maka Ho ditolak. Arti kesimpulannya adalah bahwa kemampuan berpikir kritis kelompok yang menggunakan metode penemuan terbimbing lebih tinggi dari metode pembelajaran berbasis proyek.

c. Hipotesis

Ho= Rata-rata kemampuan berpikir kritis kelompok yang menggunakan metode penemuan terbimbing lebih rendah atau sama dengan metode pembelajaran berbasis proyek.

Hi= Rata-rata kemampuan berpikir kritis kelompok yang menggunakan metode penemuan terbimbing lebih rendah atau sama dengan metode pembelajaran berbasis proyek..

Perhitungan Uji t-Dunnet

$$t_D = \frac{\bar{Y}_2 - \bar{Y}_3}{\sqrt{RJK_{DAL}\left(\frac{1}{n_2} + \frac{1}{n_3}\right)}}$$

$$t_D = \frac{7{,}4 - 5{,}30}{\sqrt{1{,}22 \times \left(\frac{1}{30} + \frac{1}{30}\right)}} = \frac{0{,}10}{0{,}29} = 0{,}44$$

Kesimpulan:

Dari hasil $T_{D.hit}$ dan $T_{D.tabel}$. Diketahui nilai $T_{D.hit} > T_{D.tabel}$ 13:27:0,05 (0,44 > 1,99), maka Ho ditolak. Arti kesimpulannya adalah bahwa kemampuan berpikir kritis kelompok yang menggunakan metode pembelajaran berbasis proyek lebih tinggi daripada metode pembelajaran berbasis teknologi.

9.3 Analisis Varians Dua Arah (*Two Way Anova*)

Pada umumnya dalam hipotesis analisis varians dua arah sama halnya dengan analisis varians satu arah yaitu sama mencari apakah ada perbedaan rata-rata antar kelompok dimana kelompok yang digunakan lebih dari 2 kelompok data. Letak perbedaan penelitian yang menggunakan uji statistik

analisis dua arah dan satu arah adalah terletak pada perlakuan kelompok dengan perlakuannya (yang pengamatan). Anova satu arah hanya memiliki variabel kolom (beberapa kelompok/populasi) sedangkan anova dua arah memiliki variabel kolom dan variabel baris (lebih dari 1 pengamatan). Pada anova satu arah terdapat lebih dari 1 kolom dan hanya 1 baris sedangkan anova dua jalur lebih dari 1 kolom dan lebih dari 1 baris.

Supardi (2014) mengatakan bahwa ada 3 jenis hipotesis penelitian yang perlu diuji yang biasanya digunakan dalam analisis 2 jalur yaitu: (1) Hipotesis efek interaksi (*interaction effect*), efek utama (*main effect*), dan interaksi sederhana (*Simpel interaction*). Linger (2010) menjelaskan bahwa Interaksi sederhana adalah merupakan kerja sama antara kelompok variabel *independen* (bebas) dalam memengaruhi variabel *dependent* (terikat) mencari tahu apakah ada pengaruh atau tidak (Ismail, 2018).

9.3.1 Desain Anova Dua Arah

Anova dua arah dapat digunakan untuk menguji hipotesis yang menyatakan perbedaan rata-rata antara kelompok sampel yang menggunakan desain dua faktor (*two factorial design*) maupun dengan desain bertingkat *treatment by level design* (kadir, 2015).

Sebagai contoh, jenis judul penelitian dengan desain 2 X 2 Pengaruh jenis pupuk terhadap kesuburan tanaman sayuran. Untuk faktornya merupakan jenis pupuk yaitu tanaman jagung dan tanaman tomat dengan perlakuan pemberian pupuk organik dan pupuk urea, sehingga desain penelitian dengan menggunakan desain faktorial sebagaimana gambar di bawah ini:

Tabel 9.6. Desain Penelitian anova dua arah

	Jenis pupuk	Perlakuan	
		Pupuk Organik (A_1)	Pupuk Urea (A_2)
Kontrol	Tanaman Jagung (B_1)	A_1B_1	A_2B_1
	Tanaman Tomat (B_2)	A_1B_2	A_2B_2

Tabel 9.7. Desain Uji Anova Dua Arah

Baris / kolom		Variabel A						Total	Nilai Tengah
		1	2	...	i	...	k		
Variabel B	1	X_{11}	X_{21}	...	X_{i1}	...	X_{k1}	Y_1	$\bar{Y}_1$
	2	X_{12}	X_{22}	...	X_{i2}	...	X_{k2}	Y_2	$\bar{Y}_2$
	⋮	⋮	⋮	⋱	⋮	⋱	⋮	⋮	⋮
	j	X_{1i}	X_{2i}	...	X_{ij}	...	X_{jk}	Y_j	$\bar{Y}_j$
	⋮	⋮	⋮	⋱	⋮	⋱	⋮	⋮	⋮
	b	X_{1b}	X_{2b}	...	X_{ib}	...	X_{kb}	X_b	$\bar{Y}_b$
Total		$Y_{.1}$	$Y_{.2}$	...	$Y_{.i}$	...	$Y_{.k}$	$\boldsymbol{Y}$	
Nilai Tengah		$\bar{Y}_1$	$\bar{Y}_2$	...	$\bar{Y}_i$	...	$\bar{Y}_k$		$\bar{\boldsymbol{Y}}$

Rumus yang digunakan pada analisis satu arah anatara lan:

1. **Jumlah Kuadrat Total (JKT)**

$$JKT = \sum Y_{tot}^2 - \frac{(\sum Y_{tot})^2}{nk}$$

2. **Jumlah Kuadrat Kolom (antar A) dan Jumlah Kuadrat Antar Baris (Antar B)**

$$JK_A = \sum_{i=1}^{k} \frac{(\sum Y_j)^2}{b} - \frac{(\sum Y_{tot})^2}{nk}$$

$$JK_B = \sum_{i=1}^{b} \frac{(\sum Y_i)^2}{k} - \frac{(\sum Y_{tot})^2}{nk}$$

3. **Jumlah Kuadrat Interaksi (AB) (JK_{AB})**

$$JK_{AB} = \sum_{i=1,j=1}^{k,b} \frac{(\sum Y_{ij})^2}{bk} - \frac{(\sum Y_{tot})^2}{bk} - JKK - JKB$$

4. **Jumlah Kuadrat Dalam (JK_{Dal})**

$JK_{Dal} = \sum_{i=1,j=1}^{k} \left[\sum Y_{ij}^2 - \frac{(\sum Y_{ij})^2}{bk} \right] = \sum y_{ij}^2$

Atau $JK_{Dal} = JKT - JKA - JKB$

5. **Menghitung derajat kebebasan**
   - $db_{tot} = n_{tot} - 1$
   - $db_a = n_a - 1$
   - $db_b = n_b - 1$
   - $db_{ab} = (n_b - 1)(n_k - 1)$
   - $db_{dal} = n_{tot} . n_b . n_k$
6. **Menghitung semua rata-rata jumlah kuadrat (RJK)**

$RJK_A = \frac{JKK_A}{db_A}$ $RJK_{AB} = \frac{JK_{ab}}{db_{ab}}$

$RJK_B = \frac{JKK_B}{db_B}$ $RJK_{dal} = \frac{JK_{dal}}{db_{dal}}$

7. **Menghitung Fhit antar Kolom, Antar Baris, dan Interaksi**
   - $F_{h(A)} = \frac{RJK_A}{RJK_{dal}}$
   - $F_{h(B)} = \frac{RJK_B}{RJK_{dal}}$
   - $F_{h(AB)} = \frac{RJK_{AB}}{RJK_{dal}}$

Tabel 9.8. Tabel Penolong Anova Dua Arah

Sumber Varians	JK	db	RJK	F_{hitung}	F_{tabel}
Antar A	JK_A	db_a	RJK_A	$F_{h(A)}$	
Antar B	JK_B	db_b	RJK_B	$F_{h(B)}$	
Interaksi AB	JK_{AB}	db_{ab}	RJK_{AB}	$F_{h(AB)}$	
Dalam	JK_{Dal}	db_{dal}	RJK_{dal}	-	
Total	JKT	db_{tot}	-	-	

8. **Membandingkan nilai F_{hitung} dengan F_{tabel},**

Kriteria pengujian,

Jika $F_{h(A)} > F_{tabel}$, maka Ho ditolak.

Jadi terdapat perbedaan rata-rata antara kelompok A yang diuji,

$F_{h(A)} \leq F_{tabel}$ maka Ho diterima.

Jadi tidak terdapat perbedaan rata-rata antara kelompok A yang diuji,

Jika $F_{h(B)} > F_{tabel}$, maka Ho ditolak.

Jadi terdapat perbedaan rata-rata antara kelompok B yang diuji,

$F_{h(B)} \leq F_{tabel}$ maka Ho diterima.

Jadi terdapat perbedaan rata-rata antara kelompok B yang diuji,

Jika $F_{h(AB)} > F_{tabel}$, maka Ho ditolak.

Jadi tidak terdapat pengaruh secara signifikan antara kelompok A dengan kelompok B

$F_{h(AB)} \leq F_{tabel}$ maka Ho diterima.

Jadi terdapat pengaruh secara signifikan antara kelompok A dengan kelompok B

maka konsekuensinya harus diuji pengaruh sederhana (*simple effect*). *Simple effect* adalah perbedaan rerata Antar A pada tiap kelompok $B\ (i = 1,2,3,..)$ atau perbedaan rerata Antar B pada tiap kelompok $A\ (i = 1,2,3,..)$.

Melakukan uji lanjut (*post hoc*) dengan tujuan untuk mengetahui perbedaan antar kelompok yang berbeda secara signifikan. Pada contoh kali ini dengan meng kan uji Scheffe rumus berikut:

$$Md_{ij} = \sqrt{(k-1)(F_{tab})(RJK_{dal})\left(\frac{1}{b}+\frac{1}{k}\right)}$$

9.3.2 Contoh aplikasi perhitungan analisis varians dua arah

Seorang peneliti ingin meneliti"Pengaruh model pembelajaran *problem based learning* dengan model pembelajaran *discovery learning* terhadap prestasi belajar siswa pada kelompok siswa berjenis kelamin perempuan dan siswa

berjenis kelamin laki-laki. Data disajikan dalam bentuk tabel dibawah ini :

Jenis Kelamin	Perlakuan	
	model pembelajaran *problem based learning* (A_1)	model pembelajaran *discovery learning* (A_2)
Laki-laki (B_1)	6,5 6,0 7,0 6,5 7,0 8,5 6,0 7,0 6,0 7,5 8,0 -	7,0 8,0 8,0 7,5 8,5 8,5 8,0 9,5 8,0 7,5 9,0 7,0
Perempuan (B_2)	3,0 4,0 3,0 4,0 5,0 3,0 4,0 3,5 4,5 3,0 4,0 3,0	7,0 6,0 6,0 7,5 6,0 7,0 6,0 8,0 7,0 6,0 7,0 -

Keterangan:

A_1B_1 = Kelompok siswa berjenis kelamin laki-laki yang diajarkan dengan model pembelajaran *problem based learning*

A_2B_1 = Kelompok siswa berjenis kelamin laki-laki yang diajarkan dengan model pembelajaran *discovery learning*

A_1B_2: = Kelompok siswa berjenis kelamin Perempuan yang diajarkan dengan model pembelajaran *problem based learning*

A_2B_2 = Kelompok siswa berjenis kelamin perempuan yang diajarkan dengan model pembelajaran *discovery learning*

Untuk mencari perbedaan rata-rata dari variabel diatas dapat menggunakan anova dua jalur, prosedurnya sebagai berikut:

1. Hipotesa Anova yaitu:
   H_0 : $\mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k$
   H_i : terdapat minimal 2 kelompok yang berbeda
2. Mencari harga ΣA_1B_1, ΣA_2B_1, ΣA_1B_2, ΣA_2B_2, $\Sigma(A_1B_1)^2$, $\Sigma(A_2B_1)^2$, $\Sigma (A_1B_2)^2$, $\Sigma(A_2B_2)^2$ dengan menggunakan tabel berikut

Tabel 9.9. Menghitung ΣA_1B_1, ΣA_2B_1, ΣA_1B_2, ΣA_2B_2, $\Sigma(A_1B_1)^2$, $\Sigma(A_2B_1)^2$, $\Sigma (A_1B_2)^2$, $\Sigma(A_2B_2)^2$

No	A_1B_1	$(A_1B_1)^2$	A_2B_1	$(A_2B_1)^2$	A_1B_2	$(A_1B_2)^2$	A_2B_2	$(A_2B_2)^2$
1	6,5	42,25	7,0	49	3,0	9	7,0	49
2	6,0	36	8,0	64	4,0	16	6,0	36
3	7,0	49	8,0	64	3,0	9	6,0	36
4	6,5	42,25	7,5	56,25	4,0	16	7,5	56,25
5	7,0	49	8,5	72,25	5,0	25	6,0	36
6	8,5	72,25	8,5	72,25	3,0	9	7,0	49
7	6,0	36	8,0	64	4,0	16	6,0	36
8	7,0	49	9,5	90,25	3,5	12,25	8,0	64
9	6,0	36	8,0	64	4,0	16	7,0	49
10	7,5	56,25	7,5	56,25	3,0	9	6,0	36
11	8,0	64	9,0	81	4,0	16	7,0	49
12	-		7,0	49	3,0	9	-	
Σ	75,5	524,25	96,5	782,25	43,5	162,25	73,5	496,25

3. Kemudian menghitung $\sum N_i \sum \boldsymbol{Y_i}$, $\sum Y_i^2, \sum y_i^2, \bar{Y}_1$

Tabel 9.10. Menghitung nilai $\sum N_i \sum \boldsymbol{Y_i}$, $\sum Y_i^2, \sum y_i^2, \bar{Y}_1$

Statistik	A_1B_1	A_2B_1	A_1B_2	A_2B_2	Jumlah
-N	11	12	12	11	46
$\sum \boldsymbol{Y_i}$	75,5	96,5	43,5	73,5	202289
$\sum Y_i^2$	524,25	782,25	162,25	496,25	14241965
$\sum y_i^2$	6,05	6,25	4,56	5,14	3321,97
$\bar{Y}_1$	6,86	8,04	3,63	6,68	20,2025,21

4. Menghitung harga dari Jumlah Kuadrat masing-masing sumber varians:
   a. JKT = $\sum Y_{tot}^2 - \frac{(\sum Y_{tot})^2}{nk} = 1965 - \frac{(289)^2}{46} = 149{,}33$
   b. $JK_A = \sum_{i=1}^{k} \frac{(\sum Y_j)^2}{b} - \frac{(\sum Y_{tot})^2}{nk}$
   $= \frac{(75{,}5 + 43{,}50)^2}{23} + \frac{(96{,}5 + 73{,}5)^2}{23} - \frac{(289)^2}{46} = 56{,}54$
   c. $JK_B = \sum_{i=1}^{b} \frac{(\sum Y_i)^2}{k} - \frac{(\sum Y_{tot})^2}{nk}$
   $= \frac{(75{,}5+96{,}5)^2}{23} + \frac{(43{,}5+73{,}5)^2}{23} - \frac{(289)^2}{46} = 67{,}76$
   d. $JK_{AB} = \sum_{i=1,j=1}^{k,b} \frac{(\sum Y_{ij})^2}{bk} - \frac{(\sum Y_{tot})^2}{bk} - JKK - JKB$
   $= \frac{(75{,}5)^2}{23} + \frac{(96{,}5)^2}{23} + \frac{(43{,}5)^2}{23} - \frac{(73{,}5)^2}{23} - \frac{(289)^2}{46}$
   $- 56{,}54 - 67{,}76 = 5{,}06$
   e. $JK_{Dal} = \sum y_{ij}^2 = 21{,}97$

5. Menghitung derajat kebebasan
   a. $db_{tot} = n_{tot} - 1 = 46 - 1 = 45$
   b. $db_a = n_a - 1 = 2 - 1 = 1$

c. $db_b = n_b - 1 = 2 - 1 = 1$

d. $db_{ab} = (n_b - 1)(n_k - 1) = (2 - 1)(2 - 1) = 1$

e. $db_{dal} = n_{tot} - (n_b - 1)(n_k - 1) = 46 - 2 \times 2 = 42$

6. Menghitung semua rata-rata jumlah kuadrat (RJK)

$RJK_A = \frac{JKK_A}{db_A} = \frac{56,54}{1} = 56,54$

$RJK_B = \frac{JKK_B}{db_B} = \frac{65,76}{1} = 65,76$

$RJK_{AB} = \frac{JK_{ab}}{db_{ab}} = \frac{5,06}{1} = 5,06$

$RJK_{dal} = \frac{JK_{dal}}{db_{dal}} = \frac{21,97}{45} = 0,52$

7. Menghitung Fhit antar Kolom, Antar Baris, dan Interaksi

a. $F_{h(A)} = \frac{RJK_A}{RJK_{dal}} = \frac{56,54}{0,52} = 108,73$

b. $F_{h(B)} = \frac{RJK_B}{RJK_{dal}} = \frac{65,76}{0,52} = 126,46$

c. $F_{h(AB)} = \frac{RJK_{AB}}{RJK_{dal}} = \frac{5,06}{0,52} = 9,73$

8. Membuat Tabel bantu Anova Dua arah

Tabel 9.11. Tabel Penolong Anova Dua Arah

Sumber Varians	JK	db	RJK	F_{hitung}	F_{tabel}
Antar A	56,54	1	56,54	108,73	4,06
Antar B	67,76	1	65,76	126,46	
Interaksi AB	5,06	1	5,06	9,73	
Dalam	21,97	42	0,52	-	
Total	149,33	45	-	-	

9. Mengambil Kesimpulan dari hasil F_{hitung} dengan F_{tabel}

$H_o: \mu_1 = \mu_2$ b) $H_o: \beta_1 \leq \beta_2$ c) $H_o: A \times B = 0$

H_i: Selain H_o $H_i: \beta_1 > \beta_2$ $H_i: A \times B \neq 0$

a. Pengaruh utama (*main effect*) A :
   Perbandingan $F_{h(A)}$ (108,73) > F_{tabel} (4,06), maka Ho ditolak, kesimpulannya adalah terdapat perbedaan rata-rata siswa dalam belajar menggunakan model pembelajaran *problem based learning* dengan model pembelajaran *discovery learning* terhadap prestasi belajar siswa.
b. Pengaruh tama (*main effect*) B :
   Perbandingan $F_{h(B)}$ (126,46) > F_{tabel} (4,06), maka Ho ditolak, kesimpulannya adalah terdapat perbedaan rata-rata prestasi belajar siswa terhadap jenis kelamin siswa laki-laki dengan Perempuan
c. Pengaruh interaksi (interaction effect) AB :
   Perbandingan $F_{h(AB)}$ (9,73)Z > F_{tabel} (4,06), maka Ho ditolak, kesimpulannya adalah terdapat pengaruh interaksi pada factor model pembelajaran (A) dengan gender (B) terhadap prestasi belajar siswa.

10. Uji Lanjut (*simple effect)*:

Uji perbedaan selanjutnya yang diselesaikan melalui *one way Anova* adalah Uji perbedaan lebih lanjut dengan menggunakan perbandingan *post hoc* seperti yang dilakukan dalam menyelsaikan uji lanjut di Analisis *One Way* ANOVA dalam pembahasan sebelumnya. Prosedur ini memerlukan konversi data ke dalam empat perlakuan atau kelompok yang akan diuji dengan prosedur *One Way* ANOVA (Kadir 2014). Kempat kelompok tersebut adalah (A_1B_1), (A_2B_1), (A_1B_2), (A_2B_2) yang akan dikomparasikan masing-masing secara bergantian berpasangan. Hipotesis yang diuji adalah:

Diperoleh JK_A= 56,54, JK_B=65,76, JK_{AB}= 5,06 maka,

$JK_{AY} = JK_A + JK_B + JK_{AB}$ = 56,54+ 65,76 + 5,06 = 127,36

$db_{AY} = n_{AY} - 1 = 4 - 1 = 3$

$RJK_{AY} = \frac{JK_{AY}}{db_{AY}} = \frac{127,36}{4-1} = 42,45$

$RJK_{DY} = RJK_D$ = 0,52 ; db_d= 42

Maka:

$F_h = \frac{RJK_{AY}}{RJK_{AY}} = \frac{42,45}{0,52} = 81,63$

$F_{(0,05:3:42)} = 2,83$

Membandingkan F_h dengan F_{tabel} Dimana, F_h $(43,80) > F_{tabel}$ $(2,72)$

maka Ho ditolak, artinya terdapat perbedaan rata-rata keempat perlakuan. Sehingga dapat dilakukan selanjutnya yaitu dengan menggunakan uji-t Dunnet untuk membandingan setiap kelompok variable (A_1B_1), (A_2B_1), (A_1B_2), (A_2B_2) secara komparasi bergantian berpasangan.

Uji lanjut dengan menggunakan t-Dunnet ($T_{tabel\ (0,05;42)} = 1,68$)

1. **Uji t-Dunnet *simpel main effect* A perbedaan antara A_1 dan A_2**

$$t_{D(A1 \times A2)} = \frac{\bar{Y}_{A1} - \bar{Y}_{A2}}{\sqrt{RJK_{DAL}\left(\frac{1}{n_{A1}} + \frac{1}{n_{A2}}\right)}} = \frac{(75,5 + 43,5) - (96,5 + 73,5)}{\sqrt{9,73\left(\frac{1}{23} + \frac{1}{23}\right)}} = -55,45$$

Dari hasil $T_{D.hit}$ dan $T_{D.tabel}$.

Diketahui nilai $T_{D.hit}$ $(-55,45) \leq T_{D.tabel(0,05:3:42)}$ $(2,83)$

maka Ho diterima. Kesimpulannya adalah bahwa prestasi belajar kelompok siswa yang menggunakan model pembelajaran *problem based learning* tidak lebih tinggi daripada kelompok yang menggunakan model pembelajaran *discovery learning.*

2. **Uji t-Dunnet *simple main effect* B perbedaan antara B_1 dan B_2**

$$t_{D(B1 \times B2)} = \frac{\bar{Y}_{B1} - \bar{Y}_{B2}}{\sqrt{RJK_{DAL}\left(\frac{1}{n_{B1}} + \frac{1}{n_{B2}}\right)}} = \frac{(75,5 + 96,5) - (43,5 + 73,5)}{\sqrt{9,73\left(\frac{1}{23} + \frac{1}{23}\right)}} = 59,80$$

Dari hasil $T_{D.hit}$ dan $T_{D.tabel}$. Diketahui nilai $T_{D.hit}$ (127,12)> $T_{D.tabel(0,05:3:42)}$ (1,68)
maka Ho ditolak. Kesimpulannya adalah bahwa kemampuan berpikir kritis kelompok yang berjenis kelamin laki-laki lebih tinggi daripada kelompok yang berjenis kelamin Perempuan.

3. Uji t-Dunnet *main effect* A perbedaan antara B_1 dan B_2
Perbedaan antara A_1 dan A_2 untuk B_1 {(A_1B_1) & (A_2B_1)}

$$t_{D(A1B1\times A2B1)} = \frac{\bar{Y}_{11} - \bar{Y}_{21}}{\sqrt{RJK_{DAL}\left(\frac{1}{n_{11}} + \frac{1}{n_{21}}\right)}}$$

$$t_{D(A1B1\times A2B1)} = \frac{6{,}91 - 8{,}04}{\sqrt{9{,}73\left(\frac{1}{11} + \frac{1}{12}\right)}} = -0{,}87$$

perbandingan $T_{D.hit}$ dan $T_{D.tabel}$.
Diketahui nilai $T_{D.hit}$ (-0,87) ≤ $T_{D.tabel(0,05:3:42)}$ (1,68)
maka Ho diterima. Kesimpulannya adalah bahwa bahwa prestasi belajar kelompok siswa yang berjenis kelamin laki-laki lebih tinggi dengan menggunakan metode pembelajaran *problem based learning*

Perbedaan antara A_1 dan A_2 untuk B_2 {(A_1B_2) & (A_2B_2)}

$$t_{D(A1B2\times A2B2)} = \frac{\bar{Y}_{12} - \bar{Y}_{22}}{\sqrt{RJK_{DAL}\left(\frac{1}{n_{12}} + \frac{1}{n_{22}}\right)}}$$

$$t_{D(A1B2\times A2B2)} = \frac{8{,}04 - 6{,}68}{\sqrt{9{,}73\left(\frac{1}{12} + \frac{1}{12}\right)}} = 1{,}07$$

perbandingan $T_{D.hit}$ dan $T_{D.tabel}$.
Diketahui nilai $T_{D.hit}$ (1,07) ≤$T_{D.tabel(0,05:3:42)}$ (1,68)
maka Ho diterima. Kesimpulannya adalah bahwa bahwa prestasi belajar kelompok siswa yang berjenis kelamin perempuan lebih tinggi dengan menggunakan metode pembelajaran problem *based learning*

4. Uji t-Dunnet *main effect* B perbedaan antara A_1 dan A_2
Perbedaan antara B_1 dan B_2 untuk A_1 {(A_1B_1) & (A_1B_2)}

$$t_{D(A1B1\times A1B2)} = \frac{\bar{Y}_{11} - \bar{Y}_{12}}{\sqrt{RJK_{DAL}\left(\frac{1}{n_{11}} + \frac{1}{n_{12}}\right)}}$$

$$t_{D(A1B1\times A1B2)} = \frac{6{,}91 - 8{,}04}{\sqrt{9{,}73\left(\frac{1}{11} + \frac{1}{12}\right)}} = -1{,}10$$

perbandingan $T_{D.hit}$ dan $T_{D.tabel}$.
Diketahui nilai $T_{D.hit}$ (-1,10) ≤ $T_{D.tabel(0,05:3:42)}$ (1,68)
maka Ho diterima. Kesimpulannya adalah bahwa bahwa prestasi belajar kelompok siswa yang berjenis kelamin laki-laki lebih tinggi dengan menggunakan metode pembelajaran *problem based learning*

Perbedaan antara B_1 dan B_2 untuk A_2 {(A_2B_1) & (A_2B_2)}

$$t_{D(A2B1\times A2B2)} = \frac{\bar{Y}_{21} - \bar{Y}_{22}}{\sqrt{RJK_{DAL}\left(\frac{1}{n_{21}} + \frac{1}{n_{22}}\right)}}$$

$$t_{D(A2B1\times A2B2)} = \frac{3{,}67 - 6{,}68}{\sqrt{9{,}73\left(\frac{1}{12} + \frac{1}{11}\right)}} = -2{,}31$$

perbandingan $T_{D.hit}$ dan $T_{D.tabel}$.
Diketahui nilai $T_{D.hit}$ (-2,31) ≤ $T_{D.tabel(0,05:3:42)}$ (1,68)
maka Ho diterima. Kesimpulannya adalah bahwa bahwa prestasi belajar kelompok siswa yang berjenis kelamin laki-laki lebih tinggi dengan menggunakan metode pembelajaran problem *discovery learning*

DAFTAR PUSTAKA

Asra, Abuzar. Rudiansyah. (2017) *Statistika Terapan Untuk Pembuat Kebijakan dan Pengambil Keputusan.* Jakarta: Creswell, Jhon W. (2012). *Educational Research: Planning, Conductingand Evaluating Quantitative and Qualitative Reasearch.* Bosston: Houghton Miffin Company.

Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics*. 5th Edition. SAGE Publications. Montgomery, D. C. (2017). *Design and Analysis of Experiments*. 9th Edition. Wiley.

Ismail, Fajri. (2018). Statistika Untuk Penelitian Pendidikan dan Ilmu-ilmu Sosial. Jakarta : Prenadamedia Group.

Kadir, M. (2015). *Statistik Terapan Edisi Ketiga*. Jakarta: PT Raja Grafindo Persada.

Supriadi. (2013). *Aplikasi Statistika dalam penelitian: Konsep Statistika yang lebih komprehensif*. Jakarta: Change Pusblication.

BAB 10
MODEL LINIER: PREDIKSI DAN INTERPRETASI DALAM KONTEKS DATA

Oleh Tedi Hartoyo dan Unang Atmaja

10.1 Pendahuluan

Model linier adalah teknik statistika dasar yang digunakan untuk menunjukkan hubungan antara satu variabel dependen dan satu atau lebih variabel independen. Secara matematis, model linier menunjukkan hubungan yang merupakan persamaan linier, di mana perubahan pada variabel independen atau variabel prediktor mengakibatkan perubahan berbanding lurus pada variabel dependen. Pada model sederhana, hubungan ini diberikan dalam bentuk persamaan berikut:

$$Y = \beta 0 + \beta 1X + \epsilon$$

Di mana Y adalah variabel dependen yang akan di prediksi, X adalah variabel independen yang digunakan untuk memprediksi, β0 adalah konstanta atau intersep, β1 adalah koefisien regresi yang memperlihatkan dampak X terhadap Y, dan ϵ adalah istilah galat yang mencerminkan kesalahan atau variabilitas yang terjadi dan tidak bisa dijelaskan oleh model.

Model linier memiliki dua varian utama yang biasa digunakan: model linier tunggal yang diterapkan pada satu variabel dan model linier ganda yang digunakan sesuai dengan lebih banyak variabel. Kedua varian ini digunakan untuk memperlihatkan hubungan dan memprediksi apa yang akan terjadi pada skenario-skenario berbeda yang dapat diukur kuantitatif. Pada ekonomi misalnya, ia memperlihatkan gambaran hubungan pada variabel ekonomi yang bisa di ukur kuantitatif seperti pengeluaran, perbelanjaan, atau investasi

pada faktor-faktor yang mempengaruhinya seperti tingkat inflasi, suku bunga, dan tingkt pengganggguran.

Model linier memainkan peran penting pada analisi data, karena hal itu mampu membuat prediksi yang akurat dan interpretasi hubungan yang mudah dimengerti. Salah satu alasan utama mengapa model tersebut digunakan begitu luas adalah bahwa model ini mudah dipahami, dan di saat yang sama, memberikan pandangan makro pada bagaimana variabel dalam data yang disurvei tergantung satu sama lain. Lingkaran linier memungkinkan analisis langsung dan terukur pada variabel dependen dan independen. Dengan menggunakan rasio regresi, contohnya, seseorang dapat menilai seberapa besar perubahan pada variabel independen mempengaruhi perubahan variabel dependen. dalam bisnis, contohnya, perusahan tersebut akan menginginkan mengetahui sebarapa besar pengubah dalam jumlah uang yang ia peruntukan untuk pemasarannya akan berpengaruh dalam penjualannya. model linier akan memberikan ide yang jelas bahwa strategi pemasaran agresif dengan jumlah yang besar akan mempengaruhi jumlah penjualan, dan seberapa banyak. Selama itu, model ini juga digunakan dalam konteks peramalian, yaitu membuat prediksi berapa sejumlah variabel dependen berdasarkan informasi yang tersedia tentang varias pada yang independen. Ini sangat bermanfaat pada banyak bidang, termasuk keuangan, ekonomi, kesehatan, dan ilmu sosial, kapan suatu keputusan dasar berdasarkan data. Bab ini diharapkan menjelaskan apa model linear baik peruntukannya, cara aplikasinya.

10.2 Definisi dan Struktur Model Linier

Model linear merupakan salah satu teknik yang paling mendasar dan umum digunakan dalam statistika untuk memodelkan hubungan antara variabel dependen dan satu atau lebih variabel independen. Dalam model linear, variabel-variabel diasumsikan memiliki hubungan linear, artinya perubahan pada variabel independen X akan mengakibatkan perubahan proporsional pada variabel dependen Y.

10.2.1 Model Linier Sederhana

Model linear sederhana merupakan bentuk model linear yang paling sederhana, di mana satu variabel dependen dihubungkan dengan satu variabel independen. Persamaan matematika untuk model linear sederhana dapat dituliskan sebagai berikut:

$Y= \beta_0 + \beta_1 X+\epsilon$

di mana Y merupakan variabel dependen yang akan diprediksi atau dijelaskan. X merupakan variabel independen yang digunakan untuk memprediksi atau menjelaskan variabel dependen. β0 adalah konstanta atau intersep, yang menunjukkan nilai Y ketika X = 0. β1 adalah koefisien regresi, yang menggambarkan perubahan rata-rata dalam Y¬ untuk setiap perubahan satu unit X. ϵ adalah istilah kesalahan, yang mencerminkan variasi dalam Y yang tidak dapat dijelaskan oleh X.

Persamaan ini menggambarkan hubungan linear antara X dan Y, dengan koefisien regresi β1 menggambarkan seberapa besar perubahan dalam Y yang dapat diharapkan untuk setiap perubahan unit dalam X. Intersep β0 menunjukkan nilai dasar Y ketika X tidak memiliki pengaruh. Misalnya, dalam konteks ekonomi, jika kita tertarik untuk memprediksi pengeluaran rumah tangga pada Y dari pendapatan pada X, maka β1 mewakili seberapa banyak pengeluaran berubah yang pendapatannya naik satu unit, dan β0 akan menjadi pengeluaran rumah tangga di mana pendapatannya nol.

10.2.2 Model Linier Berganda

Model linier berganda adalah bentuk pengembangan dari model linier sederhana, yang melibatkan lebih dari satu variabel independen. Model ini digunakan ketika kita ingin memodelkan hubungan antara satu variabel dependen dengan dua atau lebih variabel independen yang mungkin mempengaruhi variabel dependen tersebut. Persamaan matematis dari model linier berganda adalah:

$Y=\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p Xp + \epsilon$

Di mana:

1. Y tetap merupakan variabel dependen.
2. $X_1,X_2,\ldots,X_p$ adalah variabel independen yang digunakan untuk memprediksi Y.
3. $\beta_1,\beta 2,\ldots,\beta_p$ adalah koefisien regresi yang mengukur seberapa besar kontribusi masing-masing variabel independen $X_1,X_2,\ldots,Xp$ terhadap perubahan pada Y.
4. β_0 adalah konstanta atau intercept, yang menunjukkan nilai Y ketika semua X adalah nol.
5. ϵ adalah error term yang menggambarkan variabilitas Y yang tidak dijelaskan oleh variabel independen.

Model linier berganda memungkinkan analisis yang lebih kompleks dan realistik karena dapat menangani beberapa faktor yang mempengaruhi variabel dependen secara bersamaan. Misalnya, dalam analisis ekonomi, kita mungkin tertarik untuk memprediksi tingkat pengeluaran rumah tangga (Y) berdasarkan berbagai faktor seperti pendapatan (X_1), usia kepala keluarga (X_2), dan tingkat pendidikan (X_3). Model linier berganda akan memberikan kontribusi dari masing-masing faktor ini terhadap tingkat pengeluaran rumah tangga.

10.2.3 Struktur Model Linier

Secara umum, model linier dapat dibagi ke dalam dua komponen, yaitu bagian deterministik dan bagian stokastik.

1. Bagian deterministik: Bagian deterministik dari model linier adalah komponen variabel dependen yang dapat dijelaskan oleh variabel independen. Bagian ini berbentuk $\beta_0+\beta_1X$, dalam model linier sederhana, di mana komponen ini menggambarkan nilai Y yang diprediksi tergantung X. dalam model statistik lain, bagian ini tidak terbatas pada dua variabel, dan adalah $\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_p$
2. Bagian stokastik : bagian stokastik dari model linier adalah error term dari model yang ditampilkan oleh ϵ. Bagian ini menggambarkan ketidakmampuan model untuk menduga nilai Y. Komponen ini mengakomodasi ketidakpastian dan variabilitas yang tidak dapat dijelaskan oleh variabel independen. Asumsi distribusi dari error term sangat penting

dan berpengaruh terhadap validitas model statistik. Biasanya, diasumsikan bahwa error term berdistribusi normal dan memiliki mean-nol dan variasi konstan *homoscedasticity*.

10.3 Estimasi Parameter Model Linier

Pada bagian sebelumnya, seluruh struktur dari model linier telah dibahas baik untuk sederhana maupun model linier berganda. Sekarang, langkah-langkah penting dari estimasi model linier untuk parameter akan dibahas, yang termasuk utama parameter model $\beta_0,\beta_1,\ldots,\beta_p,$ Parameter-parameter ini diperkirakan menggunakan data. Data, dan estimasinya sangat penting, karena memberikan informasi pada hubungan antara variabel independen dan dependen. Model terbaik yang akan diperkirakan akan digunakan dari semua model data linier, yaitu metode kuadrat terkecil.

10.3.1 Metode Kuadrat Terkecil (*Least Squares Method*)

Least squares method adalah teknik yang paling populer dalam estimasi parameter model linier. Tujuan menghitung nilai koefisien adalah minimal dari jumlah kuadrat dari selisih antara nilai yang diprediksi oleh model dan nilai sebenarnya pada data. Dalam hal ini, kita ingin meminimalkan jumlah kuadrat residu atau sum of squared residuals (SSR). Secara matematis, jika n adalah lembar data ,, maka linier dapat ditulis sebagai:

$Y_i = \beta_0 + \beta_1 X_{1\,i} + \beta_2 X_{2i} + \cdots + \beta_p X_{p\,i} + \epsilon_i$

di mana Yi adalah nilai variabel dependen untuk observasi ke-i, dan $X_{1i}, X_{2i}, \ldots, X_{pi}$ adalah nilai independen untuk pengamatan ke-i, dan ϵi adalah error term untuk observasi ke-i. Selanjutnya, residu atau kesalahan prediksi ei dapat dihitung, sebagai selisih antara nilai observasi Yi dan Yi.nilai yang diprediksi:

$ei = Y_i - \widehat{Y}_i$

dimana $\widehat{Y}_i$ adalah nilai yang diprediksi dari model linier. Tujuan kita adalah untuk meminimalkan jumlah kuadrat dari residu, yang didefinisikan sebagai:

$SSR = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Yi - \widehat{Yi})^2$

$\beta 0, \beta 1, \ldots, \beta p$ diperkirakan dengan menghitung nilai-nilai yang akan mengurangi SSR. koefisien yang dihasilkan oleh metode ini disebut estimasi kuadrat terkecil, atau estimasi intercepts.

10.3.2 Estimasi Koefisien Regresi

Untuk model linier sederhana, koefisien regresi dapat dihitung dengan cara berikut:

$\beta_0 = Y - \beta_1 X$

dan nilai dari koefisien regresi adalah sebagai berikut:

$\beta_1 = S^2xy/S^2x$ = Kovarians xy/Varians x

Untuk model linier berganda, koefisien regresi dapat dihitung dengan cara:

$Y = Xb + \varepsilon$

Y adalah vektor $n \times 1$yang berisi nilai variabel dependen; X adalah matriks $n \times (p+1)$ yang berisi nilai variabel independen (dengan kolom pertama diisi dengan 1 untuk representasi konstanta β_0); β adalah vektor parameter regresi yang ingin kita estimasi, yaitu $\beta=(\beta_0,\beta_1,\ldots,\beta_p)^T$ ϵ adalah vektor error term.

Dalam metode kuadrat terkecil, estimasi parameter $\hat{\beta}$ diperoleh dengan rumus:

$\hat{\beta} = (X^TX)^{-1}X^TY.$

Di mana X^T adalah transpos dari matriks X, dan $(X^TX)^{-1}$ adalah invers dari matriks X^TX. Matriks ini dapat dihitung jika matriks X^TX dapat diinversi, yang biasanya memerlukan asumsi bahwa variabel independen tidak mengalami multikolinieritas

Sifat Estimasi Koefisien Regresi .

Metode kuadrat terkecil untuk menghasilkan estimasi koefisien regresi hadir dengan banyak properti penting yang menjadikannya pendekatan yang dapat dipercaya dalam banyak situasi:

1. Unbiased Estimator: Estimasi $\hat{\beta}$ adalah estimator tak bias,

   yang berarti bahwa rata-rata estimasi koefisien regresi yang dihasilkan dari sampel berulang akan sama dengan nilai

parameter sebenarnya. Yaitu E ($\hat{\beta}$) = β yang dalam notasi Matematika.

2. Efisiensi: Estimasi kuadrat terkecil adalah estimator linier blue yang memiliki varians sekecil estimator tak bias lainnya, menurut Teorema Gauss-Markov. Inilah sebabnya mengapa ini terkadang disebut metode Estimator Best Linear Unbiased Estimator (BLUE).
3. Konsistensi — Nilai $\hat{\beta}$ ke nilai parameter sebenarnya saat ukuran sampel n menjadi sangat besar, saat model liniernya benar.

Karakteristik Estimasi Koefisien Regresi

Estimasi koefisien regresi yang dihasilkan oleh metode kuadrat terkecil memiliki beberapa fitur penting:

1. Unbiased estimator: Estimasi $\hat{\beta}$ adalah estimator unbiased, yang berarti bahwa rata-rata estimator koefisien regresi yang diperoleh dari sampel berulang sama dengan nilai parameter yang sebenarnya. Secara matematis dengan notasi, E ($\hat{\beta}$) = β.
2. Efisiensi: Estimasi kuadrat terkecil memiliki varian minimal estimator unbiased lainnya di antara estimator linier. Oleh karena itu, metode kuadrat terkecil sering disebut baik estimator linier tidak bias atau estimator linier terbaik.
3. *Consistency* Estimasi $\hat{\beta}$ mendekati parameter yang sesungguhnya ketika ukuran sampel n semakin besar, dengan asumsi bahwa model linier sesuai dengan data.

Evaluasi Estimasi

Setelah dilakukan, langkah selanjutnya adalah mengevaluasi model tentang sejauh mana model linier yang telah ditentukan menjelaskan data. Salah satu cara untuk mengevaluasi kecocokan model adalah menggunakan Koefisien Determinasi. Koefisien determinasi mengukur seberapa besar proporsi variasi dalam variabel dependen yang dapat dijelaskan oleh model. Simbol koefisien determinasi r^2 (regresi sederhana) dan R^2 (regresi

berganda). Nilai berkisar antara 0 dan 1, dengan nilai yang lebih tinggi menunjukkan bahwa model lebih baik dalam menjelaskan variasi data. Koefisien determinasi dihitung dengan rumus:
Untuk regresi sederhana, sebagai berikut:

$$r^2 = (\text{Kovarian xy})^2/(\text{Varian x})(\text{Varian y})$$

Untuk regresi berganda, sebagai berikut:

$$R^{2.} = 1 - \frac{SSR}{SST}$$

Di mana:

- SSRS adalah jumlah kuadrat residu,
- SST adalah jumlah kuadrat total, yang mengukur total variasi dalam variabel dependen.

Meskipun R^2 dapat memberikan indikasi tentang kecocokan model, penting untuk diingat bahwa R2yang tinggi tidak selalu berarti model yang baik, terutama jika terdapat overfitting atau jika model tidak mematuhi asumsi-asumsi yang diperlukan.
Uji Signifikansi Parameter

Setelah estimasi parameter dilakukan, penting juga untuk melakukan uji signifikansi terhadap koefisien regresi. Uji t digunakan untuk menguji koefisien regresi sederhana apakah koefisien regresi signifikan secara statistic, dan begitu juga untuk regresi berganda apakah masing-masing koefisien regresi ($\beta_1,\beta_2,..,\beta_p$) signifikan secara statistik. Hipotesis nol untuk uji t adalah bahwa koefisien regresi sama dengan nol, yang berarti variabel independen tersebut tidak memiliki pengaruh yang signifikan terhadap variabel dependen. Begitu sebaliknya untuk hipotesis satu. Uji t dapat dilakukan dengan menghitung t-statistic:

$$t = \frac{\beta i}{SE(\hat{\beta}_1)}$$

Di mana $SE(\hat{\beta})$ adalah standard error dari estimasi koefisien regresi ($\hat{\beta}$) Nilai t-statistic dibandingkan dengan distribusi t untuk menentukan p-value, yang digunakan untuk menarik kesimpulan mengenai signifikansi parameter.

10.4 Asumsi-Asumsi dalam Model Linier

Ada beberapa asumsi dasar yang harus dipenuhi untuk model linear yang menghasilkan perkiraan yang valid dan andal. Asumsi ini penting untuk memastikan bahwa hasil dari perkiraan koefisien regresi menggunakan metode kuadrat terkecil tidak bias, efisien, dan konsisten. Jika asumsi ini tidak terpenuhi, maka hasil model mungkin tidak valid dan tidak dapat diandalkan. Secara umum, bisa dikatakan bahwa asumsi yang menjadi dasar dari model linear ini dapat dibagi menjadi beberapa kategori: homogenitas varians, multikolinearitas, dan normalitas residu. Penjelasan yang lebih rinci dari masing-masing asumsi dapat diberikan sebagai berikut.

Homogenitas varians

Asumsi homogenitas varians dapat diterjemahkan sebagai: varians dari residu harus konstan pada semua nilai yang diberikan dari variabel bebas X. Dengan kata lain, distribusi dari respon yang diestimasi tidak boleh tergantung pada nilai dari X. Jika varians dari residu berbeda-beda pada seluruh rentang data, meningkat dengan nilai X yang lebih tinggi, maka ini disebut heteroskedastisitas. Heteroskedastisitas bisa menjadi kelemahan dalam estimasi dari model, karena walaupun koefisien regresi yang dihasilkan oleh model mungkin tetap tidak bias, maka standar erorannya mungkin menjadi bias, dan ini berimbas pada uji signifikansi serta mempengaruhi hasil dari hipotesis yang diujikan di sana, akhirnya menyebabkan kesalahan dalam testing hipotesis dan keputusan yang salah. Cara untuk mendeteksi heteroskedastisitas adalah dengan menggunakan uji Breusch-Pagan atau uji White, yang menguji apakah varians dari residu akan berubah jika nilai dari variabel bebas secara bersamaan berubah. Cara lain menemukannya adalah dengan membentuk grafik dari estimasi terhadap residu dan melihat apakah pola yang dihasilkan adalah konsisten pada seluruh rentang data.

Normalitas Residu

Asumsi residual dari model linear harus terdistribusi secara normal. Normalitas dari residu ini ingin dibuktikan karena alasan testing hipotesis dan membuat prediksi interval. Hal ini memastikan bahwa uji-t akan valid untuk testing signifikansi dari estimasi koefisien, dan uji-F juga akan valid untuk whole model, dan kita akan dapat membuat confidenze interval yang sesuai. Untuk menguji normalitas dari residu, kita dapat menggunakan beberapa metode yang sudah ada, misalnya uji Kolmogorov-Smirnov, atau dengan Shapiro-Wilk. Selain itu kita juga dapat membuat grafik histogram atau grafik QQ (*Quantile-Quantile plot*), dan ini akan memberikan gambaran secara visual apakah distribusi dari estimasi sekitar distribusi normal. Jika ini tidak terjadi, maka model ini mungkin menjadi tidak valid, terutama dalam konteks dari tes-statistik parameter-model. Dalam beberapa kasus, kita juga akan dapat menggunakan transformasi dari variabel-tergantung atau variabel-penjelas sebagai cara untuk menyelesaikan ketidak-normalan.

Tidak Ada Multikolinieritas

Multikolinieritas terjadi ketika dua atau lebih variabel independen dalam model memiliki korelasi yang sangat tinggi. Hal ini nantinya akan mengakibatkan permasalahan dalam estimasi parameter regresi, karena model tersebutlah yang tidak bisa memisahkan pengaruh dari variabel-variabel yang sangat berkorelasi tersebut. Multikolinieritas sendiri dapat menghasilkan koefisien regresi yang tidak stabil, dan sangat terjadi perluasan varians koefisien yang sangat besar.

Untuk suatu kasus dapat mendeteksi multikolinieritas sendiri terdapat cara yang dapat dilakukan, yaitu dengan menggunakan VIF atau Variance Inflation Factor. VIF sendiri mengukur sejauh mana varians koefisien regresi diperbesar oleh adanya multikolinieritas, yang mana satuannya dinyatakan dalam faktor pengali. Sehingga jika mendapatkan kasus dimana VIF lebih besar dari 10, maka model tersebut sudah terdapat masalah multikolinieritas yang signifikan.

10.5 Prediksi Menggunakan Model Linier

Setelah model linier diestimasi dan koefisien-koefisien regresi diperoleh, langkah selanjutnya adalah **prediksi** atau peramalan nilai variabel dependen (Y) untuk observasi baru berdasarkan nilai variabel independen (X) yang diketahui. Prediksi ini sangat penting dalam banyak aplikasi analisis data, seperti peramalan permintaan produk, estimasi pertumbuhan ekonomi, atau prediksi harga pasar. Pada bagian ini, kita akan membahas bagaimana melakukan prediksi menggunakan model linier dan bagaimana menginterpretasi hasil prediksi tersebut.

1. **Interval Prediksi.**

   Meskipun prediksi titik ($\hat{Y}$) memberikan estimasi nilai Y, hasil tersebut tidak menginformasikan tingkat ketidakpastian yang terkait dengan prediksi tersebut. Oleh karena itu, penting untuk menghasilkan **interval prediksi**, yang memberikan rentang nilai yang mungkin untuk variabel dependen dengan tingkat keyakinan tertentu.
   Interval prediksi untuk observasi baru iii dihitung menggunakan informasi mengenai varians residual (σ^2) dan varians dari prediksi. Interval prediksi untuk nilai Y_i pada pengamatan ke-i dapat dihitung dengan rumus:

   $\hat{Y}_i \pm t_{\alpha/2} \cdot SE(\hat{Y}_i)$
   Di mana:
   a. $\hat{Y}_i$ adalah prediksi nilai variabel dependen untuk pengamatan ke-i,
   b. $t_{\alpha/2t}$ adalah nilai kritis dari distribusi t dengan derajat kebebasan n−p−1n - p - 1n−p−1, di mana α adalah tingkat signifikansi yang digunakan (misalnya, 0,05 untuk interval kepercayaan 95%),
   c. SE($\hat{Y}_i$ adalah standar error dari prediksi $\hat{Y}_i$

Standar Error Prediksi:
Standar error dari prediksi $\hat{Y}$I dapat dihitung dengan rumus:

$$SE(\hat{Y}_i) = \sqrt{\hat{\sigma}^2 \left[1 + (X_i^T (X^T X)^{-1} X_i)\right]}$$

Di mana:

- $\hat{\sigma}i^2$ adalah estimasi dari varians residual, yang dihitung dengan rumus:

$$\hat{\sigma}^2 = \frac{SSR}{n-p-1}$$

- X_i adalah vektor nilai variabel independen untuk pengamatan ke-i,
- X^TX adalah matriks kovarians antara variabel independen.

Interval prediksi memberikan rentang nilai Y_i yang mungkin untuk pengamatan ke-i dengan tingkat kepercayaan tertentu. Ini mengakomodasi ketidakpastian dalam model dan variasi residual yang tidak dapat dijelaskan oleh variabel independen.

2. Prediksi dalam Regresi Berganda

Dalam model regresi berganda, proses prediksi tetap sama seperti pada regresi sederhana, namun dengan melibatkan lebih banyak variabel independen. Misalkan kita memiliki model regresi berganda dengan p variabel independen:

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X2 + \cdots + \beta_p X_p + \epsilon$

Untuk prediksi, kita menghitung nilai $\hat{Y}$ dengan memasukkan nilai variabel independen untuk pengamatan baru dalam persamaan regresi. Proses perhitungan prediksi dan interval prediksi untuk model berganda tidak berbeda dengan model sederhana, hanya saja melibatkan lebih banyak parameter dan variabel.

3. Prediksi di Dunia Nyata

Penerapan prediksi model linier sangat luas dan mencakup berbagai bidang, seperti ekonomi, pemasaran, kesehatan, dan ilmu sosial. Di dunia nyata, prediksi sering digunakan untuk merencanakan keputusan bisnis, membuat estimasi dalam penelitian, atau mengantisipasi tren pasar. Namun, penting untuk diingat bahwa prediksi model linier bergantung pada validitas asumsi-asumsi dasar yang telah

dibahas sebelumnya. Jika asumsi-asumsi tersebut tidak dipenuhi, prediksi yang dihasilkan mungkin tidak akurat atau dapat menyesatkan.

10.6 Evaluasi Kesesuaian Model

Setelah model linier diestimasi dan digunakan untuk melakukan prediksi, langkah selanjutnya yang penting adalah mengevaluasi kesesuaian model. Evaluasi ini bertujuan untuk mengukur sejauh mana model tersebut dapat menggambarkan hubungan antara variabel dependen dan variabel independen dengan baik. Evaluasi model dilakukan untuk memastikan bahwa model yang dibangun cukup representatif dan dapat memberikan prediksi yang akurat. Tanpa evaluasi yang tepat, model yang digunakan mungkin tidak sesuai atau bahkan menyesatkan, yang dapat berakibat pada keputusan yang salah dalam praktik.

Dalam konteks regresi linier, ada beberapa pendekatan yang digunakan untuk mengevaluasi kesesuaian model, di antaranya adalah koefisien determinasi dan uji signifikansi model, dan analisis residual. Berikut ini adalah pembahasan mendalam tentang berbagai teknik evaluasi kesesuaian model.

Koefisien Determinasi

Koefisien determinasi dengan simbol r^2 untuk regresi sederhana dan R^2 untuk regresi berganda, atau r-square (r^2) dan R-squared (R^2), adalah ukuran yang paling umum digunakan untuk mengevaluasi seberapa baik model linier sesuai dengan data yang ada. R^2 mengukur proporsi variasi dalam variabel dependen Y yang dapat dijelaskan oleh model regresi berdasarkan variabel independen X. r^2 dihitung sebagai:

r^2 =(Kovarian xy)^2/(var. x)(var. y)
dan
R^2 dihitung sebagai:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

Di mana:

- Yi adalah nilai aktual dari variabel dependen,
- $\hat{Y}_i$ adalah nilai prediksi dari model,
- $\bar{Y}$ adalah rata-rata nilai dari variabel dependen.

Interpretasi r2 atau R^2:

1. R^2 berkisar antara 0 dan 1. Semakin tinggi nilai R^2, semakin baik model dalam menjelaskan variasi data. R^2 = 1 menunjukkan bahwa model dapat menjelaskan semua variasi dalam data, sementara R^2 = 0 menunjukkan bahwa model tidak dapat menjelaskan variasi sama sekali.
2. Meskipun demikian, R^2 yang tinggi tidak selalu berarti model baik, terutama jika model terlalu kompleks dan mungkin mengalami overfitting. Oleh karena itu, penting untuk mempertimbangkan ukuran evaluasi lainnya.

Uji Signifikansi Model

Selain mengukur kekuatan model dengan r^2 atau R^2, penting untuk menguji apakah model regresi secara keseluruhan signifikan atau tidak. Salah satu cara untuk melakukannya adalah dengan menggunakan uji F, yang menguji apakah setidaknya satu dari koefisien regresi $\beta_1,\beta_2,\ldots,\beta_p$ berbeda dari nol, yaitu apakah variabel independen secara keseluruhan berkontribusi signifikan dalam menjelaskan variasi variabel dependen.

Hipotesis dalam uji F adalah:

1. Hipotesis nol : Semua koefisien regresi β1, β2,..., βp sama dengan nol tidak ada pengaruh variabel independen terhadap variabel dependen.
2. Hipotesis alternatif : Setidaknya ada satu koefisien regresi yang tidak sama dengan nol ada pengaruh variabel independen terhadap variabel dependen.

Uji F dihitung menggunakan rumus berikut:

$$F = \frac{(SSR/p)}{(SSE/(n-p-1))}$$

Di mana: SSR adalah jumlah kuadrat regresi, SSE adalah jumlah kuadrat error, p adalah jumlah parameter dalam model termasuk intersep, n adalah jumlah observasi.

Interpretasi Uji F:

1. Jika nilai F yang dihitung lebih besar dari nilai F tabel pada tingkat signifikansi tertentu, maka hipotesis nol ditolak, yang berarti model secara keseluruhan signifikan.
2. Sebaliknya, jika nilai F yang dihitung lebih kecil dari nilai F tabel, maka hipotesis nol diterima, yang berarti model tidak signifikan.

Kesimpulan Evaluasi Kesesuaian Model

Evaluasi model linier membawa berbagai pendekatan di mana teknik pengukuran mencakup:

1. Pengukuran nilai koefisien determinasi R^2 yang mengukur seberapa baik model menjelaskan variasi dalam keseluruhan data,
2. Uji signifikansi model menggunakan uji F untuk memastikan model secara keseluruhan signifikan,

Dengan melakukan evaluasi yang komprehensif, kita dapat memastikan bahwa model regresi linier yang dibangun tidak hanya memberikan prediksi yang akurat, tetapi juga sesuai dengan asumsi yang mendasarinya. Evaluasi ini sangat penting agar model dapat digunakan untuk mengambil keputusan yang tepat dan berbasis bukti.

10.7 Interpretasi Hasil

Setelah model linear telah disesuaikan dan divalidasi menggunakan sejumlah pengujian, langkah selanjutnya biasanya adalah interpretasi hasil. Interpretasi sering kali berfokus pada koefisien regresi, karena ini adalah parameter estimasi model statistik. Kemampuan untuk menginterpretasikan hasil dengan tepat sangat penting karena memungkinkan seseorang untuk memahami hubungan antara variabel dependen dan independen. Interpretasi yang tepat juga dapat memberikan wawasan yang berarti dan memungkinkan kesimpulan yang

tepat untuk ditarik. Menggunakan data sebagai dasar untuk mengambil keputusan biasanya juga memerlukan kemampuan untuk menginterpretasikan data ini dengan tepat, yang mengarah pada keputusan penting.

1. Interpretasi Koefisien Regresi

   Dalam model linear, vektor koefisien regresi berisi kategori efek linear yang dimiliki setiap variabel independen terhadap variabel dependen. Koefisien menggambarkan efek yang dimiliki setiap variabel independen terhadap variabel dependen dan memberikan laju perubahan dalam variabel dependen ketika variabel independen meningkat sebesar satu, asalkan variabel independen lainnya tetap konstan. Secara matematis, model linier dapat dinyatakan sebagai:

   $Y= \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \cdots + \hat{\beta}_p X_p + \epsilon$

   Dimana:

   - β_0 adalah intersep atau konstanta model, yang menunjukkan nilai Y ketika semua variabel independen ($X_1, X_2, \ldots, X_p$) bernilai nol.
   - $\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 + \cdots + \hat{\beta}_p$ adalah koefisien regresi yang menggambarkan pengaruh masing-masing variabel independen terhadap variabel dependen.

   Contoh Interpretasi Koefisien Regresi:

   Misalkan kita memiliki model regresi linier sederhana untuk memprediksi pengeluaran rumah tangga (Y) berdasarkan pendapatan rumah tangga (X):

   $Y= \hat{\beta}_0 + \hat{\beta}_1 X$

   Jika hasil estimasi adalah $\hat{\beta}_0 = 5000$ dan $\hat{\beta}_1$^=0.6, maka

   interpretasinya adalah sebagai berikut:

   - **Intersep $\hat{\beta}_0$ = 5000:** Ketika pendapatan rumah tangga (X) adalah nol, maka pengeluaran rumah tangga diperkirakan sebesar 5000 unit.

- **Koefisien $\hat{\beta}_1$=0.6:** Setiap peningkatan satu unit dalam pendapatan rumah tangga akan meningkatkan pengeluaran rumah tangga sebesar 0.6 unit atau peningkatan sebesar koefisien regresi, dengan asumsi bahwa faktor-faktor lain tetap konstan.

2. Uji Signifikansi Koefisien Regresi

Secara statistik bertujuan untuk memastikan bahwa koefisien regresi yang diperoleh tidak dikarenakan karena kebetulan.untuk menguji hipotesis , koefisien regresi sama dengan nol, yang berarti variabel independennya tidak berpengaruh terhadap variable dependennya.

- H_0 hipotesis nol : Koefisien regresi βi=0 (variabel independen tidak berpengaruh varibel dependen).
- hipotesis alternatif H_1: Koefisien regresi βi≠0 (variabel independen berpengaruh varibel dependen).

Signifikansi diujikan secara statistik bagi masing-masing koefisien regresi dengan nilai dari t statistika adalah:

$$t_i = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)}$$

SE($\hat{\beta}_1$)adalah standar eror dari koefisien estimasi $\hat{\beta}_i$.

Pengujian hipotesis digunakan untuk menentukan apakah nilai t-value, lebih besar atau lebih kecil dari nilai kritis pada tingkat signifikansi yang diinginkan (misalnya, 5%) menggunakan distribusi t. Jika hasil statistik uji t, atau t-value, lebih besar dari t-tabel (table distribusi t), maka kita menolak hipotesis nol dan variabel independen signifikan. Sebaliknya, jika t-hitung lebih kecil nilainya dari t kritis, maka variabel tersebut tidak signifikan.

3. Pengaruh Variabel Independen terhadap Variabel Dependen

Interpretasi koefisien regresi tidak hanya mencakup nilai numerik dari koefisien tersebut, melainkan bagaimana koefisien itu juga memberikan informasi tentang pengaruh relatif dan pentingnya setiap variabel independen terhadap variasi variabel. Secara umum, koefisien regresi yang lebih besar mengindikasikan pengaruh yang lebih besar terhadap variabel dependen; koefisien yang lebih kecil merupakan pengaruh yang lebih kecil.

Sebagai contoh, dalam beberapa model regresi yang menggunakan beberapa variabel independen, misalnya: besar pendapatan, pendidikan, dan usia untuk mengantisipasi pengeluaran rumah tangga, koefisien juga memberikan informasi tentang variabel mana yang paling berpengaruh pada variabel dependen tersebut. Variabel dengan nilai koefisien tertinggi semakin mempengaruhi variabel dependen.

4. Model Linear Non-Linear

Secara umum, adalah penting untuk diingat bahwa meskipun model linier memposisikan hubungan antara variabel dependen dan variabel independen secara sederhana, tetapi tidak setiap hubungan antara suatu variabel bersifat linier. Dalam beberapa kasus, hubungan tersebut akan menunjukkan non-linieritas antara variabel, yang memerlukan tatanan lebih kompleks, seperti menggunakan model non-linier atau transformasi variatel misalnya, logaritma atau kuadrat. Mengikat model liner pada data, bahkan jika residualitas, atau skor residual, persis tidak dapat dicontohkan, jumlah dapat menunjukkan indikasi mode dan alasan lain diungkapkan di bawah ini.

Jika model linier tidak memperkirakan nominal data atau jika skor residual menunjukkan pola rangkaian yang sistematis, ini menandakan kepada kita bahwa kita sangat mungkin memiliki pola non-linier atau pola yang lebih kompleks Sedangkan model diperlukan.

5. **Validitas Eksternal. dan Keterbatasan Model**

Seiring dengan interpretasi hasil model linier, kita juga perlu memperhitungkan validitas eksternal atau sejauh mana model itu bisa digunakan dalam sebuah data yang berbeda atau sesi yang berbeda juga. Hasil dari model yang ditemukan atas sampel data tidak bisa digeneralisasikan pada seluruh populasi yang ada. Oleh karenanya, model perlu diuji berdasarkan data uji yang independens atau oleh penerapan validasi silang juga.

Selanjutnya, model linier perlu juga dibatasi. Model linier sangat berkaitan dengan asumsi-asumsi yang menjadi cakupannya, antara lain linearitas, homoskedasticity, dan normalitas residu. Jika asumsi ini separti tidak tercukupi, hasil akan berakibat tidak akurat atau dapat menyesatkan.

6. **Kesimpulan Interpretasi Hasil**

Interpretasi hasil model linier mencakup:

a. Menerjemahkan makna koefisien fefresi dan diaplikasikan terhadap variabel dependen,
b. Membuat uji dalam arti tingkat signifikansi bila ada koefisien regresi signifikan,
c. Memahami adanya interaksi antar dua variabel apabila ada,
d. Mengevaluasi validitas eksternal dan adanya keterabatasan model.

Dengan interpretasi yang tepat, hasil model linier dapat memberikan wawasan yang berharga untuk pengambilan keputusan berbasis data dan memahami dinamika hubungan antar variabel yang dianalisis.

10.8 Studi Kasus dan Aplikasi

Secara teoritis, algoritma penerapan model linier seharusnya lebih mudah diilustrasikan dengan bantuan contoh-contoh praktis. Contoh-contoh tersebut dapat diambil dari berbagai bidang kegiatan yang dapat digunakan untuk menghitung model, misalnya, pada bidang ekonomi, bisnis,

sains, teknik, dan lain lain. berdasarkan data aktual. Setiap kasus menggambarkan bagaimana model linier membantu dalam memprediksi, menemukan hubungan antara variabel, dan menilai serta menginterpretasikan hasilnya. Urutan tersebut memberikan beberapa kasus tentang bagaimana matriks kasus model linier diimplementasikan dengan tujuan pengambilan keputusan.

Kasus 1: Estimasi pengeluaran rumah tangga karena tingkat pendapatan.

Misalkan kita memiliki informasi tentang pengeluaran rumah tangga yang menjadi sampel rumah tangga dan kita mengetahui pendapatan mereka. Tujuan analisis adalah estimasi pengeluaran rumah tangga karena pendapatan mereka. Menerapkan metode regresi linier sederhana akan membantu dalam membuat estimasi ini. Variabel dependen adalah pengeluaran rumah tangga (Y) dan variabelnya adalah pendapatan keluarga (X). Persamaan regresi akan berbentuk sebagai berikut:

$Y = \hat{\beta}_0 + \hat{\beta}_1 X_1$

Setelah melakukan estimasi koefisien regresi, kita mendapatkan hasil bahwa $\hat{\beta}_0$ =5000 dan $\hat{\beta}_1$ = 0.6. Dengan hasil ini, kita dapat menginterpretasikan bahwa:

1. Intersep $\hat{\beta}_0$ =. 5000 berarti bahwa jika pendapatan rumah tangga adalah nol, maka pengeluaran rumah tangga diperkirakan sebesar 5000 unit.
2. Koefisien $\hat{\beta}_1$ =. 0.6 berarti bahwa setiap peningkatan satu unit dalam pendapatan rumah tangga akan meningkatkan pengeluaran rumah tangga sebesar 0.6 unit.

Evaluasi Model: Terdapat nilai r^2, yang dapat memberikan kita pemikiran untuk mengevakuasi model, sejauh apa pendapatan menjelaskan variasi dalam pengeluaran rumah tangga. Jika r2=0,75, maka 75 persen variasi pengeluaran rumah tangga mengikuti variasi dalam pendapatan. Kemudian, uji-t digunakan untuk menilai signifikansi koefisien regresi. Jika nilai

p untuk β_1 lebih kecil dari 0.05, maka kita dapat sebutkan bahwa pendapatan rumah tangga mempengaruhi pengeluaran.

Kasus 2: Analisis Faktor-Faktor yang Mempengaruhi Harga Properti

Dalam studi kasus ini, kami tertarik pada faktor apa yang mempengaruhi harga properti di kota besar yang populer. Kami ingin menganalisis sejauh mana ukuran properti, lokasi, dan usia properti mempengaruhi harga penjualan. Seperti disebutkan sebelumnya, variabel dependen dalam model ini adalah harga properti, yang berarti Y; variabel penjelas (independen) adalah ukuran properti, yang berupa X_1, lokasi X2, dan usianya X_3. Sebagai hasil dari semua ini, model regresi korelasi kami terlihat sebagai berikut:

$Y= \widehat{\beta}_0 + \widehat{\beta}_1 X_1 + \widehat{\beta}_2 X_2 + \widehat{\beta}_3$

Dengan data yang ada, kita memperoleh hasil estimasi koefisien sebagai berikut:

$\widehat{\beta}_0 = 15000$

$\widehat{\beta}_1 = 500$

$\widehat{\beta}_2 = 2000$

$\widehat{\beta}_3 = -300$

Interpretasi Hasil:

1. Intersep: $\widehat{\beta}_0 = 15000$: Jika ukuran properti, lokasi, dan usia properti sama dengan nol, maka harga properti akan sama dengan 15000 unit.
2. Koefisien $\widehat{\beta}_1 = 500$: Jika satu meter persegi properti disertakan, maka harga 500 unit akan naik.
3. Koefisien $\widehat{\beta}_2 = 2000$: Jika satu kategori tambahan lokasi properti lebih strategis akan meningkatkan harga properti sebesar 2000 unit atau sebaliknya.
4. Koefisien $\widehat{\beta}_3 = -300$: Jika satu tahun lagi properti ditambahkan. maka harga properti akan turun sebesar 300 unit.

Evaluasi model : Untuk mengevaluasi model ini, kita memeriksa koefisien determinasi R^2 yang mengukur seberapa baik model kita menjelaskan variasi harga properti. Contohnya, jika model kita 0,85, itu berarti bahwa 85% variasi harga dapat dijelaskan oleh variabel ukuran properti, lokasi properti, dan usia properti. Uji F dapat dijalankan atau tidak untuk menguji keseluruhan model, dan uji t dapat dijalankan atau tidak untuk memeriksa signifikansi setiap koefisien regresi.

Kasus 3: Pengaruh Iklan terhadap Penjualan Produk

Dalam kasus ini, kami ingin mendapatkan informasi tentang bagaimana pengeluaran iklan memengaruhi kuantum penjualan produk. Dalam kasus ini, variabel yang tersedia adalah variabel dependen yang berarti penjualan produk dan variabel independen, yaitu pengeluaran iklan. Model regresi linier sederhana yang digunakan adalah:

$Y= \hat{\beta}_0 + \hat{\beta}_1 X$

Dari analisis regresi, hasilnya adalah β0^=2000 dan β1^=0,75. Artinya:

1. Intersep $\hat{\beta}_0$ = 2000: Jika pengeluaran iklan nol, maka penjualan produk diperkirakan sebesar 2000 unit.
2. Koefisien $\hat{\beta}_1$ = 0,75: Untuk setiap kenaikan satu unit dalam pengeluaran iklan, penjualan produk meningkat sebesar 0,75 unit.

Evaluasi Model: Dalam model, kita perlu mengevaluasi apakah koefisien regresi signifikan dengan menggunakan uji-t. Jika p-value yang dikaitkan dengan $\hat{\beta}_1$ lebih kecil dari tingkat signifikansi yang telah ditentukan yaitu 0,05, maka pengeluaran untuk iklan berpengaruh signifikan terhadap jumlah penjualan produk. Selain itu, kekuatan hubungan juga dapat dinilai dari nilai koefisien determinasi yang dinotasikan dengan R^2.

Penggunaan Model Linier dalam Bisnis

Model linier sebagian besar digunakan dalam statistik, dan memiliki banyak aplikasi praktis dalam bisnis. Misalnya, pada penetapan harga produk, perusahaan dapat menggunakan model linier untuk membentuk hubungan antara harga dan permintaan dan dengan demikian mengoptimalkan harga agar diperoleh keuntungan paling besar. Berikut ini akan dijelaskan bagaimana model linier dapat dibagi menjadi dua jenis: variabel independen atau dependen. Dimanapun permintaan produk adalah variabel dependen, harga produk, dan variabel lain dari lingkungan eksternal seperti kualitas produk dan tren pasar adalah variabel independen.Juga analisis pasar, rencana anggaran dan evaluasi kinerja finansial menggunakan model linier, misalnya analisis regresi yang digunakan untuk memprediksi hasil ekonomi atau mengetahui apa yang menjadi faktor penggerak utama yang menentukan kinerja keuangan suatu perusahaan.

Tantangan dan Keterbatasan dalam Aplikasi Model Linier

Meskipun model linear dianggap praktis, ada beberapa keadaan dan keterbatasan yang harus diingat. Pertama, ada asumsi linearitas, yang model regresi linier harus mematuhi untuk digunakan secara benar. Sebenarnya, model linear tidak bisa melihat pola alamiah ketika hubungan antara variabel tersebut tidak linear. Oleh karena itu, metode lain, seperti regresi non-linear atau transformasi variabel, mungkin dibutuhkan.

Sebagai gantinya, multikolinearitas menghasilkan koefisien regresi yang tidak stabil, menurunkan akurasi model, dan mencegah pengujian hipotesis. Oleh karena itu, hal pertama yang harus kita lakukan adalah memeriksa apakah variabel eksplanatori satu sama lain teruji, dan jika demikian, Anda harus mulai menghilangkan variabel redudan dengan berbagai cara, beberapa dari mereka adalah pemodelan parsimonius atau PCA.

10.9 Penutup

Dari apa yang telah kita pelajari di bagian terdahulu, hampir pasti kita menyadari pentingnya model linear dalam berbagai aspek kehidupan sehari-hari, terutama ekonomi, bisnis, dan masyarakat. Selain itu, kita i juga menyadari bahwa model linear adalah cara yang baik untuk membuat prediksi dan menemukan hubungan antar variabel. Itu sebabnya direkomendasikan untuk selalu mempertimbangkan apakah model benar-benar sesuai sambil tetap memperhatikan kelemahannya. Pengalaman menggunakan model linear secara praktis akan sangat membantu kita membuat keputusan yang lebih baik berdasarkan data.

Secara keseluruhan, model linear adalah salah satu alat statistik paling mendasar yang harus diketahui dalam analisis data. Dengan memahami ide dasar, asumsi, dan cara mengestimasi serta menafsirkan hasil, kita dapat menggunakan model ini untuk tujuan prediksi dan analisis hubungan antar variabel. Tentu saja, model linear memiliki keterbatasan dan dapat digunakan secara efektif hanya dalam situasi tertentu, tetapi itu dapat memberikan wawasan yang tidak mudah bagi keputusan yang didasari data. Bahkan, Ada Banyak Cara untuk Mendapatkannya, namun Pelajari Model Linear yang Paling Murni dengan Cara Melihatnya Diterapkan dalam Konteks Praktis dan Berbasis Data.

DAFTAR PUSTAKA

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). Introduction to Linear Regression Analysis (5th ed.). Wiley.

Wooldridge, J. M. (2015). Introductory Econometrics: A Modern Approach (5th ed.). Cengage Learning.

Gujarati, D. N., & Porter, D. C. (2009). Basic Econometrics (5th ed.). McGraw-Hill Education.

Belsey, M. (1980). Regression Analysis: A Constructive Critique. Sage Publications.

BAB 11
STATISTIKA NON-PARAMETRIK: ALTERNATIF UNTUK DATA YANG TIDAK TERDISTRIBUSI NORMAL

Oleh Rektor Sianturi

11.1 Pendahuluan

Dalam dunia statistik, asumsi normalitas seringkali menjadi landasan dalam melakukan analisis data. Distribusi normal, dengan kurva berbentuk lonceng yang simetris, telah menjadi standar dalam banyak uji statistik parametrik. Namun, tidak semua data dalam kehidupan nyata mengikuti distribusi normal. Ketika kita berhadapan dengan data yang tidak berdistribusi normal, penggunaan uji statistik parametrik menjadi kurang tepat dan dapat menghasilkan kesimpulan yang bias. (Rahman et al., n.d.)

Di sinilah statistika non-parametrik hadir sebagai solusi alternatif. Statistik non-parametrik merupakan cabang statistika yang tidak bergantung pada asumsi distribusi tertentu, termasuk normalitas. Metode-metode non-parametrik ini sangat berguna dalam menganalisis data yang memiliki karakteristik seperti: (Karmini, 2020)

1. Data tidak normal: Ketika data tidak mengikuti distribusi normal, penggunaan uji parametrik dapat menghasilkan kesimpulan yang tidak valid.
2. Skala pengukuran: Untuk data nominal atau ordinal, di mana data lebih bersifat kategori atau peringkat, uji non-parametrik lebih sesuai.
3. Ukuran sampel kecil: Uji non-parametrik lebih robust terhadap ukuran sampel yang kecil.
4. Outlier: Adanya nilai ekstrim dalam data tidak terlalu memengaruhi hasil uji non-parametrik.

Konsep Dasar Statistika Non-Parametrik

Statistika non-parametrik adalah cabang statistika yang tidak bergantung pada asumsi distribusi tertentu, khususnya normalitas. Metode-metode dalam statistika non-parametrik bekerja dengan data mentah atau peringkat data, tanpa memerlukan perhitungan parameter populasi seperti rata-rata atau standar deviasi.

Adapun Kelebihan Statistika Non-Parametrik:

1. Fleksibel: Dapat digunakan untuk berbagai jenis data dan distribusi.
2. Robust: Tidak sensitif terhadap pelanggaran asumsi normalitas.
3. Mudah diinterpretasikan: Hasil uji seringkali lebih intuitif.

Adapun Kekurangan Statistika Non-Parametrik:

1. Kurang efisien: Dibandingkan dengan uji parametrik, uji non-parametrik umumnya memiliki daya uji yang lebih rendah, terutama untuk sampel besar.
2. Informasi yang hilang: Dengan mengubah data menjadi peringkat, beberapa informasi penting dalam data mungkin hilang.(Mashuri, 2015)

Jenis-Jenis Uji Non-Parametrik

Ada banyak jenis uji non-parametrik yang tersedia, masing-masing dirancang untuk menjawab pertanyaan penelitian yang berbeda. Pada pembahasan pada Bab ini akan dibahas beberapa uji non-parametrik yang umum digunakan antara lain:

1. Uji Fisher
2. Uji Kontras
3. Uji Dunnet
4. Uji Kruskall Wallis

Pemilihan uji non-parametrik yang tepat tergantung pada:

1. Jumlah kelompok yang dibandingkan
2. Apakah sampel berpasangan atau tidak
3. Jenis skala pengukuran data
4. Pertanyaan penelitian

Statistika non-parametrik memiliki banyak aplikasi dalam berbagai bidang, seperti:

1. Ilmu sosial: Membandingkan preferensi antara dua kelompok, menganalisis peringkat kinerja, dll.
2. Kedokteran: Menganalisis efektivitas pengobatan pada berbagai kelompok pasien, membandingkan tingkat keparahan penyakit, dll.
3. Biologi: Menganalisis perbedaan pertumbuhan antara dua kelompok tanaman, membandingkan keanekaragaman spesies, dll.(Setiawan, 2018)

Perbandingan Uji Parametrik dan Non-Parametrik

Untuk memahami lebih baik kapan harus menggunakan uji non-parametrik, mari kita bandingkan dengan uji parametrik:

Tabel 11.1. Uji parametrik

Fitur	**Uji Parametrik**	**Uji Non Parametrik**
Asumsi Distribusi	Membutuhkan asumsi distribusi normal	Tidak membutuhkan asumsi distribusi tertentu
Skala Pengukuran	Interval atau rasio	Nominal, ordinal, interval, atau rasio
Efisiensi	Umumnya lebih efisien jika asumsi terpenuhi	Kurang efisien jika asumsi terpenuhi
Robustness	Sensitif terhadap pelanggaran asumsi	Lebih robust terhadap pelanggaran asumsi

Perangkat Lunak Statistik untuk Analisis Non-Parametrik

Banyak perangkat lunak statistik yang dapat digunakan untuk melakukan analisis non-parametrik, antara lain: (Suyanto dan Prana Ugiana Gio, 2015)

1. SPSS: Salah satu perangkat lunak statistik yang paling populer dan menyediakan berbagai macam uji non-parametrik.

2. R: Bahasa pemrograman yang sangat fleksibel dan memiliki banyak paket untuk analisis statistik, termasuk non-parametrik.
3. Minitab: Perangkat lunak statistik yang mudah digunakan, terutama untuk analisis dasar.
4. Python: Bahasa pemrograman lain yang populer dengan pustaka seperti SciPy dan Statsmodels untuk analisis statistik.

11.2 Pengujian Hipotesis menggunakan Statistika Non Parametrik

11.2.1 Uji Fisher

Untuk menguji apakah ada perbedaan dua perlakuaan yang mungkin dari dua populasi. (Karmini, 2020)

1. Skala data nominal atau ordinal
2. Ukuran sampel kurang dari 20
3. Data disusun dalam tabel kontingensi 2 x 2

Langkah-langkah:

1. Tentukan hipotesis
   H_0 : P(I) = P(II)
   H_a : satu arah atau dua arah
2. Tentukan taraf signifikansi nya
3. Susun data dalam tabel kontingensi 2x2 sebagai berikut:

Tabel 11.2. Tabel kontingensi 2x2

	-	+	Jumlah
Kelompok A	A	B	A+B
Kelompok B	C	D	C+D
Jumlah	A+C	B+D	N

4. Tentukan uji statistiknya

$$P_{hitung} = \frac{(A+B)!(C+D)!(A+C)!(B+D)!}{N!A!B!C!D!}$$

5. Kriteria pengujian

   Terima H_0 jika $P_{hitung} \geq \alpha$ (satu arah) atau $P_{hitung} \geq \frac{\alpha}{2}$ (Dua arah) dan tolak H_0 untuk yang lainnya

6. Kesimpulan

Contoh Soal

Sebuah pertanyaan dilontarkan kepada mahasiswa baru dan lama apakah uang BOP perlu dihapus atau tidak.

No	Baru	Lama
1	S	TS
2	S	S
3	TS	TS
4	S	TS
5	S	S
6	TS	TS
7	S	TS
8	S	

Apakah terdapat perbedaan jawaban yang signifikan?
Penyelesaian:

1. Hipotesis:
   H_0 : Terdapat Perbedaan Jawaban Mahasiswa Baru dan Lama
   H_a : Tidak terdapat Perbedaan Jawaban Mahasiswa Baru dan Lama
2. Taraf Signifikansi $\alpha = 5\%$
3. Uji Statistik

	Setuju	Tidak setuju	Jumlah
Baru	6	2	8
Lama	2	5	7
Jumlah	8	7	15

(A + B)! = 8! = 8.7.6.5.4.3.2.1 = 40.320
(C + D)! = 7! = 7.6.5.4.3.2.1 = 5.040
(A + C)! = 8! = 8.7.6.5.4.3.2.1 = 40.320
(C + D)! = 7! = 7.6.5.4.3.2.1 = 5.040
A! = 6! = 6.5.4.3.2.1 = 720

B! = 2! = 2.1 = 2
C! = 2! = 2.1 = 2
D! = 5! = 5.4.3.2.1 = 120
N! = 15! = 15.14.13.12.11.10.9.8.7.6.5.4.3.2.1 = 1.307.674.368.000

$$P_{hitung} = \frac{(A+B)!(C+D)!(A+C)!(B+D)!}{N!A!B!C!D!}$$

$$P_{hitung} = \frac{(6+2)!(2+5)!(6+2)!(2+5)!}{15!6!2!2!5!}$$

$$P_{hitung} = \frac{(40.320)(5.040)(40.320)(5.040)}{(1.307.674.368.000)(720)(2)(2)(120)}$$

P_{hitung} = 0,091

4. Kriteria Pengujian
   Karena P_{hitung} = 0,091 > P_{tabel} = 0,05 maka H_0 diterima
5. Kesimpulan
   Terdapat Perbedaan Jawaban Mahasiswa Baru dan Lama

11.2.2 Uji Kontras

Untuk menguji apakah terdapat perbedaan rata-rata atau tidak dari setiap perlakuan. Skala pengukuran yang digunakan biasa nya data interval dan rasio. (Nugroho, 2008)
Adapun langkah-langkah nya adalah:

1. Tentukan hipotesis
   H_0 : $\mu_i = \mu_j$
   H_a : $\mu_i \neq \mu_j$ dimana i,j = 1,2,3, ...
2. Tentukan taraf signifikansi nya
3. Tabel uji kontras dengan menggunakan ANAVA

Tabel 11.3. Tabel uji kontras dengan menggunakan ANAVA

Sumber variasi (SV)	Derajat Bebas (Db)	Jumlah Kuadrat (JK)	Kuadrat Tengah (KT)	F_{hitung}	F_{tabel}
Perlakuan 1 2 . .	p-1	JKP	$\frac{JKP}{p-1}$	$\frac{KTP}{KTG}$	F_{α} (V_1 vs V_2)
Galat	n(p-1)	JKG	$\frac{JKG}{p(n-1)}$		
Total	N	JKT			

Dimana:
p = banyak perlakuan
n = banyak ulangan
N = jumlah total
V_1 = db perlakuan
V_2 = db galat

JKT = $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} y^2_{ij} - \frac{(y_t)^2}{N}$

JKP = $\frac{\left(\sum_{i=1}^{\infty} y_i\right)^2}{n} - \frac{(y_t)^2}{N}$

JKG = JKT – JKP

4. Tentukan uji statistiknya

   $F_{hitung} = \frac{KTP}{KTG}$

5. Kriteria pengujian
   Terima H_0 jika $F_{hitung} < F_{tabel}$ dan tolak H_0 untuk yang lainnya
6. Kesimpulan

Contoh Soal
Terdapat 4 metode diet dan 3 golongan usia peserta program diet. Berikut data rata-rata penurunan berat peserta ke empat metode dalam tiga kelompok umur.

Sampel	Penurunan berat badan (Kg)			
	Metode 1	Metode 2	Metode 3	Metode 4
Sampel 1	4	8	7	6
Sampel 2	6	12	3	5
Sampel 3	4	-	-	5

Apakah keempat metode diet tersebut memberikan rata-rata penurunan berat badan yang sama?

Penyelesaian:

1. Hipotesis
   H_0 : Terdapat 4 metode diet dalam menurunkan berat badan
   H_a : Tidak terdapat 4 metode diet dalam menurunkan berat badan
2. Taraf Signifikansi α = 1%
3. Uji Statistik

Sampel	Penurunan berat badan (Kg)				Jumlah
	Metode 1	Metode 2	Metode 3	Metode 4	
Sampel 1	4	8	7	6	25
Sampel 2	6	12	3	5	26
Sampel 3	4	0	0	5	9
Jumlah	14	20	10	16	60

$$JKT = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} y^2_{ij} - \frac{(y_t)^2}{N}$$

$$JKT = 4^2 + 8^2 + 7^2 + 6^2 + 6^2 + 12^2 + 3^2 + 5^2 + 4^2\ 5^2 - \frac{(60)^2}{10}$$

$$JKT = 420 - 360 = 60$$

$$JKP = \frac{\left(\sum_{i=1}^{\infty} y_i\right)^2}{n} - \frac{(y_t)^2}{N}$$

$$JKP = \frac{(14)^2}{3} + \frac{(20)^2}{2} + \frac{(10)^2}{2} + \frac{(16)^2}{3} - \frac{(60)^2}{10}$$

$$JKP = 400{,}67 - 360 = 40{,}67$$

$$JKG = JKT - JKP = 60 - 40{,}67 = 19{,}33$$

SV	Db	JK	KT	Fhitung	Ftabel
Perlakuan	2	40,67	20,33	$\frac{20{,}33}{2{,}42} = 8{,}41$	F0,01 (2 vs 8) = 8,65
Galat	8	19,33	2,42		
Total	10	60			

4. Kriteria Pengujian
   Karena $F_{hitung} = 8,41 < F_{tabel} = 8,65$ maka H_0 diterima
5. Kesimpulan
   Terdapat 4 metode diet dalam menurunkan berat badan

11.2.3 Uji Dunnet

Uji dunnet bertujuan untuk menentukan perbedaan yang berarti antara tiap rataan perlakuan dengan kontrol pada taraf keberartian yang sama. Adapun syarat untuk menentukan uji dunnet harus dicari dulu Analisis varians (ANAVA) nya.

Misalkan ada Perlakuan A, B, C dan D dan keempat perlakuan itu akan diujikan masing masing terhadap kontrol. Namun tidak membandingkan antara mereka sendiri. (tidak membandingkan A dengan B, B dengan C atau pun C dengan D). (Nugroho, 2008)

Langkah-langkah:

1. Menentukan Hipotesis
   $H_0 : \mu_0 = \mu_i$
   $H_a : \mu_0 \neq \mu_i$
   i = 1,2,3, …
2. Menentukan taraf signifikansi
3. Statistik Uji
   $DLSD = t_{\alpha(p;dbgalat)} \sqrt{\frac{2KTG}{n}}$
   Dimana:
   p = banyak perlakuan – 1
   dbgalat = derajat bebas galat
   KTG = Kuadrat Total Galat
   n = banyak baris
4. Kriteria pengujian
   Jika $\left|\mu_{kontrol} - \mu_j\right|$ > DLSD maka tolak H_0 (berbeda nyata)
   Jika $\left|\mu_{kontrol} - \mu_j\right| \leq$ DLSD maka Terima H_0 (tidak berbeda nyata)

Perlakuan kontrol	Perlakuan j	$\left\|\mu_{kontrol} - \mu_j\right\|$	DLSD	Notasi

Jika $\left|\mu_{kontrol} - \mu_j\right|$ < DLSD maka diberi notasi a artinya sama

Jika $\left|\mu_{kontrol} - \mu_j\right|$ > DLSD maka diberi notasi b artinya berbeda

5. Kesimpulan

   Contoh

   Terdapat 3 metode pembelajaran dan 4 kelas di SMA Kampus. Berikut data nya:

Kelas	Metode Pembelajaran		
	Metode A	Metode B	Metode C
X1	80	75	80
X2	85	75	75
X3	75	80	75
X4	80	85	80
X5	85	80	80

Dengan menggunakan uji Dunnet, Apakah ketiga metode pembelajaran tersebut memberikan rata-rata peningkatan hasil belajar siswa yang sama?

Penyelesaian:

1. Hipotesis

   H_0 = rata-rata perlakuan kontrol sama dengan rata-rata perlakuan metode A,B,C

   H_a = rata-rata perlakuan kontrol tidak sama dengan rata-rata perlakuan metode A,B,C

2. Taraf signifikansi α = 5%

3. Uji statistik

Tabel perhitungan rata-rata

Kelas	Metode Pembelajaran			Jumlah	Rata-rata
	Metode A	Metode B	Metode C		
X1	80	75	80	235	78,33
X2	85	75	75	235	78,33
X3	75	80	75	230	76,67
X4	80	85	80	245	81,67
X5	85	80	80	245	81,67
Jumlah	405	395	390	1190	1416100
Rata-rata	81	79	78		

Tabel ANOVA

SK	JK	db	KT	F_{hitung}	F_{tabel}
K	23,33	4	5,83	$F_{hitung} = 0,34$	$F_{tabel} = 3,478$
G	170	10	17		
T	193,33	14	13,81		

$$JKT = \sum_{i=1}^{k}\sum_{j=1}^{n} X^2{}_{ij} - \frac{T_i^{\,2}}{N}$$

$$= (80)^2 + (85)^2 + (75)^2 + (80)^2 + (85)^2 + (75)^2 + (75)^2 + (80)^2 + (85)^2 + (80)^2 + (80)^2 + (75)^2 + (75)^2 + (80)^2 + (80)^2 - \frac{(1190)^2}{15} = 193,3333$$

$$JKK = \sum_{i=1}^{k} \frac{T_i^2}{k} - \frac{T_i^2}{N}$$

$$= \frac{(405)^2}{5} + \frac{(395)^2}{5} + \frac{(390)^2}{5} - \frac{(1190)^2}{15} = 23{,}3333$$

JKG = JKT – JKK = 193,3333 - 23,3333 = 170

$$DLSD = t_{\alpha(p;dbgalat)} \sqrt{\frac{2KTG}{n}}$$

$$DLSD = t_{0{,}05(2;10)} \sqrt{\frac{2(17)}{5}}$$

$$DLSD = (2{,}23)\sqrt{6{,}8}$$

DLSD = (2,23)(2,67681)

DLSD = 5,82

4. Kriteria pengujian
   Absolut (rata-rata perlakuan kontrol – perlakuan j) < DLSD maka diberi notasi a artinya sama

Perlakuan control	Perlakuan j	$\lvert \mu_{kontrol} - \mu_j \rvert$	DLSD	Notasi
X1	A	2,67	5,82	a
X1	B	0,67	5,82	a
X1	C	0,33	5,82	a

5. Kesimpulan
   Apabila perlakuan kontrol diambil x1 maka rata-rata perlakuan x1 dengan perlakuan metode ABC adalah sama

11.2.4 Uji Kruskall Wallis

1. Pengembangan dari model Mann-Whitney
2. Digunakan membandingkan dua atau lebih rata-rata populasi secara bersama-sama
3. Apakah ada kesamaan nilai varians dari populasi nya
4. Populasi berdistribusi normal
5. Populasi nya mempunyai nilai standar deviasi yang sama

6. Sampel dari populasi bersifat saling bebas
7. Paling tidak data nya ordinal (Karmini, 2020)

Langkah-langkah:

1. Menentukan hipotesis
   $H_0 : \mu_1 = \mu_2$
   $H_a : \mu_1 \neq \mu_2$
2. Menentukan taraf signifikansi
3. Uji statistik

$$T_{hitung} = \frac{12}{N(N+1)} \left(\frac{\left(\sum_{j=1}^{K} R_j \right)^2}{n_j} \right)$$

   Dimana:
   N : jumlah data keseluruhan
   R_j : jumlah kolom ke-j (setelah dirangking)
   n_j : banyak data tiap kolom
4. Kriteria pengujian
   H_0 diterima jika $T_{tabel} > T_{hitung}$ atau sebaliknya H_0 Ditolak jika $T_{tabel} < T_{hitung}$ dengan T_{tabel} diperoleh dari tabel chikuadrat yaitu $\chi^2_{\alpha,n-1}$
5. Kesimpulan

Contoh Soal

Sejalan dengan adanya kecenderungan terjadinya demotivasi kerja dilingkungan perusahaan yang dipimpinnya, direktur perusahaan tersebut meminta kepada kepala bagian kepegawaian untuk meneliti apa yang menyebabkan terjadinya demotivasi tersebut. Sesuai perintah atasan selanjutnya kepala bagian kepegawaian melakukan penelitian dengan asumsi bahwa terjadinya demotivasi akibat dari sistem penggajian yang selama ini diberlakukan kepada para karyawannya. Hasil dari penyebaran angket yang disebarkan 4 divisi masing-masing diambil sampel 26

orang tiap divisi dengan menggunakan pola pertanyaan berperingkat yang tertutup, hasilnya seperti tabel berikut:

Sistem Penggajian	Frekuensi Hasil Jawaban Pertanyaan dari Divisi			
	Pemasaran	Produksi	Persediaan	Pengadaan
Sangat baik	4	2	8	1
Baik	4	6	5	3
Cukup Baik	6	4	4	7
Kurang Baik	8	6	7	9
Sangat Kurang Baik	4	8	2	6
Jumlah	26	26	26	26

Uji lah dengan menggunakan tingkat kepercayaan 95%, apakah informasi mengenai terjadinya demotivasi para karyawannya akibat dari sistem penggajian yang diberikan?

Penyelesaian:

1. H_0 : sistem penggajian yang diterapkan pada karyawannya bukan merupakan faktor utama terjadi nya demotivasi
   H_a : sistem penggajian yang diterapkan pada karyawannya merupakan faktor utama terjadi nya demotivasi
2. Taraf signifikansi 5%
3. $T_{hitung} = \frac{12}{N(N+1)} \left(\frac{\left(\sum_{j=1}^{K} R_j \right)^2}{n_j} \right)$

Perhitungan peringkat atau rangking

No	Data	Data diurutkan	Rangking
1	4	1	7
2	4	2	7
3	6	2	12,5
4	8	3	18
5	4	4	7
6	2	4	2,5

No	Data	Data diurutkan	Rangking
7	6	4	12,5
8	4	4	7
9	6	4	12,5
10	8	5	18
11	8	6	18
12	5	6	10
13	4	6	7
14	7	6	15,5
15	2	7	2,5
16	1	7	1
17	3	8	4
18	7	8	15,5
19	9	8	20
20	6	9	12,5

Daftar distribusi frekuensi

No	frekuensi	rangking
1	1	1
2	2	2,5
3	2	2,5
4	3	4
5	4	7
6	4	7
7	4	7
8	4	7
9	4	7
10	5	10
11	6	12,5
12	6	12,5
13	6	12,5
14	6	12,5

No	frekuensi	rangking
15	7	15,5
16	7	15,5
17	8	18
18	8	18
19	8	18
20	9	20

Data setelah di ubah kedalam perangkingan atau peringkat

Sistem Penggajian	Frekuensi Hasil Jawaban Pertanyaan dari Divisi			
	Pemasaran	Produksi	Persediaan	Pengadaan
Sangat baik	7	2,5	18	1
Baik	7	12,5	10	4
Cukup Baik	12,5	7	7	15,5
Kurang Baik	18	12,5	15,5	20
Sangat Kurang Baik	7	18	2,5	12,5
Jumlah	51,5	52,5	53	53
jumlah di kuadratkan	2652,25	2756,25	2809	2809

$$T_{hitung} = \frac{12}{N(N+1)}\left(\frac{\left(\sum_{j=1}^{K} R_j\right)^2}{n_j}\right) =$$

$$\frac{12}{20(20+1)}\left(\frac{(51,5)^2}{5}+\frac{(52,5)^2}{5}+\frac{(53)^2}{5}+\frac{(53)^2}{5}\right) = 63,009$$

4. $T_{tabel} = \chi^2_{tabel} = \chi^2_{(0,05;4-1)} = \chi^2_{(0,05;3)} = 7,81$
   Karena $T_{hitung} > T_{tabel}$ yaitu (63,009 > 7,81) maka H_0 ditolak
5. Kesimpulan
   sistem penggajian yang diterapkan pada karyawannya merupakan faktor utama terjadi nya demotivasi

Soal Latihan

1. Dalam satu teknik wawancara, orang yang diwawancarai berperan pasif. Wawancara berlangsung dengan kartu yang diisi dan dijawab oleh orang yang diwawancarai. Teknik wawancara kedua pewawancara berinteraksi secara hangat dan bertatap muka dengan orang yang diwawancarai. Pewawancara mengajukan pertanyaan dan memberikan komentar pada saat yang diwawancarai memberikan jawaban. Tekanan darah diastolik diukur pada saat selang waktu satu menit selama wawancara berlangsung. Berikut tabel hasil wawancara nya:

Wawancara	Perubahan Tekanan Darah	
	Cukup Besar	Sangat Kecil
Tatap Muka	6	0
Kartu	1	5

   Berdasarkan data di atas, apakah wawancara dengan tatap muka memberikan perubahan yang lebih besar terhadap tekanan darah diastolik?

2. Terdapat 3 metode pembelajaran dan 4 kelas di SMA Kampus. Berikut data nya:

Kelas	Metode Pembelajaran		
	Metode A	Metode B	Metode C
X1	70	75	80
X2	80	80	75
X3	75	75	70
X4	70	70	80

   Apakah keempat metode pembelajaran tersebut memberikan rata-rata peningkatan hasil belajar siswa yang sama?

3. Misalkan dalam suatu percobaan industri, seorang insinyur ingin menyelidiki bagaimana rataan penyerapan uap air dalam beton berubah diantara 5 adukan beton yang berbeda. Adukan beton ini berubah dalam persen berat komponen penting. Sampel dibiarkan kena uap air selama 48 jam. Dari tiap adukan diambil 6 mapel untuk di uji sehingga seluruhnya diperlukan 30 sampel. Uji lah

$\mu_1 = \mu_2 = ... = \mu_5$ pada taraf keberartian 0,05 untuk data dibawah ini mengenai penyerapan uap air oleh berbagai jenis adukan semen.

Sampel	Jenis adukan beton (% Berat)				
	1	2	3	4	5
A	551	595	639	417	563
B	457	580	615	449	631
C	450	508	511	517	522
D	731	583	573	438	613
E	499	633	648	415	656
F	632	517	677	555	679

Dengan menggunakan uji Dunnet, Apakah kelima jenis adukan beton tersebut memberikan rata-rata penyerapan uang air yang sama?

4. Pada dasarnya pemberian pelayanan kepada para pelanggan disetiap supermarket adalah sama baiknya. Untuk mengecek kebenaran adanya anggapan tersebut selanjutnya seorang mahasiswa melakukan penelitian terhadap 4 supermarket yang ada di wilayah tersebut, dengan menggunakan model jawaban tertutup (5 menyatakan sangat baik, 4 menyatakan baik, 3 menyatakan cukup baik, 2 menyatakan tidak baik dan 1 menyatakan sangat tidak baik) dan ternyata diperoleh data sebagai berikut:

No. Responden	Hasil Penilaian Responden terhadap supermarket			
	A	B	C	D
1	3	4	3	4
2	3	4	4	4
3	2	4	3	4
4	3	5	4	3
5	3	4	3	3
6	4	5	4	3
7	2	5	4	2
8	2	4	4	2
9	4	4	4	2
10	4	3	4	3
11	4	4	3	4

Ujilah anggapan diatas dengan menggunakan tingkat kepercayaan 95%, apakah betul, bahwa pelayanan di tiap super market sama baiknya.

DAFTAR PUSTAKA

Karmini. (2020). *Statistika Non Paramentrik*.

Mashuri, A. (2015). *Buku Ajar Statistika Non Parametrik*. 6.

Nugroho, S. (2008). *Statistika Nonparametrika*.

Rahman, A., Pd, S., & Si, M. (n.d.). *Statistika Parametrik*.

Setiawan, N. (2018). Statistika Nonparametrik untuk Penelitian Sosial Ekonomi Peternakan. *Statistika Non Parametrik*, 101. http://pustaka.unpad.ac.id/wp-content/uploads/2009/03/statistika_nonparametrik.pdf

Suyanto dan Prana Ugiana Gio. (2015). *Statistika Non Parametrik dengan SPSS, Minitab dan R*.

BIODATA PENULIS

Samsul Bahri Loklomin, M.Si.
Program Studi Statistika
Fakultas Sains dan Teknologi Universitas Pattimura

Penulis lahir di Seram Bagian Timur, Provinsi Maluku. Penulis adalah dosen tetap pada Program Studi Statistika Fakultas Sains dan Teknologi (FST) Universitas Pattimura. Menyelesaikan pendidikan S1 pada Program Studi Matematika Jurusan Matematika Fakultas MIPA Universitas Pattimura dan melanjutkan S2 Program Studi Statistika Jurusan Statistika Fakultas MIPA Institut Teknologi Sepuluh November-Surabaya. Beberapa mata kuliah yang pernah diampu penulis diantaranya; Statistika Dasar, Statistika Pendidikan, Statistika Sosial, Teori Probabilitas, Statistika Matematika, Analisis Data Survival, Algoritma dan Pemograman, Biostatistika, Statistika Bayesian, Pengendalian Kualitas Statistika dan Pengantar Teori Keputusan.

BIODATA PENULIS

Muh. Khaedir Lutfi, S.Pd., Gr., M.Pd.
Dosen Program Studi Pendidikan Matematika
Fakultas Keguruan dan Ilmu Pendidikan
Universitas Tangerang Raya

Penulis lahir di Ujung Pandang, 28 April 1991. Penulis adalah dosen tetap pada Program Studi Pendidikan Matematika Fakultas Keguruan dan Ilmu Pendidikan, Universitas Tangerang Raya. Penulis menyelesaikan pendidikan S1 pada Program Studi Pendidikan Matematika di Unversitas Muhammadiyah Makassar tahun 2014 dan PPG di Universitas Islam Nusantara pada tahun 2017. Penulis menyelesaikan studi S2 di Universitas Pendidikan Indonesia pada Program Studi Pendidikan Matematika. Salah satu matakuliah yang diampu oleh penulis adalah matakuliah Statistika. Penulis juga aktif dalam mempublikasikan karya ilmiah yang beberapa di antaranya menggunakan uji statistik. Selain itu, penulis juga terlibat dalam berbagai organisasi atau komunitas, seperti Indonesia Mathematics Educators Society (I-MES). Penulis juga sering membuat konten Pembelajaran Matematika melalui Media Sosial Math Couple (Instagram, Youtube, dan Tiktok). Untuk keperluan profesional, penulis dapat dihubungi melalui email di muh.khaedir.lutfi@gmail.com atau melalui Akun Instagram Muh. Khaedir Lutfi.

BIODATA PENULIS

Hanna Hilyati Aulia, M.Si.
Dosen Program Studi Ekonomi Syari'ah
Fakultas Ekonomi dan Bisnis Islam

Penulis adalah seorang akademisi dengan minat mendalam pada pemodelan matematika yang mencakup berbagai aplikasi dalam epidemiologi dan ekonomi. Keahliannya meliputi pengembangan model matematis, analisis statistik, serta penerapan kontrol optimal untuk menyelesaikan masalah-masalah kompleks di bidang tersebut. Ia juga memiliki ketertarikan pada pendekatan inovatif dalam pengajaran matematika, mengintegrasikan seni dan kreativitas sebagai alat untuk mempermudah pemahaman konsep-konsep abstrak. Dengan latar belakang yang kuat di bidang matematika industri dan keuangan, penulis terus mengeksplorasi cara-cara baru untuk menghubungkan teori dengan praktik, baik melalui penelitian maupun kegiatan pengajaran.

BIODATA PENULIS

Yusup Junaedi, S.Pd., M.Pd.
Dosen Program Studi Pendidikan Matematika
FKIP, Universitas La Tansa Mashiro

Penulis lahir di Lebak, 8 Januari 1995. Penulis adalah dosen tetap pada Program Studi Pendidikan Matematika di Universitas La Tansa Mashiro. Menyelesaikan pendidikan S1 pada Jurusan Pendidikan Matematika di Universitas Sultan Ageng Tirtayasa pada tahun 2017 dan melanjutkan S2 pada Jurusan Pendidikan Matematika di Universitas Pendidikan Indonesia yang lulus pada tahun 2020. Penulis memiliki keahlian dibidang aljabar, aljabar linear, dan aljabar abstrak. Selain menjadi pengajar aktif, penulis memiliki beberapa prosiding terindeks scopus, jurnal terindeks SINTA. Memperoleh hibah penelitian dan pengabdian kepada masyarakat serta peraih dana Kosabangsa 2023 dan 2024. Menjadi koordinator dosen pembimbing Kampus Mengajar Provinsi Banten, Pengajar Praktik program pendidikan guru penggerak, dan team leader pada program survei nasional.

BIODATA PENULIS

Agus Winarji, M.Pd.
Dosen Program Studi Pendidikan Matematika
Fakultas Keguruan dan Ilmu Pendidikan

Penulis lahir di Ketapang tanggal 13 Agustus 1986. Penulis adalah dosen pada Program Studi Pendidikan Matematika Fakultas Keguruan dan Ilmu Pendidikan, Universitas Tanjungpura. Menyelesaikan pendidikan S1 pada program studi Pendidikan Matematika jurusan PMIPA di Universitas Tanjungpura dan melanjutkan S2 pada Pendidikan Matematika Departemen PMIPA di Universitas Pendidikan Indoensia.

BIODATA PENULIS

Ridha Yuniara, S.Pd., M.Pd.
Dosen Program Studi Ekonomi Pembangunan
Fakultas Ekonomi dan Bisnis Universitas Gajah Putih

Penulis lahir di Takengon tanggal 11 Juni 1993. Penulis adalah dosen tetap pada Program Studi Program Studi Ekonomi Pembangunan Fakultas Ekonomi dan Bisnis Universitas Gajah Putih. Menyelesaikan pendidikan S1 pada Jurusan Pendidikan Matematika Universitas Syiah Kuala dan melanjutkan S2 pada Jurusan yang sama juga di Universitas yang sama.

BIODATA PENULIS

Mia Audina Musyadad, S.Pd., Gr., M.Pd.
Dosen Program Studi Pendidikan Guru Madrasah Ibtidaiyah
Sekolah Tinggi Ilmu Tarbiyah Rakeyan Santang

Penulis lahir di Karawang tanggal 29 Mei 1994. Penulis adalah dosen tetap Matematika pada Program Studi Pendidikan Guru Madrasah Ibtidaiyah, Sekolah Tinggi Ilmu Tarbiyah Rakeyan Santang Karawang. Menyelesaikan pendidikan S1 pada Jurusan Pendidikan Matematika di UIN Sunan Gunung Djati Bandung dan melanjutkan S2 pada Jurusan Pendidikan Matematika di Universitas Pendidikan Indonesia. Penulis juga merupakan lulusan PPG Prajabatan dari Universitas Pasundan Bandung.

BIODATA PENULIS

Fitri Anisa Kusumastuti, M.Pd
Dosen Program Studi Pendidikan Matematika
Universitas Tangerang Raya

Penulis lahir di Wonogiri tanggal 29 Maret 1995. Penulis adalah dosen tetap pada Program Studi Pendidikan Matematika Fakultas Keguruan dan Ilmu Pendidikan, Universitas Tangerang Raya. Menyelesaikan pendidikan S1 pada Jurusan Pendidikan Matematika di Universitas Negeri Surabaya dan melanjutkan S2 pada Jurusan Pendidikan Matematika Universitas Pendidikan Indonesia. Saat ini menekuni bidang literasi-numerasi, literasi matematis, matematika sekolah, serta pendidikan dan pengajaran matematika.

BIODATA PENULIS

Novela Wulandari, M.Pd.
Dosen Program Studi Pendidikan Matematika
FKIP Universitas Tanggerang Raya

Penulis lahir di Lais Bengkulu, tanggal 11 November 1992. Anak ke dua dari empat bersaudara dari pasangan bapak Zulkifli dan Ibu Rosmani. Dosen tetap Program Studi Pendidikan Matematika FKIP di Universitas Tanggerang Raya.

Riwayat pendidikan SD Negeri 01 Medan Jaya tahun 1999-2005, SMP Negeri 02 Muko-muko tahun 2005-2008, SMA Negeri 02 Muko-muko tahun 2008-2011, kemudian melanjutkan pendidikan S1 pada Jurusan Pendidikan Matematika pada tahun 2011 dan melanjutkan S2 pada Jurusan Pendidikan Matematika pada tahun 2018.

Beberapa pengalaman kerja penulis antara lain : Guru di SMA Negeri 11 Kota Bengkulu (2016-2018) kemudian penulis Resign karena melanjutkan kuliahnya di UPI pada tahun 2018 tersebut. Guru Matematika SMP Al- Azhar Syifabudi Kemang Jakarta Selatan (2022-2023). Dosen program studi Pendidikan matematika FKIP Universitas Tanggerang Raya (2023-2024). Mengikuti pada Program kampus UAD Yogyakrta Guru matematika di Sekolah songserm Wittaya Muniti School dan Suksasat Thailand (2014).

BIODATA PENULIS

Tedi Hartoyo, Ir. MSc.
Dosen Program Studi Agribisnis
Fakultas Pertanian Universitas Siliwangi

Penulis lahir di Tasikmalaya tanggal 26 Juli 1962. Penulis adalah Dosen Program Studi Agribisnis Fakultas Pertanian Universitas Siliwangi. Menyelesaikan pendidikan S1 pada Jurusan Statistika IPB Bogor dan melanjutkan S2 pada Agriculture Development, Gent Unversity Belgium. Saat ini penulis mengampu beberapa mata kuliah, diantaranya Matematika, Statistika, Ekonometrika, Metode Penelitian Sosial Ekonomi dan Perancangan Percobaan.

BIODATA PENULIS

Rektor Sianturi, S.Pd., M.Si
Dosen Program Studi Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam

Rektor Sianturi, Lahir di Pansur, 26 Mei 1981 sebagai anak ke 6 dari 7 bersaudara dari pasangan Alm. Bapak Mustar Sianturi dan Alm. Ibu Selma Silitonga. Pendidikan dasar penulis tempuh di SD Inpres Pansur di Kecamatan Tanah Jawa Kabupaten Simalungun, selesai pada Tahun 1993 melanjutkan sekolah di SMP Swasta HKBP Tanah Jawa Kabupaten Simalungun, selesai pada tahun 1996 melanjutkan sekolah di SMA Swasta Dharma Bakti Tanah Jawa, selesai pada Tahun 1999 melanjutkan kuliah di Fakultas Keguruan dan Ilmu Pendidikan Universitas HKBP Nommensen dan selesai pada tahun 2004. Pada tahun 2013 melanjutkan ke Program Pasca Sarjana Universitas Sumatera Utara jurusan Matematika dan lulus pada tahun 2015. Pada tahun 2005 sampai dengan tahun 2021, penulis pernah bekerja sebagai guru matematika di SMA Swasta Teladan Pematangsiantar. Dan pada tahun 2009 sampai dengan tahun 2021, pernah bekerja sebagai guru di SMA Negeri 5 kota Pematangsiantar. Kemudian pada tahun 2021 sampai sekarang menjadi Dosen di Universitas HKBP Nommensen Pematangsiantar.