catatan kuliah

analisis data

robby

Catatan Kuliah Analisis Data

Edisi 2015-2016

-Robby-

Special thanks to:

Cover Designer : Feliana Eka Dewi

Editor : Emmanuel Arya

Kenan Mahesa Sudiro

Catatan:

Buku ini adalah ringkasan dari pustaka utama, semua gambar dan materi mengacu pada pustaka utama tersebut.

Daftar Isi

Prakata

Puji syukur kepada Tuhan yang Maha Esa atas kebaikan-Nya sehingga penyusun dapat menyelesaikan **Catatan Kuliah Analisis Data** ini.
Buku ini berisi rangkuman materi yang diajarkan berdasarkan mata kuliah *MA2181 - Analisis Data*. Buku ini berisi 13 bab yang merujuk pada pustaka utama "*Probability & Statistics for Engineers & Scientists* 9^{th} *Edition*" karangan Walpole, Myers. Buku ini berisi teorema-teorema penting, contoh soal, berikut soal-soal latihan dalam tiap babnya yang merujuk pada kumpulan soal latihan Analisis Data KK Statistika FMIPA ITB. Adapun rujukan semua gambar dalam buku ini berdasarkan pustaka utama yang digunakan penyusun.
Penyusun tentu tidak dapat mengerjakan buku ini atas kerjanya sendiri. Penyusun ingin mengucapkan terima kasih kepada Bapak Sapto Wahyu Indratno selaku dosen penyusun pada saat penyusun mengambil kuliah ini. Penyusun juga ingin menyampaikan terima kasih kepada Emmanuel Arya dan Kenan Mahesa Sudiro selaku editor dari buku ini, Stephanus Ardyanto yang telah memberikan kritik dan saran, serta Feliana Eka Dewi yang telah membuat cover buku ini.
Penyusun menyadari bahwa buku ini belum sempurna. Sehingga penyusun mengharapkan kritik dan saran dari para pembaca mengenai materi, cara penyajian, dan soal latihan guna memperbaiki buku ini.

Bandung, Agustus 2015

Penyusun

Bab **1**

Statistika Deskriptif

Pada SMA, kita sudah mengetahui maksud dari mean, median, modus, dan lain sebagainya. Hal tersebut termasuk dalam bidang statistika deskriptif, yaitu pengolahan data dari sampel yang diberikan, mulai dari ukuran pemusatan data hingga ukuran penyebaran data. Dalam bab ini kita akan mempelajari tentang hal-hal mendasar dalam statistika serta penyajian data dan manipulasi data, berikut bagaimana cara kita menjelaskan hasil pengolahan data yang didapat kepada orang awam yang tidak mengerti tentang statistika.

1.1 Beberapa Terminologi

Pertama-tama kita bahas beberapa istilah yang seringkali disalahpahami oleh beberapa orang, yaitu statistik dan statistika. **Statistik** adalah nilai-nilai ukuran data yang mudah dipahami, sedangkan **statistika** adalah ilmu yang berkaitan dengan cara pengumpulan, pengolahan, analisis, dan penarikan kesimpulan tentang data. Bidang statistika terbagi menjadi dua bagian, yaitu:

- Statistika deskriptif: pengumpulan dan penyajian data.
- Statistika inferensi: analisis sampel untuk penarikan kesimpulan tentang populasi.

Dalam bab ini, kita akan mempelajari mengenai statistika deskriptif terlebih dahulu. Namun kita akan melihat terlebih dahulu gambaran kecil tentang statistika inferensi. Pertama-tama kita ketahui terlebih dahulu apa itu sampel dan populasi. Prinsip dasarnya, **Sampel** adalah bagian dari **populasi**. Dalam sebuah penelitian, misalnya tinggi badan seluruh warga Indonesia, kita tidak mungkin mendata seluruh warga Indonesia, namun kita dapat mengambil beberapa orang yang menjadi *perwakilan* dari populasi warga Indonesia tersebut. Beberapa orang itu kita sebut sebagai **sampel**. Namun, pengambilan sampel haruslah secara acak, agar dapat mewakilkan populasi dengan baik.

Terminologi terakhir adalah mengenai jenis-jenis observasi. Berikut rincian dari jenis observasi:

- Observasi kualitatif
  - Nominal: *tidak mengenal* urutan dan operasi aritmetika. Contoh: jenis bidang datar: persegi, lingkaran, dll.
  - Ordinal/Rank: *mengenal* urutan dan operasi aritmetika. Contoh: kelas penumpang pesawat ekonomi, bisnis.
- Observasi kuantitatif
  - Diskrit: himpunan yang meliputi semua bilangan bulat (*terhitung*). Contoh: banyaknya peserta kuliah analisis data.
  - Kontinu: himpunan yang meliputi semua bilangan real (*terukur*). Contoh: waktu tempuh pelari dalam sebuah perlombaan.

1.2 *Sari Numerik*

Selanjutnya kita akan membahas mengenai elemen terpenting dalam statistika deskriptif, yaitu sari numerik. Sari numerik adalah statistik yang mendeskripsikan sifat dari suatu data secara numerik (dalam bentuk angka, bukan narasi). Sari numerik dibagi menjadi dua, yaitu ukuran pemusatan data dan ukuran penyebaran data.

1.2.1 *Ukuran Pemusatan Data*

Dalam ukuran pemusatan data, hal yang dilihat adalah di manakah data-data tersebut berkumpul dengan jumlah tertentu.

Sari numerik yang pertama yaitu *mean* atau rataan ($\bar{x}$). Jika terdapat data $x_1, x_2, \cdots, x_n$, maka rataannya adalah:

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

Selanjutnya *median* atau nilai tengah yang membagi dua kelompok data sama banyak ($\tilde{x}$).

$$\tilde{x} = \begin{cases} x_{(n+1)/2} & n \text{ ganjil} \\ \frac{1}{2}(x_{n/2} + x_{n/2+1}) & n \text{ genap} \end{cases}$$

Bentuk median dapat diperumum menjadi nilai yang membagi empat kelompok data sama banyak, yang disebut **kuartil** (Q). Kuartil terbagi menjadi kuartil bawah Q_1 yaitu nilai yang membagi seperempat pertama data, kuartil tengah Q_2 (yang sama dengan median), dan kuartil atas Q_3 yaitu nilai yang membagi tigaperempat data. Variasi lain yang analog dengan kuartil yaitu desil (membagi sepuluh data sama banyak) dan persentil (membagi seratus data sama banyak). Coba tentukan bentuk umum dari kuartil, desil, dan persentil.

Terakhir, *modus* atau nilai yang paling sering keluar. Perlu diperhatikan bahwa nilai modus boleh lebih dari satu.

Contoh Soal 1. Diberikan databanyaknya pelanggan yang datang ke sebuah mini market di 15 hari tertentu pada bulan Juli 2011 sebagai berikut.

26	37	76	49	95	69	83	87	39	95	59	83	83	87	46

Tentukan mean, median, dan modusnya.
Solusi. Data diurutkan terlebih dahulu untuk mempermudah perhitungan.

26	37	39	46	49	59	69	76	83	83	83	87	87	95	95

Rataannya:

$$\bar{x} = \frac{26 + 37 + 39 + \cdots + 95}{15} = 67.60$$

Nilai tengahnya:

$$\tilde{x} = x_8 = 76$$

Modusnya adalah 83. ■

1.2.2 *Ukuran Penyebaran Data*

Selanjutnya mengenai ukuran penyebaran data. Jika pusat dari suatu data adalah rataan, ukuran penyebaran data adalah statistik yang menjelaskan seberapa jarak data-data dari pusatnya. Secara wajar, kita dapat menghitung jarak data dari pusatnya dengan menjumlahkan semua selisih tiap data dengan rataannya, yaitu:

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

Namun hal tersebut tidak memberikan informasi apapun karena:

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

Maka dari itu, untuk menghitung jarak dari suatu data ke pusatnya, kita gunakan jumlah kuadratnya:

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

Beberapa statistik yang menjelaskan ukuran penyebaran data antara lain:
Jangkauan (*range*):

$$R = x_{\max} - x_{\min}$$

Ragam (*variansi*):

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

Simpangan baku (*standar deviasi*):

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

dan Jangkauan Antar Kuartil:

$$dq = Q_3 - Q_1$$

Jangkauan antar kuartil digunakan untuk mendeteksi pencilan. **Pencilan** adalah data yang berbeda secara drastis dengan data-data lainnya. Cara mendeteksi sebuah pencilan adalah dengan mendefinisikan batas atas pencilan (BAP) dan batas bawah pencilan (BBP) sebagai berikut:

$$BBP = Q_1 - k \cdot dq \qquad\qquad BAP = Q_3 + k \cdot dq$$

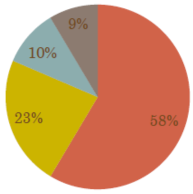
dengan $k = \frac{3}{2}$ (standar, tidak mutlak). Pencilan adalah data yang di bawah BBP atau di atas BAP.
Statistik lainnya lebih berkontribusi pada pembentukan distribusi data, yaitu kemiringan (*skewness*) dan kelancipan (*kurtosis*). Perhitungan kemiringan dan kelancipan distribusi data cukup sulit untuk dilakukan secara manual, namun kita dapat mengetahui kemiringan suatu distribusi data melalui mean dan median.

- Distribusi data **simetris** jika $\bar{x} = \tilde{x}$.
- Distribusi data **miring ke kanan** jika $\bar{x} > \tilde{x}$.
- Distribusi data **miring ke kiri** jika $\bar{x} < \tilde{x}$.

1.3 Penyajian Data

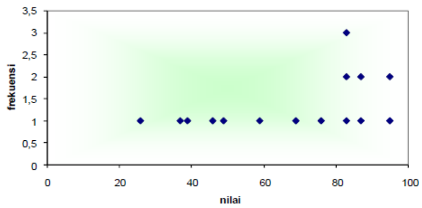
Dari sari numerik yang sudah didapat, kita dapat menyajikan data tersebut dengan berbagai macam bentuk. Contoh diagram yang dakan diberikan berdasarkan dari data pada contoh soal 1.

- **Pie Chart**. Pie chart merupakan grafik yang berbentuk lingkaran yang setiap potongannya mewakili proporsi atau persentase suatu komponen dari sebuah kelompok data (100%). Pemakaian pie chart hanya cocok ketika menyatakan data dalam bentuk proporsi dari satu kelompok data. Berikut contohnya:



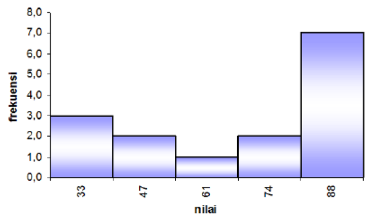
Gambar 1.1: Pie Chart untuk Data Contoh Soal 1

- **Dot Plot**. Cara menggambarkan data dalam bentuk titik, dengan memperhatikan frekuensi (jumlah suatu data) dari data yang bersangkutan. Titik ditumpuk di atas nilai data yang digambarkan. Contohnya sebagai berikut:



Gambar 1.2: Dot Plot untuk Data Contoh Soal 1

- **Histogram**. Histogram adalah gambar berdasarkan distribusi frekuensi. Setiap frekuensi dipresentasikan oleh suatu segi empat. Daerah setiap segi empet sebanding dengan frekuensinya. Berikut contohnya:



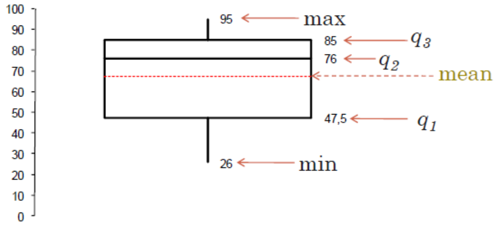
Gambar 1.3: Histogram untuk Data Contoh Soal 1

- **Diagram Batang-Daun (Stem-Leaf Plot)**. Stem atau batang, mirip dengan grup data pada histogram, sedangkan leaf atau daun, mirip dengan frekuensi. Stem atau batang adalah digit pertama yang terpenting yang ada dalam bilangan yang membentuk harga data, sedangkan digit di belakangnya akan merupakan leaf atau daun. Stem-leaf masih dapat dilihat nilai data mentahnya. Contohnya adalah sebagai berikut:

2	6
3	7 9
4	6 9
5	9
6	9
7	6
8	3 3 3 7 7
9	5 5

Keterangan: batang menunjukkan puluhan, daun menunjukkan satuan.

- **Diagram Kotak-Titik (Box Plot)**. Box Plot digunakan untuk menyelidiki distribusi tanpa menggunakan grup data seperti pada histogram dan diagram batang daun. Box Plot terdiri dari: $x_{\min}$, Q_1, Q_2, Q_3, dan $x_{\max}$ yang disusun secara terurut dengan membentuk kotak. Contohnya adalah sebagai berikut:



Gambar 1.4: Box Plot untuk Data Contoh Soal 1

1.4 *Soal Latihan*

1. Diberikan data besar klaim (*severity*) asuransi kendaraan bermotor (dalam puluhan juta rupiah) yang dilaporkan masing-masing pemiliki 10 (sepuluh) mobil yang dipilih secara acak dari suatu portofolio asuransi kendaraan bermotor:

20	24	19	94	25	19	30	44	25	88

   a) Tentukan mean dan simpangan baku dari data besar klaim tersebut.
   b) Gambarkan *box plot* dari data *severity* tersebut. Beri komentar.

2. Diagram batang dan daun dari diameter sejenis pohon di suatu hutan adalah sebagai berikut (dalam mm):

44	3 5
43	9 9
43	0 0 1 2 2 3 3
42	5 6 6 6 6 8 8 8 9 9 9 9 9
42	0 1 1 1 2 2 3 3 4
41	5 6 7 8 8 9 9
41	2 3 3

   Keterangan: batang menunjukkan satuan, daun menunjukkan sepersepuluhan.

   a) Tentukan ukuran pemusatan data (median dan rata-rata).
   b) Tentukan ukuran penyebaran data (variansi dan standar deviasi).
   c) Selidiki adakah pencilan pada data di atas dan berikan penjelasan.
   d) Buat diagram kotak dan titik.

3. Tabel berikut menyatakan lama waktu yang diperlukan (dalam jam) untuk mengerjakan tiga tipe soal ujian yang berbeda yang diberikan pada 25 peserta kuliah Analisis Data tahun 2006.

Tipe	Waktu pengerjaan (jam)							
I	2.5	2	2.6	1.5	2.25	1.75	2.3	2	
II	2.75	2.5	3.25	2.5	2.35	2.3	2.5	2	2.25
III	3	3.15	3.75	4	2.5	3.5	2.75	1.75	

   a) Hitung rataan dan variansi masing-masing tipe soal.
   b) Dengan skala yang sama, buatlah boxplot untuk masing-masing tipe soal di atas, kemudian bandingkan. Berikan ulasan Anda.

Bab 2

Peluang dan Aturan Bayes

Statistika tentu tidak terlepas dari peluang. Peluang tersebut yang membuat data-data menjadi beragam. Pada bab ini kita akan mempelajari mengenai hal-hal dasar mengenai peluang serta Aturan Bayes sebagai perluasan dari peluang bersyarat.

2.1 Terminologi

Dalam membahas peluang, kita harus mengetahui istilah-istilah yang digunakan terlebih dahulu serta elemen-elemen penting yang digunakan. Elemen penting yang pertama adalah eksperimen. Eksperimen dalam bidang statistika seringkali digunakan dalam statistika inferensi, yang berkaitan erat dengan peluang. Eksperimen yang digunakan adalah eksperimen acak dalam membuat sampel agar dapat mewakili populasi. Ciri-cirinya antara lain:

- Dapat diulangi baik oleh si pengamat sendiri maupun orang lain.
- Proporsi keberhasilan dapat diketahui dari hasil-hasil sebelumnya.
- Bisa diukur.
- Hasilnya tidak bisa ditebak karena adanya galat (*error*).

Setelah kita melakukan eksperimen acak, kita akan mendapatkan sebuah himpunan dari semua kemungkinan hasil dari eksperimen acak tersebut, yang kita sebut seagai **ruang sampel** (biasanya dinotasikan $\mathcal{S}$). Ruang sampel dibagi menjadi dua kategori:

- **Ruang sampel diskrit**: banyaknya elemen pada $\mathcal{S}$ dapat dihitung/dicacah (*countable*). Hasil pencacahannya mungkin berhingga atau tidak terhingga. Contoh: percobaan survei warga Bandung, lalu dikelompokkan melalui jenis kelaminnya sebagai laki-laki atau perempuan.
- **Ruang sampel konitnu**: elemen-elemen dari $\mathcal{S}$ adalah bagian dari suatu interval. Contoh: percobaan pengukuran berat badan warga Bandung dalam kg, misal $\mathcal{S} = \{x : 30 < x < 100\}$.

Tentu saja kita dapat melakukan percobaan yang menghasilkan subset dari ruang sampel. Himpunan tersebut kita namakan kejadian.

Untuk sebuah kejadian (kita notasikan E), kita dapat mengelompokkan semua elemen yang tidak termasuk dalam kejadian, namun masih termasuk ke dalam ruang sampel $\mathcal{S}$. Himpunannya kita sebut **komplemen** dari kejadian E terhadap ruang sampel $\mathcal{S}$, dinotasikan E' atau E^C.

Jika kita meninjau dua kejadian, ada suatu kemungkinan di mana kita harus membandingkan kedua elemen dalam kejadian tersebut. Sebagai ilustrasi, misalkan dalam sebuah ruang sampel pelemparan dadu $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$, A menyatakan kejadian mata dadu prima $A = \{2, 3, 5\}$, dan B menyatakan kejadian mata dadu genap $B = \{2, 4, 6\}$.

- Jika kita ingin menentukan kejadian C yang menyatakan mata dadu prima **dan** genap, maka kita menyatakan **irisan** dari kedua kejadian, dinotasikan $A \cap B$, yaitu kejadian yang menyatakan semua elemen yang ada di kedua kejadian A dan B tersebut. Dalam kasus ini, $A \cap B = \{2\}$.

- Jika kita ingin menentukan kejadian C yang menyatakan mata dadu prima **atau** genap, maka kita menyatakan **gabungan** dari kedua kejadian, dinotasikan $A \cup B$, yaitu kejadian yang menyatakan semua elemen yang ada di salah satu kejadian A dan B tersebut, atau keduanya. Dalam kasus ini, $A \cup B = \{2, 3, 4, 5, 6\}$.

Catatan: pendefinisian himpunan $\mathcal{S}$ dengan mengubah semua kejadian (banyaknya mata dalam dadu) menjadi suatu bilangan di himpunan $\mathcal{S}$ untuk mempermudah penulisan (Akan dipelajari di bab selanjutnya).

Beberapa istilah terkait gabungan dan irisan dari dua buah kejadian:

- Dua kejadian dikatakan **saling bebas** (*independent*) jika terjadinya kejadian yang satu tidak mempengaruhi kemungkinan terjadinya kejadian yang lain. Contoh: ketika melempar koin dua kali, hasil lemparan pertama tidak mempengaruhi hasil lemparan kedua. Dalam notasi himpunan, dua kejadian dikatakan saling bebas jika $A \cap B = \emptyset$ (himpunan kosong).

- Dua kejadian dikatakan **saling lepas** jika kedua kejadian tersebut tidak dapat terjadi secara bersamaan. Contoh: ketika melempar sekeping koin, kejadian 'mendapat kepala' dan kejadian 'mendapat ekor' adalah saling terpisah, sebab keduanya tidak mungkin terjadi secara bersamaan.

Salah satu cara untuk mengidentifikasi apakah kedua kejadian itu saling bebas adalah dengan membuat *diagram Venn* (tidak dibahas di sini). Beberapa sifat terkait irisan dan gabungan berikut dapat dibuktikan dengan mudah melalui diagram Venn:

- $A \cap \emptyset = \emptyset$
- $A \cup \emptyset = A$
- $A \cap A' = \emptyset$
- $A \cup A' = \mathcal{S}$
- $\mathcal{S}' = \emptyset$
- $\emptyset' = \mathcal{S}$
- $(A')' = A$
- $(A \cap B)' = A' \cup B'$
- $(A \cup B)' = A' \cap B'$

2.2 *Aturan Pencacahan*

Berfokus pada ruang sampel diskrit, kita akan belajar menghitung berapa banyak elemen pada ruang sampel tersebut dengan mempelajari aturan pencacahan. Pertama-tama kita memulai dengan aturan pencacahan dasar:

- Jika suatu prosedur terdiri dari barisan pekerjaan $T_1, T_2, \cdots, T_m$ yang dapat dilakukan dalam $n_1, n_2, \cdots, n_m$ cara, secara berurutan, maka terdapat $n_1 \cdot n_2 \cdot \cdots \cdot n_m$ cara untuk melaksanakan prosedur tersebut.

- Jika suatu prosedur terdiri dari barisan pekerjaan $T_1, T_2, \cdots, T_m$ yang dapat dilakukan dalam $n_1, n_2, \cdots, n_m$ cara dan tidak ada dua di antara pekerjaan tersebut yang dapat dilakukan dalam waktu yang bersamaan, maka terdapat $n_1 + n_2 + \cdots + n_m$ cara untuk melakukan salah satu pekerjaan tersebut.

Selanjutnya, kita akan mengenal permutasi dan kombinasi. **Permutasi** adalah pengaturan dari semua elemen pada himpunan yang memperhatikan urutan. Dari definisi tersebut, kita ketahui bahwa banyaknya permutasi dari n objek adalah $n(n-1)(n-2)\cdots 2 \cdot 1$, atau dapat dituliskan sebagai n faktorial, dinotasikan $n!$. Didefinisikan $0! = 1$.

Jika kita perumum, banyaknya permutasi dari n objek berbeda diambil r adalah:

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

Sebagai kasus khusus, banyaknya permutasi n objek bila disusun membentuk sebuah siklis adalah $(n-1)!$.

Kasus khusus selanjutnya, banyaknya permutasi n buah objek di mana n_1 adalah jenis pertama, n_2 adalah jenis kedua, $\cdots$, dan n_k adalah jenis ke-k adalah:

$$\frac{n!}{n_1!n_2!\cdots n_k!}$$

Contoh Soal 1. Dalam sebuah latihan sepak bola di sebuah universitas, seorang pelatih mengatur barisan dari 10 orang terlebih dahulu. Sepuluh orang tersebut di antaranya: 1 orang tingkat satu, 2 orang tingkat dua, 4 orang tingkat tiga, dan 3 orang tingkat empat. Berapa banyak cara penyusunan barisan jika hanya tingkatannya yang dapat dibedakan?
Solusi. Dengan menggunakan permutasi dengan banyak jenis, kita dapatkan:

$$\frac{10!}{1!2!4!3!} = 12600 \text{ cara}$$

■

Permasalahan selanjutnya adalah permasalahan distribusi objek yang dapat dibedakan ke dalam kotak yang dapat dibedakan. Banyaknya cara untuk mendistribusikan n objek yang dapat dibedakan ke dalam r kotak yang dapat dibedakan sehingga n_i buah objek ditempatkan ke dalam kotak i (dengan $i = 1, 2, \cdots, r$) adalah:

$$\binom{n}{n_1, n_2, \cdots, n_r} = \frac{n!}{n_1!n_2!\cdots n_r!}$$

dengan $n_1 + n_2 + \cdots + n_r = n$.

Dari informasi tersebut, kita dapat menghitung banyaknya cara untuk memilih r objek dari n objek tanpa memperhatikan urutan, yang kita sebut dengan **kombinasi**. Kombinasi dapat dianalogikan sebagai pendistribusian n objek ke dalam dua kotak, yaitu r objek dan $(n-r)$ objek. Banyaknya cara tersebut adalah:

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

Contoh Soal 2. Berapa banyak permutasi huruf yang berbeda dari kata STATISTIKA?
Solusi. Gunakan distribusi objek yang dapat dibedakan (dalam hal ini semua huruf) ke dalam kotak yang dapat dibedakan (dalam hal ini huruf S, T, A, I, K).

$$\binom{10}{2, 3, 2, 2, 1} = \frac{10!}{2!3!2!2!1!} = 75600 \text{ cara}$$

■

2.3 *Peluang*

Dalam subbab sebelumnya, kita sudah mempelajari mengenai aturan pencacahan, yang tidak lain adalah kardinalitas dari kejadian di ruang sampel diskrit. Kardinalitas dari suatu himpunan A, dinotasikan $|A|$ atau $n(A)$, adalah banyaknya elemen pada sebuah himpunan. Informasi mengenai kardinalitas tersebut yang dapat digunakan untuk menghitung peluang.

Misalkan $\mathcal{S}$ adalah ruang sampel hingga yang kemungkinan terjadinya setiap keluaran sama dan E adalah kejadian yang merupakan himpunan bagian dari $\mathcal{S}$, maka peluang kejadian E adalah:

$$P(E) = \frac{|E|}{|\mathcal{S}|}$$

Dengan mengetahui bahwa E adalah subset dari $\mathcal{S}$, maka $|E| \leq |\mathcal{S}|$, sehingga:

$$0 \leq P(E) \leq 1$$

$P(E) = 0$ diartikan sebagai kejadian yang tidak pernah terjadi (dalam hal ini $E = \emptyset$) dan $P(E) = 1$ diartikan sebagai kejadian yang selalu terjadi apapun eksperimen yang dilakukan (dalam hal ini $E = \mathcal{S}$).

Ketika membicarakan mengenai himpunan, maka hal yang penting untuk diketahui adalah operasi pada himpunan, dalam hal ini akan dibahas megenai prinsip inklusi eksklusi. Misalkan A dan B adalah dua kejadian, maka:

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

Hal tersebut dengan mudah dapat dibuktikan dengan diagram Venn.

Akibat dari prinsip tersebut untuk kejadian yang saling lepas adalah:

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

dan hal tersebut analog untuk kejadian yang jamak (dengan catatan tiap kejadian saling lepas).

Salah satu contoh kejadian yang saling lepas adalah kejadian A dengan komplemennya, yaitu A'. Perhatikan bahwa:

$$P(A \cup A') = P(\mathcal{S}) = P(A) + P(A') = 1 \Leftrightarrow P(A') = 1 - P(A)$$

Prinsip inklusi eksklusi dapat digeneralisasi untuk 3 kejadian atau lebih dengan menggunakan diagram Venn.

2.4 *Peluang Bersyarat dan Aturan Bayes*

Peluang kejadian B yang terjadi ketika kejadian A terjadi disebut **peluang bersyarat** dan dinotasikan $P(B|A)$. Dalam peluang bersyarat $P(B|A)$, kejadian A adalah kejadian yang terjadi terlebih dahulu, lalu kejadian B. Besar peluangnya adalah:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \text{ dengan } P(A) > 0$$

Dalam peluang bersyarat, kejadian A dapat kita misalkan sebagai ruang sampel yang memuat kejadian B di dalamnya.

Contoh Soal 3. Diberikan tabel jenis rambut sebagai berikut:

Jenis Rambut	Warna	
	Hitam	Tidak Hitam
Lurus	2	0
Ikal	2	4
Keriting	1	2

Berapakah peluang rambut lurus di antara rambut hitam?

Solusi.

$$P(\text{Lurus}|\text{Hitam}) = \frac{P(\text{Lurus} \cap \text{Hitam})}{P(\text{Hitam})} = \frac{2/11}{5/11} = \frac{2}{5}$$

■

Sekarang kita akan melihat kasus jika A dan B saling bebas. Dua kejadian A dan B akan saling bebas jika dan hanya jika:

$$P(B|A) = P(B) \text{ atau } P(A|B) = P(A)$$

Perhatikan bahwa peluang bersyarat dapat dituliskan sebagai berikut:

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

Jika A dan B saling bebas maka:

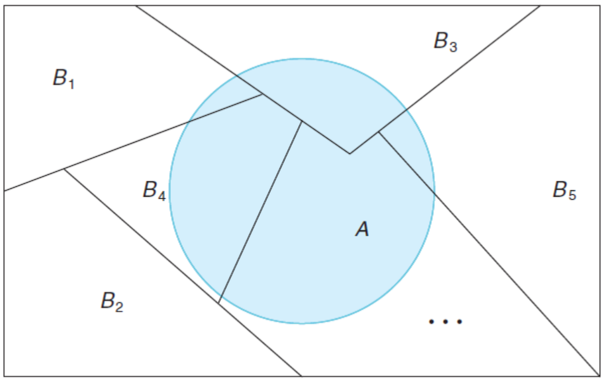
$$P(A \cap B) = P(A)P(B)$$

Persamaan tersebut dapat diperumum untuk kejadian jamak (jika semua kejadian saling bebas).

Bentuk perumuman dari peluang kondisional/peluang bersyarat adalah Aturan Bayes. Misalkan ruang sampel $\mathcal{S}$ dipartisi menjadi himpunan-himpunan kejadian B_i, $i = 1, 2, \cdots, n$ dengan $P(B_i) \neq 0$, atau dapat dituliskan:

$$\mathcal{S} = B_1 \cup B_2 \cup \cdots \cup B_n = \bigcup_{i=1}^{n} B_i$$

dengan $B_i \cap B_j = \emptyset$ untuk $i \neq j$, $i, j = 1, 2, \cdots, n$. Ilustrasinya dapat dilihat pada gambar 2.1.



Gambar 2.1: Diagram Venn Partisi $\mathcal{S}$

Jika A subset dari $\mathcal{S}$, maka:

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_n) = \bigcup_{i=1}^{n} (A \cap B_i)$$

Karena B_i untuk setiap i adalah saling lepas, maka:

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i)$$

Ingat kembali bahwa bentuk $P(A \cap B_i)$ dapat diubah menjadi:

$$P(A \cap B_i) = P(A|B_i)P(B_i) = P(B_i|A)P(A)$$

Teorema 1 Aturan Bayes

Misalkan $\{B_1, B_2, \cdots, B_n\}$ adalah partisi dari ruang sampel $\mathcal{S}$ dan misalkan A adalah kejadian yang terobservasi di $\mathcal{S}$. Peluang kejadian B_j diberikan A adalah:

$$P(B_j|A) = \frac{P(A \cap B_j)}{\sum_{i=1}^{n} P(A \cap B_i)}$$

Contoh Soal 5. Suatu perusahaan besar menggunakan 3 hotel sebagai tempat menginap para langganannya. Dari pengalaman yang lalu diketahui bahwa 20% langganannya ditempatkan di Hotel I, 50% di Hotel B, dan 30% di Hotel S. Bila 5% kamar mandi di Hotel i tidak berfungsi dengan baik, 4% di Hotel B, dan 8% di Hotel S, berapa peluang bahwa seseorang yang mendapat kamar mandi yang tidak baik ditempatkan di Hotel S?

Solusi. Misalkan:

- A: kejadian seorang pelanggan mendapat kamar mandi yang tidak baik.
- B_1: kejadian seorang pelanggan mendapat kamar di Hotel I.
- B_2: kejadian seorang pelanggan mendapat kamar di Hotel B.
- B_3: kejadian seorang pelanggan mendapat kamar di Hotel S.

Kita bentuk tabel terlebih dahulu untuk mempermudah:

i	1	2	3
$P(B_i)$	0.2	0.5	0.3
$P(A\|B_i)$	0.05	0.04	0.08

Pertama-tama kita hitung $P(A)$:

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) = (0.05)(0.2) + (0.04)(0.5) + (0.08)(0.3) = 0.054$$

Sehingga peluang seseorang mendapat kamar mandi yang tidak baik di Hotel S adalah:

$$P(B_3|A) = \frac{P(A|B_3)P(B_3)}{P(A)} = \frac{(0.08)(0.3)}{0.054} = \frac{4}{9}$$

■

2.5 Soal Latihan

1. Berapa banyak cara penyusunan huruf yang berbeda yang dapat dibuat dari huruf-huruf pada kata MATEMATIKA?

2. Peluang bahwa jadwal keberangkatan regular sebuah maskapai penerbangan tepat waktu adalah 0.83, peluang kedatangannya tepat waktu adalah 0.82, dan peluang bahwa keberangkatan dan kedatangannya tepat waktu adalah 0.78. Berapakah peluang bahwa kedatangannya tepat waktu jika keberangkatannya tepat waktu dan peluang bahwa keberangkatannya tepat waktu jika kedatangannya tepat waktu?

3. Seorang peneliti mencatat bahwa dalam 6 bulan (180 hari) pada musim hujan di suatu daerah di Jawa Barat, hujan turun selama 150 hari. Dalam kurun waktu yang sama, banyaknya kejadian bahwa kecepatan angin di atas batas kecepatan normal adalah 120 hari. Sedangkan banyak kejadian bahwa kecepatan angin di atas batas normal saat hari hujan adalah 108 hari. Diasumsikan bahwa kejadian hujan maupun kejadian kecepatan angin yang di atas batas normal tersebut adalah sekali dalam sehari. Dalam 6 bulan musim hujan di tahun berikutnya:

   a) Hitung berturut-turut peluang hujan dan peluang kecepatan angin di atas batas normal.

   b) Apakah kejadian hari hujan dan kecepatan angin di atas batas normal adalah saling bebas?

   c) Hitung peluang terjadinya hujan jika kecepatan angin di atas batas normal.

Bab 3

Peubah Acak

Dalam pelantunan sebuah dadu kita mengetahui bahwa nilai dari dadu secara acak dari 1 sampai 6. Pada bab ini kita akan melihat definisi dari peubah acak dan hubungannya dengan ilustrasi yang diberikan. Peubah acak yang akan dibahas adalah peubah tunggal dan jamak.

3.1 *Definisi Peubah Acak*

Misalkan dalam pelemparan dua buah koin, kita mendapatkan ruang sampel kepala-ekor, ekor-kepala, kepala-kepala, dan ekor-ekor. Misal $\mathcal{S}$ menyatakan ruang sampel pelemparan dua buah koin, maka:

$$\mathcal{S} = \{KE, EK, EE, KK\}$$

Hal ini cukup menyulitkan apabila ruang sampel bertambah besar. Salah satu solusi untuk mengatasinya dengan memetakan setiap elemen pada ruang sampel menjadi suatu bilangan real.

Definisi 1 Peubah Acak

Peubah acak adalah fungsi yang memetakan setiap elemen pada ruang sampel ke bilangan real.

Pada ilustrasi sebelumnya, kita dapat misalkan X sebagai peubah acak yang menyatakan banyaknya ekor pada hasil keluaran pelemparan dua buah koin. Maka:

Ruang Sampel	X
KE	1
EK	1
EE	2
KK	0

Peubah acak dapat dibedakan menjadi dua jenis, yaitu peubah acak diskrit dan peubah acak kontinu. Salah satu hal yang membedakan kedua peubah acak tersebut adalah melalui fungsi peluangnya.

3.2 *Peubah Acak Diskrit*

Pada contoh ilustrasi sebelumnya, kita memetakan hasil pelemparan dua buah koin dengan bilangan bulat. Karena ruang sampelnya *countable*, maka lustrasi tersebut merupakan contoh dari peubah acak diskrit.

Definisi 2 Fungsi Massa Peluang

Pasangan terurut $(x, f(x))$ disebut fungsi massa peluang dari peubah acak diskrit X jika untuk setiap keluaran x:

1. $f(x) \geq 0$

2. $\sum_{x} f(x) = 1$
3. $P(X = x) = f(x)$

Contoh Soal 1. Tunjukkan bahwa:

$$f(x) = \begin{cases} \frac{1}{16}\binom{4}{x} & \text{untuk } x = 0, 1, 2, 3, 4 \\ 0 & \text{untuk } x \text{ lainnya} \end{cases}$$

merupakan fungsi massa peluang.
Solusi. Kita cek terlebih dahulu nilai fungsi massa peluang tiap x:

$$f(0) = \frac{1}{16} \qquad f(1) = \frac{1}{4} \qquad f(2) = \frac{3}{8} \qquad f(3) = \frac{1}{4} \qquad f(4) = \frac{1}{16}$$

Karena nilai $f(x)$ setiap nilai x tidak ada yang di bawah nol, maka syarat pertama terpenuhi. Syarat kedua yang akan kita cek adalah penjumlahan semua nilai $f(x)$. Dapat diuji bahwa penjumlahan nilai $f(x)$ sama dengan 1 sehingga dapat disimpulkan bahwa $f(x)$ merupakan fungsi massa peluang. ■

Satu lagi yang akan kita definisikan mengenai fungsi kepadatan massa adalah fungsi distribusi kumulatif.

Definisi 3 Fungsi Distribusi Kumulatif untuk Peubah Acak Diskrit

Fungsi distribusi kumulatif dari $f(x)$ (dinotasikan $F(x)$) dari peubah acak diskrit X adalah:

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t) \qquad -\infty < x < \infty$$

Contoh Soal 2. Tentukan fungsi distribusi kumulatif pada fungsi massa peluang pada contoh soal 1 dan tunjukkan bahwa $f(2) = \frac{3}{8}$ menggunakan fungsi kumulatif tersebut.
Solusi. Kita tentukan nilai-nilai $F(x)$ untuk $x = 0, 1, 2, 3, 4$.

$$F(0) = f(0) = \frac{1}{16}$$

$$F(1) = f(0) + f(1) = \frac{5}{16}$$

$$F(2) = f(0) + f(1) + f(2) = \frac{11}{16}$$

$$F(3) = f(0) + f(1) + f(2) + f(3) = \frac{15}{16}$$

$$F(4) = f(0) + f(1) + f(2) + f(3) + f(4) = 1$$

Jadi, kita dapatkan fungsi distribusi kumulatif nya:

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{16} & 0 \leq x < 1 \\ \frac{5}{16} & 1 \leq x < 2 \\ \frac{11}{16} & 2 \leq x < 3 \\ \frac{15}{16} & 3 \leq x < 4 \\ 1 & x \geq 4 \end{cases}$$

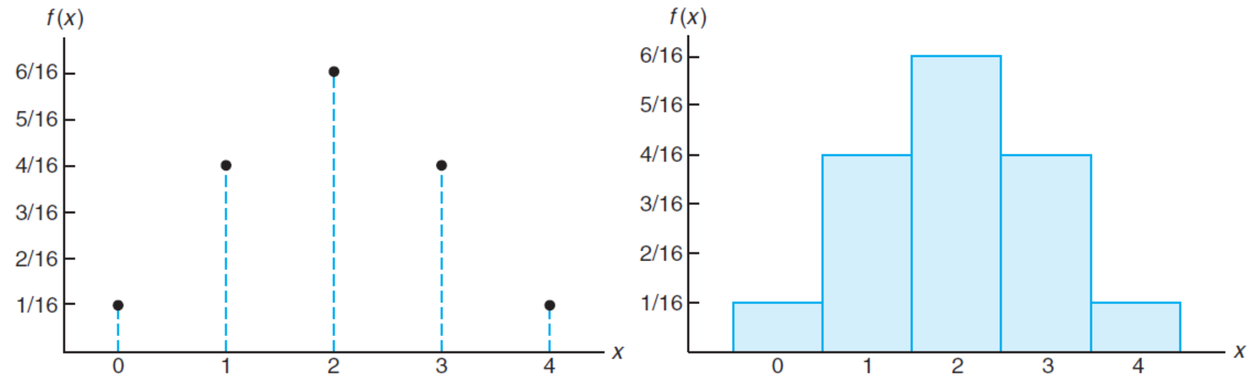
Untuk menghitung $f(2)$ kita dapat gunakan persamaan:

$$f(2) = F(2) - F(1) = \frac{11}{16} - \frac{5}{16} = \frac{3}{8}$$

■

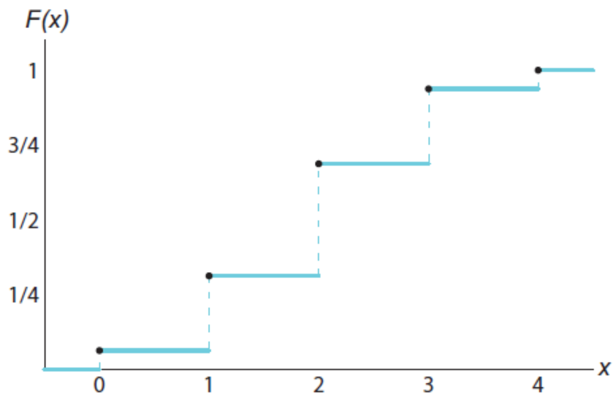
Contoh Soal 3. Gambarkan grafik fungsi massa peluang dan fungsi distribusi kumulatif untuk peubah acak contoh soal 1.

Solusi. Kita dapat membuat dua jenis grafik fungsi untuk fungsi massa peluang, pertama dengan titik dan kedua dengan histogram. Karena peubah acak bersifat diskrit, maka nilai $f(x)$ hanya terdefinisi pada suatu titik pada x, bukan menjadi selang.



Gambar 3.1: Grafik Fungsi Massa Peluang ($f(x)$) dalam Bentuk Diagram Titik (kiri) dan Histogram (kanan)

Selanjutnya untuk grafik fungsi distribusi kumulatif, kita ketahui bahwa $F(x)$ akan selalu monoton naik dari 0 ke 1 seiring bertambahnya nilai x dan nilainya **akan berubah** ketika $f(x) \neq 0$. Sehingga grafik fungsinya tersaji pada gambar 3.2.



Gambar 3.2: Grafik Fungsi Distribusi Kumulatif ($F(x)$)

■

3.3 *Peubah Acak Kontinu*

Setelah kita membahas semua permasalahan peubah acak dari ruang sampel yang elemennya *countable*, sekarang kita akan membahas ruang sampel yang terukur, yaitu kontinu. Pada peubah acak kontinu, semua nilai x pada sebuah selang akan mendapatkan suatu nilai. Definisi dari fungsi massa peluang akan secara analog kita gunakan untuk peubah acak kontinu, dengan mengganti perhitungan diskrit (notasi sigma) dengan perhitungan kontinu (integral tentu).

Definisi 4 Fungsi Kepadatan Peluang

Fungsi $f(x)$ disebut fungsi massa peluang dari peubah acak kontinu X jika untuk setiap keluaran x:

1. $f(x) \geq 0$ untuk semua $x \in \mathbb{R}$
2. $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = 1$
3. $P(a < X < b) = \displaystyle\int_a^b f(x)\,dx$

Hal yang serupa terjadi untuk fungsi distribusi kumulatif.

Definisi 5 Fungsi Distribusi Kumulatif untuk Peubah Acak Kontinu
Fungsi distribusi kumulatif dari $f(x)$ (dinotasikan $F(x)$) dari peubah acak kontinu X adalah:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt \qquad -\infty < x < \infty$$

Dari definisi tersebut kita dapat ketahui bahwa:

$$P(a < X < b) = F(b) - F(a) \text{ dan } f(x) = \frac{dF(x)}{dx}$$

Contoh Soal 4. Diberikan fungsi kepadatan peluang sebagai berikut:

$$f(x) = \begin{cases} \frac{x^2}{3} & -1 < x < 2 \\ 0 & x \text{ lainnya} \end{cases}$$

Tunjukkan terlebih dahulu bahwa $f(x)$ adalah fungsi kepadatan peluang, lalu tentukan $P(0 < X \leq 1)$ dengan menenntukan fungsi distribusi kumulatifnya terlebih dahulu. Solusi. Karena fungsi termasuk definit positif pada interval yang diberikan, maka syarat pertama terpenuhi. Untuk syarat kedua:

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-1}^{2} \frac{x^2}{3}\,dx = 1$$

Fungsi distribusi kumulatif untuk $f(x)$ adalah:

$$F(x) = \int_{-1}^{x} \frac{t^2}{3}\,dt = \frac{x^3 + 1}{9}$$

Sehingga:

$$P(0 < X \leq 1) = F(1) - F(0) = \frac{2}{9} - \frac{1}{9} = \frac{1}{9}$$

■

Perlu diperhatikan: Ada atau tidaknya notasi sama dengan pada perhitungan peubah acak kontinu tidak berpengaruh (secara signifikan), maka dapat diasumsikan sama saja pada perhitungan.

3.4 *Fungsi Peluang Gabungan*

Setelah membahas peubah acak dan berbagai fungsi peluangnya, kita akan membahas fungsi yang melibatkan dua peubah acak sekaligus, yang disebut dengan fungsi peluang gabungan. Seperti biasa, fungsi peluang gabungan pun dibedakan menjadi diskrit dan kontinu.

Definisi 6 Fungsi Massa Peluang Gabungan

Fungsi $f(x,y)$ disebut fungsi massa peluang gabungan untuk peubah acak diskrit X dan Y jika:

- $f(x,y) \geq 0$ untuk semua (x,y)
- $\sum_x \sum_y f(x,y) = 1$
- $P(X = x, Y = y) = f(x,y)$
- Untuk setiap daerah A di bidang xy, $P[(X,Y) \in A] = \sum \sum_A f(x,y)$

Contoh Soal 4. Diberikan fungsi massa peluang gabungan sebagai berikut:

$$f(x,y) = \begin{cases} f(x,y) = \frac{\binom{3}{x}\binom{2}{y}\binom{3}{2-x-y}}{\binom{8}{2}} & x = 0,1,2;\ y = 0,1,2;\ 0 \leq x+y \leq 2 \\ 0 & (x,y) \text{ lainnya} \end{cases}$$

Daftarkan semua $f(x,y)$ untuk semua pasangan terurut (x,y) yang memenuhi syarat dalam bentuk tabel dan tentukan $P[(X,Y) \in A]$ dengan $A = \{(x,y)|x+y \leq 1\}$.

Solusi. Tabel yang terbentuk adalah sebagai berikut:

$f(x,y)$		x = 0	1	2	Jumlah
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
Jumlah		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

Terbukti bahwa $f(x,y)$ merupakan fungsi massa peluang. Selanjutnya akan dicari $P[(X,Y) \in A]$.

$$\begin{aligned} P[(X,Y) \in A] &= P(X + Y \leq 1) \\ &= f(0,0) + f(0,1) + f(1,0) \\ &= \frac{3}{28} + \frac{3}{14} + \frac{9}{28} = \frac{9}{14} \end{aligned}$$

■

Selanjutnya kita lihat fungsi kepadatan peluang gabungan.

Definisi 7 Fungsi Kepadatan Peluang Gabungan

Fungsi $f(x,y)$ disebut fungsi kepadatan peluang gabungan untuk peubah acak kontinu X dan Y jika:

- $f(x,y) \geq 0$ untuk semua (x,y)
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\, dx\, dy = 1$

- Untuk setiap daerah A di bidang xy, $P[(X,Y) \in A] = \iint_A f(x,y)\, dA$

Contoh Soal 5. Diberikan fungsi kepadatan peluang gabungan sebagai berikut:

$$f(x,y) = \begin{cases} \frac{2}{5}(2x+3y) & 0 \le x \le 1, 0 \le y \le 1 \\ 0 & (x,y) \text{ lainnya} \end{cases}$$

Tentukan $P[(X,Y) \in A]$ jika $A = \{(x,y) | 0 < x < \frac{1}{2}, \frac{1}{4} < y < \frac{1}{2}\}$.
Solusi.

$$\begin{aligned} P\left(0 < X < \frac{1}{2}, \frac{1}{4} < Y < \frac{1}{2}\right) &= \int_{1/4}^{1/2} \int_0^{1/2} \frac{2}{5}(2x+3y)\, dx\, dy \\ &= \int_{1/4}^{1/2} \left(\frac{1}{10} + \frac{3y}{5}\right) dy \\ &= \frac{13}{160} \end{aligned}$$

■

3.5 *Peluang Bersyarat*

Karena kita melibatkan lebih dari satu peubah acak, maka lebih dari satu ruang sampel pun terlibat. Maka dari itu, kita dapat mengecek sifat saling bebas secara statistik (bebas statisik) dari kedua ruang sampel tersebut dengan menggunakan peluang bersyarat.

Sebelumnya, kita definisikan suatu perhitungan yang akan membantu kita dalam menentukan peluang bersyarat, yaitu distribusi marjinal.

Definisi 8 Distribusi Marjinal
Distribusi marjinal untuk masing-masing peubah acak diskrit X dan Y adalah:

$$g(x) = \sum_y f(x,y) \qquad h(y) = \sum_x f(x,y)$$

Distribusi marjinal untuk masing-masing peubah acak kontinu X dan Y adalah:

$$g(x) = \int_{-\infty}^{\infty} f(x,y)\, dy \qquad h(y) = \int_{-\infty}^{\infty} f(x,y)\, dx$$

Sebagai contoh, tabel pada contoh soal 4 sudah menyatakan distribusi marjinal di dalamnya. Baris dan kolom yang menyatakan "jumlah" berturut-turut menyatakan distribusi marjinal untuk peubah acak X dan Y.

Selanjutnya, kita definisikan peluang bersyarat bagi peubah acak.

Definisi 9 Peluang Bersyarat
Misal X dan Y dua buah peubah acak (diskrit atau kontinu), dengan fungsi peluang gabungan $f(x,y)$

dan fungsi marjinal masing-masingnya $g(x)$ dan $h(y)$. Peluang bersyarat dari peubah acak Y diberikan $X = x$ adalah:

$$f(y|x) = \frac{f(x, y)}{g(x)} \text{ untuk } g(x) > 0$$

Sebaliknya, peluang bersyarat dari peubah acak X diberikan $Y = y$ adalah:

$$f(x|y) = \frac{f(x, y)}{h(y)} \text{ untuk } h(y) > 0$$

Definisi berikut dapat diperumum untuk peubah acak lebih dari dua (mirip seperti Aturan Bayes).

Sifat yang serupa pun terjadi ketika kita membahas dua kejadian yang saling bebas bila meninjau dari peluang bersyaratnya.

Definisi 10 Bebas Statistik

Misal X dan Y dua buah peubah acak (diskrit atau kontinu), dengan fungsi peluang gabungan $f(x, y)$ dan fungsi marjinal masing-masingnya $g(x)$ dan $h(y)$. Dua peubah acak tersebut dikatakan bebas statistik jika dan hanya jika:

$$f(x, y) = g(x)h(y)$$

untuk semua $f(x, y)$ di range yang diberikan.

3.6 *Soal Latihan*

1. Diketahui suatu peubah acak X mempunyai fungsi distribusi peluang sebagai berikut:

$$F(x) = \begin{cases} 0 & x < 0 \\ 0.5 & 0 \leq x < 1 \\ 0.75 & 1 \leq x < 2 \\ 0.87 & 2 \leq x < 3 \\ 0.95 & 3 \leq x < 4 \\ 1 & x \geq 4 \end{cases}$$

   a) Gambarkan fungsi distribusi dari peubah acak X.
   b) Tentukan fungsi peluang peubah acak X.
   c) Gambarkan fungsi peluang peubah acak X.

2. Lamanya pemakaian sparepart (dalam bulan) dari suatu jenis motor Jepang, misalkan W, dianggap mempunyai fungsi distribusi peluang sebagai berikut

$$f(w) = \begin{cases} Kwe^{-w/2} & w > 0 \\ 0 & w \leq 0 \end{cases}$$

   a) Berapakah nilai K?
   b) Hitung peluang sparepart tersebut sudah terpakai paling sedikit 5 bulan.

Bab 4

Ekspektasi

Nilai dari sebuah fungsi peluang yang didapat dari peubah acak yang bersangkutan tentu beragam. Namun kita tentu mempunyai nilai yang seharusnya didapat. Dalam bab ini kita akan mempelajari mengenai ekspektasi dari sebuah fungsi peluang serta variansi yang dihasilkan dari suatu fungsi peluang. Dalam fungsi berpeubah acak jamak kita akan melihat padanan dari variansi yang disebut dengan kovariansi dan korelasi.

4.1 Ekspektasi dan Sifatnya

Pada bab sebelumnya kita mengetahui bahwa peubah acak adalah suatu pemetaan dari tiap elemen pada ruang sampel ke sebuah bilangan real. Karena kita membicarakan mengenai bilangan, maka kita dapat mengetahui nilai yang diharapkan (*ekspektasi*) atau biasa kita kenal dengan rataan.

Definisi 1 Ekspektasi

Misal X peubah acak dengan fungsi peluang $f(x)$. Ekspektasi dari X:

$$\mu = E(X) = \begin{cases} \sum_x x f(x) & X \text{ diskrit} \\ \int_{-\infty}^{\infty} x f(x)\, dx & X \text{ kontinu} \end{cases}$$

Definisi ini secara analog berlaku untuk dua peubah acak atau lebih.

Perlu diperhatikan: notasi $\bar{x}$ dan μ pada rataan mengartikan sesuatu yang berbeda. Notasi $\bar{x}$ menyatakan rataan sampel sedangkan notasi μ menyatakan rataan populasi.

Contoh Soal 1. Diberikan fungsi peluang:

$$f(x) = \begin{cases} \frac{\binom{4}{x}\binom{3}{3-x}}{\binom{7}{x}} & x = 0, 1, 2, 3 \\ 0 & \text{lainnya} \end{cases}$$

Tentukan nilai ekspektasi bagi $f(x)$.

Solusi. Kita dapatkan:

$$f(0) = \frac{1}{35} \qquad f(1) = \frac{12}{35} \qquad f(2) = \frac{18}{35} \qquad f(3) = \frac{4}{35}$$

Sehingga:

$$\mu = E(X) = (0)\left(\frac{1}{35}\right) + (1)\left(\frac{12}{35}\right) + (2)\left(\frac{18}{35}\right) + (3)\left(\frac{4}{35}\right) = \frac{12}{7}$$

■

Selanjutnya kita lihat sifat-sifat dari ekspektasi:

Teorema 1 Sifat Ekspektasi

- Untuk peubah acak X dan Y:

$$E(X+Y) = E(X) + E(Y)$$

- Untuk a dan b sebuah konstanta:

$$E(aX+b) = aE(X) + b$$

- Untuk $g(x)$ sebuah fungsi:

$$E(g(x)) = \begin{cases} \sum_x g(x)P(X=x) & X \text{ diskrit} \\ \int_{-\infty}^{\infty} g(x)f(x)\,dx & X \text{ kontinu} \end{cases}$$

Catatan: Untuk $g(x) = x^p$ maka $E(X^p)$ disebut momen ke-p. Untuk $g(x) = (x-a)^p$ maka $E((x-a)^p)$ disebut momen ke-p di sekitar a.

Secara umum kita dapat simpulkan bahwa ekspektasi memiliki sifat kelinieran.

4.2 *Variansi dan Sifatnya*

Setelah kita mengetahui mengenai nilai harapan dan sifat-sifatnya, kita dapat menentukan keberagaman suatu peubah acak melalui variansi.

Definisi 2 Variansi

Misal X peubah acak dengan fungsi peluang $f(x)$ dan rataan μ. Variansi dari X adalah:

$$\text{Var}(X) = \sigma^2 = E[(X-\mu)^2] = \begin{cases} \sum_x (x-\mu)^2 P(X=x) & X \text{ diskrit} \\ \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx & X \text{ kontinu} \end{cases}$$

Nilai $(x-\mu)$ disebut deviasi dari sebuah observasi.

Perlu diperhatikan: sama seperti rataan, notasi s^2 menyatakan variansi sampel sedangan notasi σ^2 menyatakan variansi populasi. Hal ini harus dibedakan karena rumus perhitungan variansi sampel dengan populasi berbeda.

Selanjutnya, kita lihat teorema yang memudahkan kita untuk menghitung variansi:

Teorema 2

Variansi dari peubah acak X adalah:

$$\sigma^2 = \text{Var}(X) = E(X^2) - \mu^2$$

Bukti. Untuk kasus diskrit, melalui sifat kelinieran notasi sigma kita dapatkan:

$$\begin{aligned}\sigma^2 &= \sum_x (x-\mu)^2 f(x) = \sum_x (x^2 - 2\mu x - \mu^2) f(x) \\ &= \sum_x x^2 f(x) - 2\mu \sum_x x f(x) + \mu^2 \sum_x f(x)\end{aligned}$$

Karena $\mu = \sum_x x f(x)$ dan $\sum_x f(x) = 1$ untuk semua fungsi massa peluang, kita dapatkan:

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

Hal yang sama terjadi untuk kasus kontinu karena operasi integral tentu juga memiliki sifat kelinieran. ■

Contoh Soal 2. Diberikan fungsi kepadatan peluang untuk peubah acak X sebagai berikut:

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

Tentukan mean dan variansi untuk X.

Solusi. Kita cari meannya terlebih dahulu:

$$\mu = E(X) = 2\int_1^2 x(x-1)\,dx = \frac{5}{3}$$

Selanjutnya kita cari $E(X^2)$:

$$E(X^2) = 2\int_1^3 x^2(x-1)\,dx = \frac{17}{6}$$

Sehingga variansinya:

$$\text{Var}(X) = \frac{17}{6} - \left(\frac{5}{3}\right)^2 = \frac{1}{18}$$

■

Beberapa sifat yang dimiliki oleh variansi adalah sebagai berikut:

Teorema 3 Sifat Variansi

- Untuk a dan b sebuah konstanta:

$$\text{Var}(aX + b) = a^2 \cdot \text{Var}(X)$$

- Untuk $g(x)$ sebuah fungsi:

$$\text{Var}(g(x)) = \begin{cases} \sum_x (g(x)-\mu)^2 P(X = x) & X \text{ diskrit} \\ \int_{-\infty}^{\infty} (g(x)-\mu)^2 f(x)\,dx & X \text{ kontinu} \end{cases}$$

Bukti teorema 3 dengan mudah dapat diperoleh dengan menggunakan sifat ekspektasi.

4.3 *Kovariansi dan Korelasi*

Selanjutnya kita akan membicarakan mengenai derajat keacakan peubah acak bivariat (peubah acak gabungan) melalui kovariansi.

Definisi 3 Kovariansi
Misalkan X dan Y adalah peubah acak dengan fungsi peluang gabungan $f(x, y)$. Kovariansi dari X dan Y adalah:

$$\sigma_{XY} = \text{Cov}(X, Y) = E[(X - \mu_x)(Y - \mu_Y)]$$

Dalam perhitungan, kita gunakan teorema berikut:

Teorema 4
Kovariansi dari peubah acak X dan Y adalah:

$$\sigma_{XY} = E(XY) - E(X)E(Y)$$

dengan $E(X)$ dan $E(Y)$ masing-masing merupakan ekspektasi fungsi marjinal masing-masing peubah acak.

Bukti. Untuk kasus diskrit, kita gunakan sifat kelinieran sehingga didapat:

$$\begin{aligned}\sigma_{XY} &= \sum_x \sum_y (x - \mu_X)(y - \mu_Y) f(x, y) \\ &= \sum_x \sum_y xyf(x, y) - \mu_X \sum_x \sum_y yf(x, y) - \mu_Y \sum_x \sum_y xf(x, y) + \mu_X \mu_Y \sum_x \sum_y f(x, y)\end{aligned}$$

Karena untuk semua fungsi peluang gabungan berlaku:

$$\mu_X = \sum_x xf(x, y) \qquad \mu_Y = \sum_y yf(x, y) \qquad \sum_x \sum_y f(x, y) = 1$$

Jadi:

$$\sum_{XY} = E(XY) - \mu_X\mu_Y - \mu_Y\mu_X + \mu_X\mu_Y = E(XY) - \mu_X\mu_Y$$

∎

Perhatikan jika $Y = X$ maka $\text{Cov}(X, X) = E(X^2) - [E(X)]^2$ =Var(X). Sehingga kita mendapat hubungan kovariansi dan variansi sebagai berikut:

$$\text{Cov}(X, X) = \text{Var}(X)$$

Contoh Soal 3. Perhatikan kembali contoh soal 4 pada bab 3. Tentukan kovariansi dari peubah acak X dan Y.
Solusi. Pertama-tama kita cari terlebih dahulu nilai $E(XY)$:

$$\begin{aligned}E(XY) &= \sum_{x=0}^{2} \sum_{y=0}^{2} xyf(x, y) \\ &= (0)(0)f(0, 0) + (1)(0)f(1, 0) + (0)(1)f(0, 1) + (1)(1)f(1, 1) + (2)(0)f(2, 0) = f(1, 1) = \frac{3}{14}\end{aligned}$$

Selanjutnya kita tentukan ekspektasi masing-masing fungsi marjinal:

$$\mu_X = \sum_{x=0}^{2} xg(x) = (0)\left(\frac{5}{14}\right) + (1)\left(\frac{15}{28}\right) + (2)\left(\frac{3}{28}\right) = \frac{3}{4}$$

$$\mu_Y = \sum_{y=0}^{2} yh(y) = (0)\left(\frac{15}{28}\right) + (1)\left(\frac{3}{7}\right) + (2)\left(\frac{1}{28}\right) = \frac{1}{2}$$

Jadi kovariansi yang didapat:

$$\sigma_{XY} = \frac{3}{14} - \frac{3}{4}\cdot\frac{1}{2} = -\frac{9}{56}$$

■

Setelah mempelajari mengenai korelasi, kita dapat melihat hubungan linier dari dua buah peubah acak melalui sebuah paramaeter yang kita sebut korelasi.

Definisi 4 Koefisien Korelasi Pearson
Misal X dan Y peubah acak dengan kovariansi σ_{XY} dan standar deviasi (akar dari variansi) σ_x dan σ_Y. Korelasi dari koefisien X dan Y adalah:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Perlu diketahui bahwa nilai korelasi berkisar pada interval $-1 \leq \rho_{XY} \leq 1$. Korelasi pada dasarnya memberikan informasi mengenai hubungan linier antar dua peubah. Misalkan hubungan linier tersebut $Y = a + bX$ untuk dua peubah X dan Y. Jika korelasi semakin mendekat ke 1, maka nilai konstanta b akan positif karena terjadi hubungan linier yang berbanding lurus. Sebaliknya, jika korelasi semakin mendekat ke -1, maka nilai konstanta b akan negatif karena terjadi hubungan linier yang berbanding terbalik. Peubah acak X dan Y dianggap tidak berhubungan apabila korelasinya mendekati nol (atau kovariansinya mendekati nol).

Contoh Soal 4. Perhatikan kembali contoh soal 4 pada bab 3. Tentukan koefisien korelasi dari peubah acak X dan Y.

Solusi. Dapat dicek bahwa:

$$\sigma_X^2 = \frac{27}{28} - \left(\frac{3}{4}\right)^2 = \frac{45}{112} \qquad \sigma_Y^2 = \frac{4}{7} - \left(\frac{1}{2}\right)^2 = \frac{9}{28}$$

Jadi korelasinya:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{-9/56}{\sqrt{(45/112)(9/28)}} = -\frac{1}{\sqrt{5}}$$

■

4.4 *Soal Latihan*

1. Misalkan X adalah variabel acak dengan fungsi kepadatan/kerapatan peluang sebagai berikut:

$$f(x) = \begin{cases} x+1 & -1 < x < 0 \\ 1-x & 0 \leq x < 1 \\ 0 & x \text{ lainnya} \end{cases}$$

   Tentukan mean dan variansi dari variabel acak X.

2. Misalkan kita melakukan survey mengenai waktu hidup suatu produk elektronik. Waktu hidup alat elektronik ini dinyatakan oleh peubah acak (*random variable*) kontinu X. Dari hasil pengamatan diperoleh fungsi padat peluangnya (*probability density function*) berbentuk:

$$f_x(x) = \begin{cases} k \cdot \exp\left(-\frac{1}{2}x\right) & x \geq 0 \\ 0 & x < 0 \end{cases}$$

   dengan k sebagai bilangan yang masih harus ditetapkan.

   a) Tentukan nilai k.
   b) Hitunglah $\mathrm{Var}\left(\frac{1}{2}x - 1\right)$.

3. Mengacu pada soal nomor 2 bab 3, hitung ekspektasi dan simpangan baku dari lamanya pemakaian sparepart tersebut.

Bab 5

Distribusi Peluang Diskrit

Setelah kita mempelajari peluang, jenis-jenisnya, peubah acak, ekspektasi, serta variansi, kita akan lihat bermacam-macam distribusi dari peluang yang bersifat diskrit serta nilai ekspektasi dan variansi yang dihasilkan. Tentu saja distribusi yang akan kita bahas berkaitan dengan peluang di kehidupan sehari-hari.

5.1 *Distribusi Uniform*

Distribusi pertama yang akan kita pelajari adalah distribusi uniform. Seperti dengan namanya, distribusi ini berisikan peubah acak X yang setiap nilainya $(x_1, x_2, \cdots, x_n)$ memiliki peluang yang sama. Distribusi peluang dari X dituliskan sebagai berikut:

$$P(X = x) = \frac{1}{k} \qquad x = x_1, x_2, \cdots, x_k$$

Teorema 1 Ekspektasi dan Variansi Distribusi Uniform

Ekspektasi dan Variansi dari Distribusi Uniform berturut-turut adalah

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

Bukti. Berdasarkan definisi dari ekspektasi dan variansi kita dapatkan

$$\mu = \sum_{i=1}^{k} x_i \cdot P(X = x_i) = \sum_{i=1}^{k} \frac{x_i}{k} = \frac{1}{k}\sum_{i=1}^{k} x_i$$

$$\sigma^2 = \sum_{i=1}^{k}(x_i - \mu)^2 \cdot P(X = x_i) = \frac{1}{k}\sum_{i=1}^{k}(x_i - \mu)^2$$

■

Contoh Soal 1. Dalam pelantunan sebuah dadu, kita mengetahui bahwa peluang munculnya mata dadu 1 hingga 6 adalah sama, yaitu $\frac{1}{6}$. Tentukan ekspektasi dan variansi dari pelantunan sebuah dadu.

Solusi. Misalkan X peubah acak yang menyatakan mata dadu, maka X berdistribusi uniform dengan peluang

$$P(X = x) = \frac{1}{6} \qquad x = 1, 2, 3, 4, 5, 6$$

Ekspektasi dan variansinya adalah

$$\mu = \frac{1+2+3+4+5+6}{6} = 3.5 \qquad \sigma^2 = \frac{1^2+2^2+3^2+4^2+5^2+6^2}{6} - (3.5)^2 = 2.92$$

■

5.2 *Distribusi Bernoulli*

Distribusi selanjutnya adalah mengenai peluang yang sukses atau gagal dalam suatu kejadian. Misal X menyatakan peubah acak suatu kejadian yang bernilai 0 ketika kejadian gagal dan 1 ketika kejadian sukses, serta peluang kejadian sukses adalah p. Jika kejadian tersebut terjadi hanya satu kali, maka X akan berdistribusi Bernoulli dengan peluang

$$P(X = x) = Ber(x; p) = \begin{cases} p^x(1-p)^{1-x} & x = 0, 1 \\ 0 & \text{lainnya} \end{cases}$$

Kita dapat mengecek nilai tersebut dengan cara distribusi langsung nilai x. Jika kejadian itu sukses, maka peluangnya $p^1(1-p)^{1-1} = p$ dan jika kejadian itu gagal, maka kita dapatkan peluangnya $p^0(1-p)^{1-0} = 1-p$.

Teorema 2 Ekspektasi dan Variansi Distribusi Bernoulli

Ekspektasi dan Variansi dari Distribusi Bernoulli berturut-turut adalah

$$\mu = p \qquad \sigma^2 = p(1-p)$$

Bukti. Berdasarkan definisi dari ekspektasi dan variansi kita dapatkan

$$\begin{aligned} \mu &= \sum_{i=1}^{k} x_i \cdot P(X = x_i) = 0(1-p) + 1(p) = p \\ \sigma^2 &= \sum_{i=1}^{k} (x_i - \mu)^2 \cdot P(X = x_i) = (-p)^2(1-p) + (1-p)^2(p) \\ &= p^2 - p^3 + p - 2p^2 + p^3 = p(1-p) \end{aligned}$$

■

5.3 *Distribusi Binomial*

Distribusi berikut adalah perluasan dari distribusi Bernoulli. Distribusi Binomial adalah percobaan Bernoulli sebanyak n kali dengan peluang tiap percobaan yang saling bebas. Misal X adalah peubah acak berdistribusi Binomial. Kita notasikan $X \sim B(n, p)$ dengan n jumlah usaha yang dilakukan dan p adalah peluang sukses. Peluangnya

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

Dengan $\binom{n}{x}$ disebut koefisien binomial

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} \text{ untuk } x = 0, 1, \cdots, n$$

Nama binomial sendiri diambil dari ekspansi binomial dengan p sebagai peluang sukses dan $q = 1 - p$ sebagai peluang gagal

$$\begin{aligned} (p+q)^n &= \binom{n}{0} q^n + \binom{n}{1} pq^{n-1} + \binom{n}{2} p^2 q^{n-2} + \cdots + \binom{n}{n} p^n \\ &= b(0; n, p) + b(1; n, p) + b(2; n, p) + \cdots + b(n; n, p) \end{aligned}$$

Kita ketahui bahwa

$$\sum_{x=0}^{n} b(x;n,p) = (p+1-p)^n = 1$$

Maka syarat suatu distribusi peluang terpenuhi.

Contoh Soal 2. Peluang sebuah pasien sembuh dari penyakit kronis adalah 0.4. Jika 15 orang divonis terkena penyakit kritis tersebut, tentukan peluang:

(a) tepat 5 orang sembuh.

(b) dari 3 sampai 8 orang sembuh.

(c) paling tidak 5 orang sembuh.

Solusi. Misalkan X peubah acak yang menyatakan banyaknya orang yang sembuh dari penyakit kronis, maka $X \sim B(15, 0.4)$.

(a) Peluang tepat 5 orang sembuh adalah

$$P(X=5) = b(5;15,0.4) = \binom{15}{5}(0.4)^5(1-0.4)^{10} = 0.1859$$

(b) Peluang dari 3 sampai 8 orang sembuh adalah

$$\begin{aligned} P(3 \le x \le 8) &= \sum_{x=3}^{8} b(x;15,0.4) \\ &= \sum_{x=0}^{8} b(x;15,0.4) - \sum_{x=0}^{2} b(x;15,0.4) \\ &= 0.9050 - 0.0271 = 0.8879 \end{aligned}$$

(c) Peluang paling sedikit 5 orang yang selamat adalah

$$\begin{aligned} P(x \ge 5) &= 1 - P(x < 5) \\ &= 1 - \sum_{x=0}^{4} b(x;15,0.4) \\ &= 0.7827 \end{aligned}$$

■

Selanjutnya kita lihat nilai ekspektasi dan variansi dari Distribusi Binomial.

Teorema 3 Ekspektasi dan Variansi Distribusi Binomial

Ekspektasi dan Variansi dari Distribusi Binomial berturut-turut adalah

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

Bukti. Karena Distribusi Binomial tersusun dari Distribusi Bernoulli yang diulang sebanyak n usaha yang saling bebas, maka peubah acak Binomial X dapat dituliskan sebagai penjumlahan dari peubah-peubah acak Bernoulli $I_1, I_2, \cdots, I_n$.

$$X = I_1 + I_2 + \cdots + I_n$$

Dengan menggunakan sifat dari ekspektasi maka didapat

$$\mu = E(I_1) + E(I_2) + \cdots + E(I_n) = p + p + \cdots + p = np$$

Dengan sifat dari variansi juga kita dapatkan

$$\sigma_X^2 = \sigma_{I_1}^2 + \sigma_{I_2}^2 + \cdots + \sigma_{I_n}^2 = p(1-p) + p(1-p) + \cdots + p(1-p) + np(1-p)$$

■

Ekspektasi dan variansi ini akan sering digunakan untuk melakukan hampiran bila n terlalu besar (dipelajari di bab selanjutnya).

5.4 *Distribusi Multinomial*

Setelah kita membahas mengenai Distribusi **Bi**nomial, sekarang kita membahas mengenai distribusi yang keluarannya tidak hanya sukses atau gagal saja, melainkan lebih dari dua keluaran, yang kita sebut dengan Distribusi **Multi**nomial. Pada dasarnya dari definisi, distribusi ini hanya perluasan dari sisi peubah dari Distribusi Binomial.

Jika n usaha saling bebas menghasilkan keluaran $E_1, E_2, \cdots, E_k$ dengan peluang $p_1, p_2, \cdots, p_k$, maka distribusi peluang dari peubah acak $X_1, X_2, \cdots, X_k$ adalah

$$f(x_1, x_2, \cdots, x_k; p_1, p_2, \cdots, p_k, n) = \binom{n}{x_1, x_2, \cdots, x_n} p_1^{x_1} p_2^{x_2} \cdots p_n^{x_n}$$

dengan syarat

$$\sum_{i=1}^{k} x_i = n \text{ dan } \sum_{i=1}^{n} p_i = 1$$

dan

$$\binom{n}{x_1, x_2, \cdots, x_k} = \frac{n!}{x_1! x_2! \cdots x_k!}$$

Pembuktian dari syarat distribusi peluang pun masih sama, yaitu dengan memandang ekspansi dari bentuk:

$$(p_1 + p_2 + \cdots + p_k)^n$$

Contoh Soal 3. Suatu program dalam komputer dibuat untuk memunculkan angka 1, 2, dan 3 secara acak dengan peluang sebagai berikut: peluang muncul angka 1 adalah $\frac{2}{9}$, angka 2 adalah $\frac{1}{6}$ dan angka 3 sebesar $\frac{11}{18}$. Tentukan peluang ketika program tersebut dieksekusi 6 kali, maka terdapat keluaran angka 1 sebanyak 2 kali, angka 2 sebanyak 1 kali, dan angka 3 sebanyak 3 kali.

Solusi. Dengan menggunakan distribusi multinomial ktia dapatkan

$$f\left(2, 1, 3; \frac{2}{9}, \frac{1}{6}, \frac{11}{18}, 6\right) = \binom{6}{2, 1, 3} \left(\frac{2}{9}\right)^2 \left(\frac{1}{6}\right)^1 \left(\frac{11}{18}\right)^3 = 0.1127$$

■

5.5 *Distribusi Hipergeometrik*

Distribusi yang akan kita lihat selanjutnya adalah Distribusi Hipergeometrik. Pada prinsipnya, distribusi ini tidak jauh berbeda dengan Distribusi Binomial, yaitu mempunyai keluaran sukses dan gagal. Perbedaan mendasar adalah pada Binomial percobaan dilakukan **dengan pengembalian** sedangkan Hipergeometrik, percobaan dilakukan **tanpa pengembalian**.

Jika X peubah acak yang menyatakan banyaknya sukses dalam sampel acak ukuran n yang diambil dari N benda yang mengandung k bernama sukses dan $N-k$ bernama gagal. Maka X berdistribusi Hipergeometrik (notasi $X \sim h(N,n,k)$) dengan peluang:

$$P(X=x) = h(x;N,n,k) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

dengan syarat $\max\{0, n-(N-k)\} \leq x \leq \min\{n,k\}$.

Contoh Soal 4. Dari 50 gedung di sebuah kawasan industri, 12 gedung mempunyai kode pelanggaran. Jika 10 gedung dipilih secara acak dalam suatu inspeksi, hitung peluang bahwa 3 dari 10 gedung mempunyai kode pelanggaran.

Solusi. Misal X peubah acak yang menyatakan banyak gedung yang dipilih mempunyai kode pelanggaran.

$$P(X=3) = h(3;50,10,12) = \frac{\binom{12}{3}\binom{38}{7}}{\binom{50}{10}} = 0.2703$$

■

Seperti biasa, kita akan lihat ekspektasi dan variansi dari distribusi ini.

Teorema 4 Ekspektasi dan Variansi Distribusi Hipergeometrik

Ekspektasi dan Variansi dari Distribusi Hipergeometrik berturut-turut adalah

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

Bukti secara kasar dari teorema 4 adalah dengan menggunakan hampiran dari Distribusi Binomial.

Teorema 5 Hampiran Distribusi Hipergeometrik dengan Distribusi Binomial

Untuk ukuran sampel acak n yang diambil semakin kecil terhadap N (standarnya $n/N \leq 0.05$), maka Distribusi Hipergeometrik dapat dihampiri oleh Distribusi Binomial, dengan peluang sukses $\frac{k}{N}$.

$$h(N,n,k) \xrightarrow{n<<<N} b\left(n, \frac{k}{N}\right)$$

Bukti secara kasar pada ekspektasi didapat dengan menggunakan substitusi pada peluang Binomial sebesar $\frac{k}{N}$. Sedangkan pada variansi, kita lihat bahwa terdapat faktor koreksi $\frac{N-n}{N-1}$ yang membedakan distribusi ini dari Distribusi Binomial. Perhatikan bahwa jika $n/N \leq 0.05$ maka faktor koreksi tersebut dapat diabaikan.

Pertanyaan lanjutan: Bagaimana untuk Distribusi Hipergeometrik Multivariat (Distribusi Multinomial dengan percobaan tanpa pengembalian)?

Jika N barang dapat dipartisi menjadi k klasifikasi $A_1, A_2, \cdots, A_k$ dengan $a_1, a_2, \cdots, a_k$ elemen, maka distribusi peluang dari peubah acak $X_1, X_2, \cdots, X_k$ dalam sampel acak n adalah:

$$f(x_1, x_2, \cdots, x_k; a_1, a_2, \cdots, a_k, N, n) = \frac{\binom{a_1}{x_1}\binom{a_2}{x_2}\cdots\binom{a_k}{x_k}}{\binom{N}{n}}$$

dengan

$$\sum_{i=1}^{k} x_1 = n \text{ dan } \sum_{i=1}^{k} a_i = N$$

5.6 Distribusi Poisson

Distribusi terakhir masih mengenai Distribusi Binomial, yaitu memiliki keluaran sukses atau gagal. Namun percobaannya dilakukan pada suatu selang waktu atau daerah. Contoh: Banyak pengunjung yang datang ke dalam suatu toko dari pukul 8 hingga 9 pagi, atau banyaknya suatu jenis tanaman di daerah Bandung Barat. Percobaan tersebut disebut juga dengan Percobaan Poisson. Ciri-cirinya adalah sebagai berikut:

(a) Selang waktu atau daerahnya saling bebas.

(b) Peluang pada Proses Poisson tergantung pada selang waktu atau besarnya daerah.

(c) Peluang untuk selang yang pendek atau daerah yang sempit dalat diabaikan.

Distribusi Poisson sendiri mempunyai parameter λ sebagai perbandingan sukses terhadap suatu waktu atau daerah. Misal peubah acak X berdistribusi Poisson (notasi: $X \sim Poi(\lambda t)$) maka peluangnya adalah:

$$P(X = x) = Poi(x; \lambda t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}$$

untuk $x = 0, 1, 2, \cdots$ dan $e = 2.71828\ldots$ merupakan tetapan Euler.

Hal yang menarik dapat terlihat dari ekspektasi dan variansi Distribusi Poisson.

Teorema 6 Ekspektasi dan Variansi Distribusi Poisson
Ekspektasi dan Variansi dari Distribusi Poisson berturut-turut adalah

$$\mu = \lambda t \qquad \sigma^2 = \lambda t$$

Bukti. Untuk ekspektasi:

$$\begin{aligned}\mu &= \sum_{x=0}^{\infty} x \frac{e^{-\lambda t}(\lambda t)^x}{x!} \\ &= 0 + \sum_{x=1}^{\infty} x \frac{e^{-\lambda t}(\lambda t)^x}{x!} \\ &= \lambda t \sum_{x=1}^{\infty} \frac{e^{-\lambda t}(\lambda t)^{x-1}}{(x-1)!}\end{aligned}$$

Misalkan $y = x - 1$ maka didapat

$$\mu = \lambda t \sum_{y=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^y}{y!}$$

Karena suku jumlahan tak hingga merupakan fungsi massa peluang dari Distribusi Poisson maka hasil dari jumlahan tersebut adalah 1. Jadi,

$$\mu = \lambda t$$

Selanjutnya untuk variansi kita lakukan sedikit manipulasi aljabar menggunakan sifat dari ekspektasi:

$$\sigma^2 = E(X^2) - \mu^2 = E(X^2 - X) + \mu - \mu^2$$

Lalu nilai dari $E(X^2 - X)$ adalah:

$$\begin{aligned} E(X^2 - X) &= E(X(X-1)) \\ &= \sum_{x=0}^{\infty} x(x-1)\frac{e^{-\lambda t}(\lambda t)^x}{x!} \\ &= 0 + 0 + \sum_{x=2}^{\infty} x(x-1)\frac{e^{-\lambda t}(\lambda t)^x}{x!} \\ &= (\lambda t)^2 \sum_{x=2}^{\infty} \frac{e^{-\lambda t}(\lambda t)^{x-2}}{(x-2)!} \end{aligned}$$

Gunakan pemisalan $y = x - 2$ sehingga didapat

$$E(X^2 - X) = (\lambda t)^2 \sum_{y=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^y}{y!} = (\lambda t)^2 = \mu^2$$

Substitusi ke persamaan variansi menghasilkan

$$\sigma^2 = E(X^2 - X) + \mu - \mu^2 = \mu^2 - \mu^2 + \mu = \lambda t$$

■

Contoh Soal 5. Rata-rata banyaknya pelanggan yang datang setiap jam di sebuah jasa mobil adalah 7.

(a) Hitung peluang bahwa lebih dari 2 pelanggan akan datang dalam suatu periode 30 menit.

(b) Berapa rata-rata banyaknya pelanggan yang datang dalam suatu periode 2 jam?

Solusi. Misal X menunjukkan banyaknya pelanggan yang datang setiap jam di sebuah jasa mobil, maka $X \sim Poi(7)$. Didapat informasi bahwa $\mu_X = 7$ dan $\sigma_X^2 = 7$.

(a) Karena waktunya 30 menit maka $t = 0.5$ (t yang digunakan dalam satuan jam).

$$\begin{aligned} P(X > 2) &= 1 - P(X \leq 2) \\ &= 1 - \sum x = 0^2 \frac{e^{3.5}(3.5)^x}{x!} \\ &= 0.494 \end{aligned}$$

(b) Jika dalam 1 jam rata-rata banyak pelanggan yang datang adalah 7 orang, maka dalam 2 jam rata-ratanya akan menjadi 14 orang.

■

Karena Distribusi Poisson cukup berkaitan dengan Distribusi Binomial, maka perhitungan Distribusi Binomial dapat dihampiri dengan menggunakan Distribusi Poisson.

Teorema 7 Hampiran Distribusi Binomial dengan Distribusi Poisson

Misal X peubah acak Binomial dengan peluang $b(x; n, p)$. Saat $n \to \infty$, $p \to 0$, dan $np \xrightarrow{n\to\infty} \mu$ konstan maka:

$$b(x; n, p) \xrightarrow{n\to\infty} p(x; \mu)$$

5.7 *Distribusi Geometrik*

Distribusi Geometrik memberikan cara pandang yang berbeda terhadap usaha yang menghasilkan keluaran sukses atau gagal. Pada distribusi sebelumnya kita mencari peluang berapa banyak yang sukses atau gagal. Pada Distribusi Geometrik kita mencari peluang banyaknya usaha agar terjadi sukses yang pertama dari usaha-usaha yang saling bebas.

Misalkan X menyatakan peubah acak Distribusi Geometrik (notasi: $X \sim Geom(p)$) dengan p menyatakan peluang sukses, maka peluangnya:

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

untuk $x = 1, 2, \ldots$.

Teorema 8 Ekspektasi dan Variansi Distribusi Geometrik

Ekspektasi dan Variansi dari Distribusi Geometrik berturut-turut adalah

$$\mu = \frac{1}{p} \qquad \sigma^2 = \frac{1-p}{p^2}$$

Contoh Soal 6. Suatu tes hasil pengelasan logam meliputi proses pengelasan sampai suatu patahan terjadi. Pada jenis pengelasan tertentu, patahan terjadi 80% disebabkan oleh logam itu sendiri dan 20% oleh penyinaran pada pengelasan. Beberapa hasil pengelasan dites. Misalkan X adalah banyak tes yang dilakukan sampai ditemukan patahan pertama pada hasil pengelasan. Hitung peluang pada tes ketiga ditemukan patahan pertama.

Solusi. $X \sim Geom(0.2)$

$$P(X = 3) = Geom(3; 0, 2) = 0.2(0.8)^2 = 0.128$$

Cara lain: dengan menggunakan analisis kasus kita dapatkan informasi bahwa tes pertama dan kedua harus gagal dan tes ketiga harus sukses. Peluangnya adalah sebagai berikut

$$P = 0.8 \times 0.8 \times 0.2 = 0.128$$

■

5.8 *Distribusi Binomial Negatif*

Distribusi yang terakhir adalah perluasan dari Distribusi Geometrik. Pada distribusi ini, banyaknya sukses yang ditentukan lebih dari 1, sebutlah k. Maka peubah acak dari Binomial Negatif adalah jumlahan dari peubah acak Geometrik yang saling bebas.

$$X = Y_1 + Y_2 + \cdots + Y_k$$

Dinotasikan $X \sim b^*(k, p)$ maka peluangnya adalah:

$$P(X = x) = b^*(x; k, p) = \binom{x-1}{k-1} p^k (1 - p)^{x-k}$$

dengan $x = k, k + 1, k + 2, \ldots$.

Teorema 9 Ekspektasi dan Variansi Distribusi Binomial Negatif

Ekspektasi dan Variansi dari Distribusi Binomial Negatif berturut-turut adalah

$$\mu = \frac{k}{p} \qquad \sigma^2 = \frac{k(1-p)}{p^2}$$

Contoh Soal 7. Mengacu pada contoh soal 6, tentukan peluang bahwa dilakukan 8 tes sehingga ditemukan 3 patahan pertama.

Solusi. $X \sim b^*(3, 0.2)$

$$P(X = 8) = b^*(8; 3, 0.2) = \binom{7}{2}(0.2)^3(0.8)^5 = 0.05505$$

■

5.9 *Soal Latihan*

1. Suatu hasil penelitian menyebutkan bahwa tingkat kematian bayi di Indonesia adalah 37 bayi dari 1000 kelahiran. Misalkan pada tanggal 13 Maret 2006 di RSHS ada 25 ibu yang mau melahirkan, berapakah peluang:

   a) tepat 20 bayi akan selamat.

   b) bayi yang meninggal paling banyak 3 orang.

2. Sebuah jenis biskuit tertera pada label pembungkusnya bahwa berat biskuit tersebut 20.7 gram. Misal peluang sebuah biskuit yang beratnya lebih besar dari 20.7 gram adalah 0.85. Jika X menyatakan banyaknya biskuit yang beratnya lebih dari 20.7 gram yang diambil dari sebuah sampel acak yang berisi 8 buah biskuit:

   a) sebutkan distribusi X jika diasumsikan bebas.

   b) hitung peluang semua biskuit beratnya lebih dari 20.7 gram.

   c) hitung peluang banyaknya biskuit yang beratnya lebih dari 20.7 gram tidak lebih dari 6.

3. Diketahui bahwa banyaknya kesalahan cetak di dalam sebuah buku referensi berdistribusi Poisson dengan rataan (mean) kesalahan 0.01 kesalahan cetak per halaman. Berapa peluang terdapat paling banyak 3 kesalahan cetak dalam 100 halaman?

Bab 6

Distribusi Peluang Kontinu

Setelah kita melihat distribusi peluang diskrit, pada bab ini kita akan melihat beberapa distribusi peluang kontinu yang menyangkut masalah sehari-hari. Distribusi kontinu pun berguna untuk melakukan statistika inferensi, yang akan dipelajari di bab selanjutnya.

6.1 *Distribusi Uniform*

Seperti pada distribusi peluang diskrit, dalam distribusi peluang kontinu pun terdapat Distribusi Uniform. Misal peubah acak X beridistribusi Uniform (notasi: $X \sim U[A, B]$) maka fungsi peluangnya:

$$f(x; A, B) = \begin{cases} \frac{1}{B-A} & A \leq x \leq B \\ 0 & \text{lainnya} \end{cases}$$

Interval dalam fungsi kepadatan peluang Distribusi Uniform tidak harus dalam interval tertutup. Namun, dalam aplikasinya kita gunakan asumsi kemungkinan dalam interval $[A, B]$ konstan.

<u>Contoh Soal 1.</u> Dalam sebuah konferensi dari suatu perusahaan hanya dapat dipesan tidak lebih dari 4 jam. Kita asumsikan X sebagai peubah acak distribusi uniform yang menyatakan lamanya konferensi tersebut.

(a) Tentukan fungsi kepadatan peluangnya.

(b) Tentukan peluang konferensi berlangsung paling tidak 3 jam.

<u>Solusi.</u>

(a) $f(x) = \begin{cases} \frac{1}{4} & 0 \leq x \leq 4 \\ 0 & \text{lainnya} \end{cases}$

(b) $P(X \geq 3) = \displaystyle\int_3^4 \frac{1}{4}\, dx = \frac{1}{4}.$

■

Selanjutnya kita lihat ekspektasi dan variansi dari distribusi tersebut.

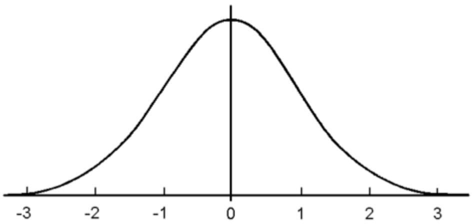
<u>Teorema 1</u> Ekspektasi dan Variansi Distribusi Uniform

Ekspektasi dan Variansi dari Distribusi Uniform berturut-turut adalah

$$\mu = \frac{A+B}{2} \qquad \sigma^2 = \frac{(B-A)^2}{12}$$

6.2 Distribusi Normal

Distribusi yang paling penting dan banyak menyangkut dalam berbagai topik dalam statistika yaitu Distribusi Normal. Distribusi ini ditemukan oleh Karl Friedrich Gauss sehingga terkadang distribusi ini disebut juga Distribusi Gauss. Ciri khas dari Kurva Distribusi Normal adalah bentuk kurva yang **simetris** dengan garis simetris $z = \mu$.



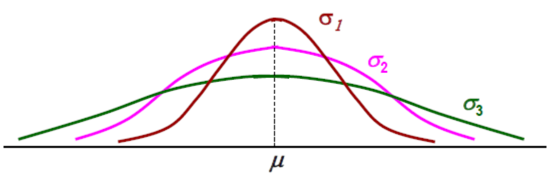
Gambar 6.1: Bentuk Kurva Distribusi Normal dengan $\mu = 0$ dan $\sigma = 1$ (Normal Baku)

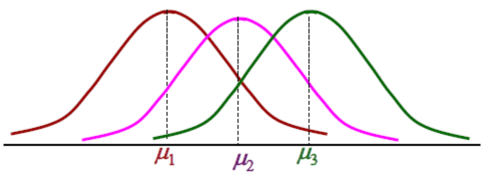
Misalkan X peubah acak berdistribusi normal (notasi: $X \sim N(\mu, \sigma)$) maka fungsi peluangnya:

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

dengan $-\infty \leq x \leq \infty$, $\pi = 3.14159\ldots$, dan $e = 2.71828\ldots$.

Dalam parameternya, dapat terlihat bahwa Distribusi Normal dipengaruhi oleh parameter μ dan σ. Perhatikan pada gambar 6.2 bahwa perubahan μ berdampak pada pergeseran kurva ke kiri atau ke kanan dan perubahan σ berdampak pada kurtosis (kelancipan) kurva normal.





Gambar 6.2: Bentuk Perubahan Kurva Distribusi Normal (Keterangan: $\sigma_1 < \sigma_2 < \sigma_3$ dan $\mu_1 < \mu_2 < \mu_3$)

Seperti biasa, kita tentukan terlebih dahulu ekspektasi dan variansi dari Distribusi Normal.

Teorema 2 Ekspektasi dan Variansi Distribusi Normal
Ekspektasi dan Variansi dari Distribusi Normal berturut-turut adalah μ dan σ^2.

Bukti. Kita hitung ekspektasinya terlebih dahulu. Untuk mempermudah perhitungan, kita hitung $E(X - \mu)$.

$$E(X-\mu) = \int_{-\infty}^{\infty} \frac{x-\mu}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

Gunakan pemisalan $z = \frac{x-\mu}{\sigma}$ sehingga $dx = \sigma\, dz$ menghasilkan:

$$E(X-\mu) = \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{-z^2/2}\, dz = 0$$

Nilai integral menjadi nol karena integran (fungsi yang diintegralkan) adalah fungsi ganjil. Tinjau ruas kiri dari persamaan di atas.

$$E(X) - E(\mu) = E(X) - \mu = 0 \Leftrightarrow E(X) = \mu$$

Selanjutnya untuk variansi kita hitung $E((X-\mu)^2)$.

$$E((X-\mu)^2) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}(x-\mu)^2 e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\,dz$$

Dengan menggunakan pemisalan yang sama, kita dapatkan:

$$E((X-\mu)^2) = \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty} z^2 e^{-\frac{z^2}{2}}\,dz$$

Menggunakan integral parsial dengn $u = z$ dan $dv = ze^{-z^2/2}$ sehingga didapat

$$E((X-\mu)^2) = \sigma^2$$

(verifikasi.) ■

Permasalahan kita sekarang adalah bagaimana cara menghitung peluang dari suatu kejadian yang berdistribusi Normal. Kita sudah melihat bahwa fungsi kepadatan peluang Distribusi Normal tidak mudah untuk diintegralkan, bahkan butuh perhitungan secara numerik untuk mengetahui luas di bawah kurva normal dalam interval tertentu.

Salah satu prinsip yang diberikan adalah dengan menggunakan **kurva normal baku**, yaitu kurva normal dengan $\mu = 0$ dan $\sigma = 1$ (lihat gambar 6.1). Misalkan X suatu peubah acak yang berdistribusi normal dengan mean dan variansinya berturut-turut μ dan σ. Maka untuk mencari $P(x_1 < X < x_2)$, kita definisikan peubah acak Z dengan aturan $Z = \dfrac{X-\mu}{\sigma}$. Maka Z berdistribusi normal baku (dengan $\mu = 0$ dan $\sigma = 1$). Maka kita dapatkan:

$$z_1 = \frac{x_1-\mu}{\sigma} \qquad\qquad z_2 = \frac{x_2-\mu}{\sigma}$$

Kita dapat hitung peluangnya sebagai berikut:

$$\begin{aligned}
P(x_1 < X < x_2) &= \frac{1}{\sqrt{2\pi}\sigma}\int_{x_1}^{x_2} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\,dx \\
&= \frac{1}{\sqrt{2\pi}}\int_{z_1}^{z_2} e^{-\frac{z^2}{2}}\,dz \\
&= \int_{z_1}^{z_2} n(z;0,1)\,dz = P(z_1 < Z < z_2)
\end{aligned}$$

Proses transformasi kurva normal menjadi normal baku disebut dengan *normalisasi*. Nilai luas di bawah kurva normal baku tersaji dalam tabel di lampiran 1.

Contoh Soal 2. Suatu mesin pembuat resistor mempunyai rataan resistansi sebesar 40 ohm dan standar deviasi 2 ohm. Asumsikan resistansi berdistribusi Normal dan dapat diukur untuk semua derajat akurasi, tentukan persentase resistor yang mempunyai resistansi lebih dari 43 ohm.

Solusi. Misal X peubah acak yang menyatakan resistansi dari mesin pembuat resistor. Tugas kita adalah mencari nilai luas di bawah kurva normal dan di sebelah kanan nilai $x = 43$. Kita normalisasi nilai x tersebut sehingga menghasilkan:

$$z = \frac{43-40}{2} = 1.5$$

Sehingga peluangnya adalah:

$$P(X > 43) = P(Z > 1.5) = 1 - P(Z \le 1.5) = 1 - 0.9332 = 0.0668$$

Jadi, 6.68% dari resistor yang diproduksi resistansinya akan melebihi 43 ohm. ■

Contoh Soal 3. Dalam sebuah ujian, rata-rata nilainya 74 dan standar deviasinya 7. Jika 12% dari kelas tersebut akan diberikan nilai A, dan penilaian mengikuti distribusi normal, tentukan nilai A terkecil dalam kelas tersebut. Asumsikan nilai ujian dalam bilangan bulat.
Solusi. Sekarang, kita bekerja secara terbalik. Diketahui $P(Z < z) = 0.12$. Kita mulai mencari z yang memenuhi nilai tersebut. Dari tabel lampiran 1, kita dapatkan bahwa $z = 1.18$. Setelah itu kita cari nilai x yang memenuhi.

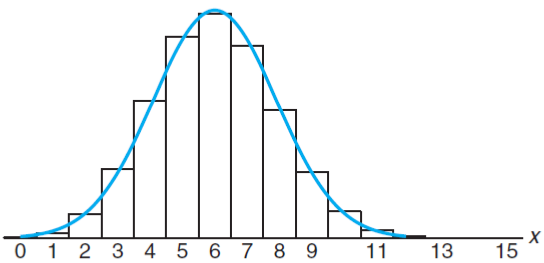
$$z = \frac{x-\mu}{\sigma} \Leftrightarrow x = z\sigma + \mu = 82.26$$

Maka nilai A terkecil adalah 83 (karena asumsi nilai ujian bilangan bulat). ■

Hal terakhir yang akan dilihat mengenai Distribusi Normal adalah hampiran Binomial melalui Distribusi Normal. Perhatikan teorema berikut.

Teorema 3 Hampiran Distribusi Binomial dengan Distribusi Normal
Misal X peubah acak Binomial dengan ekspektasi $\mu = np$ dan variansi $\sigma = np(1-p)$ serta $n \to \infty$ atau $p \to \frac{1}{2}$ maka kita dapat bentuk peubah acak $Z \sim N(0,1)$ dengan nilai:

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



Gambar 6.3: Ilustrasi Hampiran Binomial dengan Normal

Pada dasarnya bila kita gambarkan histogram dari peubah acak X dan kurva normal seperti pada gambar 6.3, terdapat galat dari x. Untuk perhitungan daerah di sebelah kiri x ($P(X \leq x)$), kita gunakan perubahan $x + 0.5$ karena hampiran ini melibatkan distribusi diskrit ke kontinu. Faktor koreksi $+0.5$ ini disebut juga **koreksi kontinuitas**. Namun hal lain terjadi jika kita mencari $P(X < x)$, kita gunakan koreksi kontinuitas sebanyak -0.5. Hal yang sepadan terjadi untuk daerah di sebelah kanan x.

Contoh Soal 4. Peluang seorang pasien sembuh dari penyakit kronis adalah 0.4. Jika 100 orang terkena penyakit kronis, berapa kemungkinan kurang dari 30 orang sembuh?
Solusi. Misal X peubah acak yang menyatakan banyaknya pasien yang sembuh. Karena $n = 100$ kita bisa dapatkan nilai hampiran yang baik dengan menggunakan hampiran normal.

$$\mu = np = 40 \text{ dan } \sigma = \sqrt{np(1-p)} = 4.899$$

Sehingga kita dapatkan

$$P(X < 30) \approx P\left(Z < \frac{29.5 - 40}{4.899}\right) = P(Z < -2.14) = 0.0162$$

■

6.3 Distribusi Gamma dan Eksponensial

Distribusi terkahir yang mempunyai peranan penting dalam bidang sains dan insinyur adalah Distribusi Gamma. Misal peubah acak X beridstribusi Gamma (notasi: $X \sim Gamma(\alpha, \beta)$) dengan peluang:

$$f(x; \alpha, \beta) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} & x > 0 \\ 0 & \text{lainnya} \end{cases}$$

Dengan $\Gamma(\alpha)$ didefinisikan sebagai fungsi Gamma

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx \qquad \alpha > 0$$

Sifat-sifat yang dimiliki oleh fungsi Gamma adalah sebagai berikut.

(a) $\Gamma(n) = (n-1)!$ untuk bilangan bulat positif n.

(b) $\Gamma(1) = 1$

(c) $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

Parameter α dan β pada Distribusi Gamma dapat membuat distribusi-distribusi lainnya, seperti Weilbull, Khi Kuadrat, dan lainnya. Distribusi yang akan kita bahas lainnya adalah Distribusi Eksponensial, yaitu distribusi Gamma saat $\alpha = 1$.

Misal X peubah acak yang berdistribusi Eksponensial (notasi $X \;\exp(\beta)$) maka peluangnya:

$$f(x; \beta) = \begin{cases} \frac{1}{\beta} e^{-\frac{x}{\beta}} & x > 0 \\ 0 & \text{lainnya} \end{cases}$$

Kita lihat ekspektasi dan variansi dari kedua distribusi tersebut.

Teorema 4 Ekspektasi dan Variansi Distribusi Gamma

Ekspektasi dan Variansi dari Distribusi Normal berturut-turut adalah

$$\mu = \alpha\beta \qquad \sigma^2 = \alpha\beta^2$$

Akibatnya, mean dan variansi dari Distribusi Eksponensial adalah $\mu = \beta$ dan $\sigma^2 = \beta^2$

Interpretasi parameter β dalam Distribusi Eksponensial berasal dari Distribusi Poisson. Ingat kembali bahwa Distribus Poisson menghitung peluang banyaknya sukses dari suatu rentang waktu atau daerah. Dalam aplikasinya, suatu rentang waktu atau daerah tersebut merupakan peubah acak. Parameter λ pada Distribusi Poisson bertindak sebagai rataan sukses per selang waktu atau daerah.

Sekarang kita akan mencoba mencari waktu atau daerah yang diperlukan untuk mendapat sukses pertama. Menggunakan Distribusi Poisson, maka peluang tidak ada sukses dari suatu waktu atau daerah t adalah:

$$p(0; \lambda t) = e^{-\lambda t}$$

Karena sekarang kita membuat waktu atau daerah menjadi sebuah peubah acak, misalkan X, maka didapatkan peluang waktu atau daerah yang dibutuhkan untuk mendapat sukses pertama adalah:

$$P(X > x) = e^{-\lambda x}$$

Kita dapatkan fungsi distribusi kumulatif untuk X:

$$P(0 \leq X \leq x) = 1 - e^{-\lambda x}$$

Persamaan tersebut tak lain menjadi fungsi distribusi kumulatif untuk distribusi eksponensial. Kita dapat mengeceknya dengan menurunkan fungsi tersebut sehingga didapat:

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

dengan $\lambda = \frac{1}{\beta}$.

Contoh Soal 5. Misalkan lama pembicaraan telepon dapat dimodelkan oleh distribusi eksponensial dengan rataan 10 menit/orang. Bila seseorang tiba-tiba mendahului Anda di suatu telepon umum, carilah peluangnya bahwa Anda harus menunggu antara 10 sampai 20 menit.

Solusi. Misalkan X peubah acak yang menyatakan lama pembicaraan telepon. Diketahui $X \sim \exp(\frac{1}{10})$ sehingga

$$f(x) = \frac{1}{10}e^{-\frac{x}{10}}$$

Lama pembicaraan setara dengan waktu menunggu sehingga

$$P(10 < X < 20) = \int_{10}^{20} \frac{1}{10}e^{-\frac{x}{10}}\, dx = 0.233$$

■

6.4 *Soal Latihan*

1. Banyaknya peserta kuliah *online* suatu mata kuliah tertentu dianggap berdistribusi normal dengan rataan 100 dan standa deviasi 25. Tentukan
   a) Peluang peserta *online* mata kuliah lebih dari 175 orang.
   b) Peluang peserta *online* mata kuliah tersebut kurang dari 105 orang tapi tidak kurang dari 90 orang.
   c) Di bawah banyak peserta berapakah terdapat 15% dari seluruh peserta kuliah tersebut?

2. Misalkan variabel acak X menyatakan diameter 'dot' (titik) hasil cetak dari sebuah mesin printer. Kita modelkan distribusi X sebagai distribusi normal dengan mean diameter 0.002 inci dengan simpangan baku 0.0004 inci.
   a) Berapa peluang diameter 'dot' melebihi 0.0026 inci?
   b) Tentukan c agar $P(X > c) = 0.95$.
   c) Dengan menggunakan mean 0.002, tentukan simpangan baku dari diameter 'dot' agar $P(0.0014 < X < 0.0026) = 0.995$.

3. Umur suatu komponen elektronik berdistribusi eksponensial dengan tingkat kegagalan $\beta = 2$. Seratus alat dipasang pada sistem yang berlainan. Tentukan
   a) model distribusi banyaknya alat yang rusak pada tahun pertama.
   b) peluang paling banyak 5 gagal pada tahun pertama.

Bab 7

Statistika Inferensi: Penaksiran

Pada bab ini dan bab selanjutnya kita akan mempelajari tentang statistika inferensi. Pada dasarnya, kita tidak dapat menentukan secara eksak rataan atau variansi dari suatu populasi. Cara lainnya yaitu melakukan penaksiran melalui sampel dari sebuah populasi yang diambil secara acak.

7.1 Penaksiran Titik dan Selang

Dalam suatu populasi, kita tidak bisa menentukan parameter-parameter statistik secara tepat. Namun kita dapat *menaksir* parameter tersebut melalui sampel acak yang berasal dari populasi tersebut. Penaksiran dapat dibagi menjadi dua kategori, yaitu penaksiran titik dan selang.

Ciri-ciri penaksir yang baik adalah sebagai berikut:

- **Tak Bias**
  Sebuah statistik dapat dihitung bias nya dengan persamaan berikut:
  $$\text{Bias}(\theta) = E(\hat{\theta}) - \theta$$
  Statistik $\hat{\theta}$ dikatakan sebagai penaksir tak bias bagi parameter θ apabila ekspektasi dari statistik tersebut berupa parameter itu sendiri.
  $$\text{Bias}(\theta) = 0 \Leftrightarrow E(\hat{\theta}) = \theta$$
- **Variansi Minimum**
  Penaksir yang paling efisien bagi θ adalah $\hat{\theta}$ yang memiliki variansi paling kecil.
- **Konsisten**
  Jika $\hat{\theta}$ merupakan penaksir untuk parameter θ yang didasarkan pada sampel acak berukuran n, maka $\hat{\theta}$ dikatakan penaksir konsisten bagi parameter θ apabila:
  $$\lim_{n\to\infty} P(|\hat{\theta} - \theta| < \epsilon) = 1$$
- **Cukup Secara Statistik**
  Statistik $Y = Y(x_1, x_2, \cdots, x_n)$ dikatakan cukup bagi parameternya jika fungsi kepadatan peluang bersyarat $P(x_1, x_2, \cdots, x_n) = x_n|Y(x_1, x_2, \cdots, x_n) = t$ tidak bergantung pada θ.

Beberapa contoh penaksir yang baik yaitu $\bar{x}$ untuk menaksir ekspektasi μ dan s^2 untuk menaksir σ^2.

Dalam penaksiran titik, kita menentukan suatu nilai tunggal dari suatu parameter dengan metode-metode tertentu. Sebagai contoh: seekor ayam betina kira-kira dapat bertelur sebanyak 10 butir setiap minggunya. Nilai tunggal yang dimaksud dalam contoh adalah 10 butir, yang didapat melalui metode penelitian tertentu. Contoh penaksiran titik lainnya yaitu $\bar{x}$ untuk menaksir ekspektasi μ dan s^2 untuk menaksir σ^2 (buktinya dengan menunjukkan bahwa $\bar{x}$ dan s^2 memenuhi ciri-ciri penaksir yang baik).

Dalam penaksiran selang, kita akan menentukan nilai sebenarnya dari suatu parameter, namun dalam bentuk interval tertentu. Interval tersebut diatur melalui suatu variabel yang disebut tingkat signifikansi/keberartian (notasi: α). Sebagai contoh: seekor sapi dapat menghasilkan susu antara 1 hingga 2 liter dalam satu minggu. Dalam bab ini akan lebih ditekankan mengenai penaksiran selang melalui selang kepercayaan untuk parameter μ (rataan) dan σ^2 (variansi) dalam satu populasi dan dua populasi (berpasangan atau tidak berpasangan).

Kembali kepada pembahasan mengenai tingkat signifikansi dalam penaksiran selang. Nilai dari tingkat signifikansi α berkisar dari $0 \leq \alpha \leq 1$ menyatakan galat yang ditoleransi dalam sebuah penaksiran. Semakin kecil α maka akan semakin pendek selang yang akan didapatkan. Prinsip dasar dari penaksiran selang untuk sebuah parameter, sebutlah θ, adalah menentukan batas bawah $\hat{\theta}_L$ dan batas atas $\hat{\theta}_U$ yang memenuhi:

$$P(\hat{\theta}_L < \theta < \hat{\theta}_U) = 1 - \alpha$$

Dengan kata lain, kita akan mencari suatu interval yang membuat peluang terambilnya suatu nilai dalam interval tersebut adalah $1 - \alpha$. Interval yang dimaksud disebut dengan *selang kepercayaan.*

7.2 Selang Kepercayaan untuk Rataan

Sebelumnya, akan dikenalkan terlebih dahulu teorema yang menjadi kunci utama dalam penaksiran selang, yaitu Teorema Limit Pusat.

Teorema 1 Teorema Limit Pusat

Jika $\bar{x}$ rataan untuk sampel acak sebanyak n buah yang diambil dari populasi dengan rataan μ dan variansi σ^2, maka untuk $n \to \infty$

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

Populasi akan berdistribusi normal baku $n(z; 0, 1)$.

Teorema Limit Pusat banyak digunakan bagi sampel-sampel berukuran besar, khususnya dalam statistika inferensi. Inti dari teorema tersebut adalah, untuk n yang sangat besar (standarnya $n \geq 30$) maka bentuk distribusi dari suatu sampel berukuran besar tersebut akan mirip dengan Distribusi Normal.

7.2.1 Satu Populasi

Dalam penaksiran rataan untuk satu populasi, kita bagi menjadi dua kategori berdasarkan variansi populasinya.

1. **Variansi populasi diketahui**. Melalui Teorema Limit Pusat, penaksiran selang bagi rataan didefinisikan sebagai berikut:

   $$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha \text{ dengan } Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

   Dengan menyelesaikan ruas kiri persamaan, maka didapat selang kepercayaan untuk tingkat signifikansi α sebagai berikut:

   $$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

   Contoh Soal 1. Konsentrasi seng rata-rata yang diambil dari 36 tempat berbeda dalam sebuah sungai adalah 2.6 gram per mm. Tentukan selang kepercayaan 95% untuk rataan konsentrasi seng dalam sungai tersebut. Diketahui bahwa standar deviasi populasi adalah 0.3 gram per mm.
   Solusi. Diketahui $\bar{x} = 2.6$, $n = 36$, dan $\sigma = 0.3$. Selang kepercayaan 95% ($\alpha = 5\%$) rataannya adalah:

   $$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Untuk menentukan nilai $z_{\alpha/2} = z_{0.025}$, kita cari nilai z yang luas di bawah kurvanya sebesar 0.025 (gunakan lampiran 1). Didapat $z_{0.025} = 1.96$ sehingga selang kepercayaannya menjadi:

$$2.6 - 1.96\frac{0.3}{\sqrt{36}} < \mu < 2.6 + 1.96\frac{0.3}{\sqrt{36}} \Leftrightarrow 2.47 < \mu < 2.73$$

■

Kita juga dapat menentukan jumlah data yang dibutuhkan agar tidak melebihi suatu galat ($e = |\mu - \bar{x}|$) dari rataan. Perhatikan teorema berikut.

Teorema 2

Jika $\bar{x}$ penaksir μ, kita dapat buat selang kepercayaan $100(1 - alpha)\%$ dengan galat tidak akan melebihi e ketika ukuran sampel memenuhi

$$n = \left(\frac{z_{\alpha/2}\sigma}{e}\right)^2$$

Bukti. Akan dicari nilai n sehingga memenuhi $z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = e$ dengan $e = |\bar{x} - \mu|$. Dengan memindahkan ruas maka teorema dapat terbukti. ■

Contoh Soal 2. Tentukan ukuran sampel yang dibutuhkan agar galat taksiran rataan yang diinginkan pada kasus contoh soal 1 lebih kecil daripada 0.05.

Solusi. Dengan menggunakan teorema 2 kita dapatkan

$$n = \left(\frac{(1.96)(0.3)}{0.05}\right)^2 = 138.3$$

Agar galat yang dihasilkan lebih kecil, kita ambil jumlah sampel sebanyak 139 buah (lakukan pembulatan ke atas). ■

2. **Variansi populasi tidak diketahui.**

   a) *Untuk* $n \geq 30$ ingat kembali Teorema Limit Pusat. Karena variansi populasi tidak diketahui, kita dapat menggunakan variansi sampel sebagai penaksir yang baik bagi variansi populasi sehingga kita dapatkan selang kepercayaannya adalah:

   $$\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}$$

   b) *Untuk* $n < 30$ kita tidak dapat menggunakan Teorema Limit Pusat dikarenakan ukuran sampel yang kecil. Namun, kita dapat gunakan distribusi *t-Student* untuk membentuk selang kepercayaan rataan bagi variansi populasi yang tidak diketahui. Adapun statistik bagi distribusi t-Student sebagai berikut:

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

   dengan s sebagai standar deviasi sampel (terlihat perbedaannya dengan statistik Z pada bagian sebelumnya). Dengan menggunakan prosedur yang sama, didapat selang kepercayaan untuk rataan satu populasi sebagai berikut:

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

dengan $\nu = n - 1$ sebagai derajat kebebasan untuk distribusi t-Student. Nilai untuk titik kritis distribusi ini dapat dilihat pada lampiran 2. Dalam kasus ini, α bertindak sebagai luas dari sebelah kanan pada kurva distribusi t-Student dengan derajat kebebasan ν.

Contoh Soal 3. Volume dari 7 botol larutan asam sulfat adalah: 9.8, 10.2, 10.4, 9.8, 10.0, 10.2, dan 9.6 liter. Tentukan selang kepercayaan 95% untuk rataan volume larutan asam sulfat tersebut.
Solusi. Pertama-tama kita tentukan terlebih dahulu rataan dan standar deviasi dari sampel yang diberikan.

$$\bar{x} = 10.0 \qquad\qquad s = 0.283$$

Dengan menggunakan tabel pada lampiran 2, didapat $t_{0.025,6} = 2.447$ karena $\nu = n - 1 = 6$. Selang kepercayaannya menjadi:

$$10.0 - 2.447\frac{0.283}{\sqrt{7}} < \mu < 10.0 - 2.447\frac{0.283}{\sqrt{7}} \Leftrightarrow 9.74 < \mu < 10.26$$

■

7.2.2 Dua Populasi

Dalam penaksiran dua populasi, kita menaksir selisih dari rataan populasi pertama dan kedua $(\mu_1 - \mu_2)$. Untuk kategorinya kita bagi sebagai berikut:

1. **Variansi kedua populasi diketahui**. Seperti pada satu populasi, kita gunakan kembali distribusi normal dengan statistik sebagai berikut:

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Ilustrasi dari statistik tersebut adalah sebagai berikut: karena kasus menyangkut dua populasi, maka rataan antar populasi kita kurangkan karena mengikuti sifat kelinearan dari rataan (lihat kembali topik mengenai sifat ekspektasi) namun standar deviasi menjadi ditambahkan karena keragaman dua populasi tentu berasal dari keragaman dari tiap-tiap populasinya (lihat kembali topik mengenai sifat variansi).

Dengan menggunakan prosedur yang sama, kita dapatkan selang kepercayaan untuk selisih rataan dua populasi tidak berpasangan adalah sebagai berikut:

$$(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Contoh Soal 4. Sebuah studi dilakukan untuk membandingkan efisiensi bensin (dalam satuan mil per galon) tipe A dan B. Lima puluh eksperimen dilakukan untuk bensin tipe A dan tujuh pulh lima percobaan dilakukan untuk bensin tipe B. Didapat rataan efisiensi untuk bensin tipe A adalah 36 mil per galon dan bensin tipe B adalah 42 mil per galon. Tentukan selang kepercayaan 96% untuk selisih rataan bensin tipe B terhadap bensin tipe A dengan informasi bahwa standar deviasi populasi bensin tipe A dan B berturut-turut sebesar 6 dan 8.
Solusi. Kita cari informasi yang dibutuhkan terlebih dahulu. Diketahui $\bar{x}_B - \bar{x}_A = 42 - 36 = 6$ dan $\alpha = 0.04$ sehingga nilai $z_{0.02} = 2.05$. Dengan menggunakan definisi yang diberikan maka selang kepercayaan yang terbentuk adalah:

$$6 - 2.05\sqrt{\frac{36}{50} + \frac{64}{75}} < \mu_B - \mu_A < 6 + 2.05\sqrt{\frac{36}{50} + \frac{64}{75}} \Leftrightarrow 3.43 < \mu_B - \mu_A < 8.57$$

■

2. **Variansi kedua populasi tidak diketahui**. Kita bagi kategori untuk bagian ini berdasarkan asumsi.

   a) *Asumsi variansi kedua populasi berbeda.* Diberikan asumsi bahwa variansi kedua populasi sama, namun tidak diketahui ($\sigma_1^2 = \sigma_2^2 = \sigma^2$), kita gunakan statistik t-Student yang diperluas:

   $$T = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

   Sehingga didapatkan selang kepercayaannya:

   $$(\bar{x}_1 - \bar{x}_2) - t_{\frac{\alpha}{2},\nu}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\frac{\alpha}{2},\nu}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

   dengan $t_{\frac{\alpha}{2},\nu}$ mempunyai derajat kebebasan:

   $$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{(s_1^2/n_1)^2}{n_1-1}\right) + \left(\frac{(s_2^2/n_2)^2}{n_2-1}\right)}$$

   Karena derajat kebebasan ν haruslah bilangan bulat, maka jika didapat bilangan desimal, ν *harus dibulatkan ke bilangan bulat terdekat.*

   Catatan: Asumsi ini digunakan apabila tidak terdapat keterangan apapun dalam soal mengenai variansi populasi (karena kemungkinan dua populasi berbeda mempunyai variansi populasi yang sama adalah sangat kecil).

   b) *Asumsi variansi kedua populasi sama.* Hal yang berbeda terjadi pada bagian ini. Kita mulai dari statistik distribusi Normal.

   $$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

   Karena $\sigma_1^2 = \sigma_2^2 = \sigma^2$ maka statistik dapat dituliskan sebagai berikut:

   $$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

   Parameter σ kita ubah menjadi parameter s_p (prosedur pengubahan dapat dilihat di teorema dalam buku teks bab 8) sehingga statistik yang digunakan adalah:

   $$T = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \text{dengan} s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

   Parameter s_p^2 seakan-akan menjadi "rata-rata berbobot" dari variansi populasi yang bersangkutan (karena asumsi variansi yang sama). Selanjutnya didapat selang kepercayaan:

   $$(\bar{x}_1 - \bar{x}_2) - t_{\frac{\alpha}{2},\nu}s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\frac{\alpha}{2},\nu}s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

   dengan s_p seperti persamaan yang diberikan sebelumnya dan $\nu = n_1 + n_2 - 2$.

7.2.3 Dua Populasi Berpasangan

Selanjutnya akan dibahas terlebih dahulu mengenai pengamatan berpasangan. Berbeda dengan kasus sebelumnya, di mana dua populasi yang dibandingkan rataannya tidak "berhubungan", dalam populasi berpasangan terdapat ciri-ciri sebagai berikut:

- Setiap satu percobaan/observasi mempunyai sepasang pengamatan.
- Data berasal dari satu populasi yang sama.

Sebagai contoh: nilai tes awal dan tes akhir suatu mata pelajaran, pemberian serum A dan serum B pada pasien penyakit tertentu, dan lain-lain. Seperti yang terlihat pada contoh, data berpasangan dapat berasal dari satu populasi yang dikenakan dua perlakuan yang berbeda, serta data yang dihasilkan berpasangan.

Untuk mencari selang kepercayaan bagi selisih rataan dua populasi berpasangan, kite perlu mendefinisikan terlebih dahulu populasi baru, yaitu populasi d dengan data-datanya sebagai berikut:

$$d_i = X_{1i} - X_{2i}$$

dengan X_1 dan X_2 berturut-turut sebagai sampel dari populasi 1 dan populasi 2 berukuran n dan $i = 1, 2, \cdots, n$. Maksud dari persamaan tersebut adalah *kurangkan setiap data pada sampel populasi 1 dengan setiap data pada sampel populasi 2.*

Setelah itu, kita gunakan selang kepercayaan:

$$\bar{d} - t_{\frac{\alpha}{2},\nu}\frac{s_d}{\sqrt{n}} < \mu < \bar{d} + t_{\frac{\alpha}{2},\nu}\frac{s_d}{\sqrt{n}}$$

dengan $\bar{d}$ sebagai rataan sampel d, s_d sebagai standar deviasi sampel d, serta $\nu = n - 1$ sebagai derajat kebebasan distribusi t-Student.

Contoh Soal 5. Suatu pengamatan dilakukan untuk mengetahui tingkat dioksin TCDD dari 20 warga Vietnam. Tingkat TCDD dalam plasma darah dan dalam jaringan lemak disajikan dalam tabel berikut:

Plasma	Lemak
2.5	4.9
3.1	5.9
2.1	4.4
3.5	6.9
3.1	7.0
1.8	4.2
6.0	10.0
3.0	5.5
36.0	41.0
4.7	4.4

Plasma	Lemak
6.9	7.0
3.3	2.9
4.6	4.6
1.6	1.4
7.2	7.7
1.8	1.1
20.0	11.0
2.0	2.5
2.5	2.3
4.1	2.5

Tentukan selang kepercayaan 95% bagi selisih rataan tingkat TCDD di kedua jaringan tersebut.

Solusi. Kita bentuk terlebih dahulu sampel d sebagai selisih dari tiap pasangan data sebagai berikut:

-2.4	-2.8	-2.3	-3.4	-3.9	-2.4	-4.0	2.5	-5.0	0.3
-0.1	0.4	0.0	0.2	-0.5	0.7	9.0	-0.5	0.2	1.6

Dari sampel tersebut, kita cari rataan dan standar deviasinya dan didapat $\bar{d} = -0.87$ dan $s_d = 2.9773$. Untuk $\alpha = 5\%$ dan $\nu = 19$ kita dapatkan nilai $t_{0.025,19} = 2.093$. Sehingga didapat selang kepercayaannya:

$$-0.87 - (2.093)\frac{2.9773}{\sqrt{20}} < \mu_D < -0.87 + (2.093)\frac{2.9773}{\sqrt{20}} \Leftrightarrow -2.2624 < \mu_D < 0.5234$$

■

7.3 Selang Kepercayaan untuk Variansi

7.3.1 Satu Populasi

Dalam satu populasi, selang kepercayaan untuk variansi (ragam) dapat ditaksir dengan menggunakan statistik khi kuadrat.

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

Dengan prosedur yang sama kita dapatkan selang kepercayaan untuk variansi satu populasi sebagai berikut:

$$\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},\nu}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},\nu}}$$

dengan $\nu = n-1$ menunjukkan derajat kebebasan untuk Distribusi Khi Kuadrat, dan nilai $\chi^2_{\frac{\alpha}{2},\nu}$ dan $\chi^2_{1-\frac{\alpha}{2},\nu}$ dapat dilihat pada lampiran 3.

Contoh Soal 6. Data berikut merupakan berat (dalam gram) dari 10 paket biji dalam suatu perusahaan: 46.4, 46.1, 45.8, 47.0, 46.1, 45.9, 45.8, 46.9, 45.2, dan 46.0. Tentukan selang kepercayaan 95% bagi variansi data tersebut. Asumsi data berat biji berdistribusi normal.

Solusi. Dari data, kita tentukan variansi bagi sampel

$$s^2 = \frac{\sum_{i=1}^{10}(x_i - \bar{x})^2}{n-1} = 0.286$$

Gunakan $\nu = 10 - 1 = 9$ dan $\alpha = 5\%$ sehingga didapat nilai $\chi^2_{0.975,9} = 2.700$ dan $\chi^2_{0.025,9} = 19.023$ dari lampiran 3. Kita bentuk selang kepercayaannya menjadi:

$$\frac{(9)(0.286)}{19.023} < \sigma^2 < \frac{(9)(0.286)}{2.700} \Leftrightarrow 0.135 < \sigma^2 < 0.953$$

■

7.3.2 Dua Populasi

Terakhir untuk dua populasi, kita akan membandingkan variansinya (rasio variansi). Kita gunakan statistik bagi Distribusi Fisher:

$$F = \frac{\sigma_1^2/\sigma_2^2}{s_1^2/s_2^2}$$

Dengan prosedur yang sama maka didapat selang kepercayaan bagi rasio variansi sebagai berikut:

$$\frac{s_1^2}{s_2^2}\frac{1}{f_{\frac{\alpha}{2},\nu_1,\nu_2}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2}\frac{1}{f_{1-\frac{\alpha}{2},\nu_1,\nu_2}}$$

Salah satu ciri khas bagi nilai distribusi Fisher adalah sebagai berikut:

$$f_{1-\frac{\alpha}{2},\nu_1,\nu_2} = \frac{1}{f_{\frac{\alpha}{2}}, \nu_2, \nu_1}$$

Sehingga kita dapat substitusikan ke dalam selang kepercayaan menghasilkan:

$$\frac{s_1^2}{s_2^2}\frac{1}{f_{\frac{\alpha}{2},\nu_1,\nu_2}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} f_{1-\frac{\alpha}{2},\nu_2,\nu_1}$$

Dengan $\nu_1 = n_1 - 1$, $\nu_2 = n_2 - 1$, dan nilai $f_{\frac{\alpha}{2},\nu_1,\nu_2}$ dapat dilihat dalam lampiran 4 dan 5.

Contoh Soal 7. Studi dalam program studi Zoologi dalam suatu universitas dilakukan untuk mengetahui perbedaan ragam dari banyaknya senyawa orthophosphorus (dalam mg per liter) dari dua tempat

di suatu sungai. Lima belas sampel diambil pada daerah I dan 12 sampel dari daerah II. Dalam daerah I, didapat standar deviasi sebesar 3.07 dan daerah II sebesar 0.80. Tentukan selang kepercayaan 98% bagi rasio variansi daerah I banding daerah II.
Solusi. Dari tabel lampiran 5 kita dapatkan $f_{0.01,14,11} \approx 4.30$ dan $f_{0.01,11,14} \approx 3.87$. Selang kepercayaannya adalah:

$$\frac{3.07^2}{0.80^2}\frac{1}{4.30} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{3.07^2}{0.80^2}3.87 \Leftrightarrow 3.425 < \frac{\sigma_1^2}{\sigma_2^2} < 56.991$$

■

7.4 *Soal Latihan*

1. Dua puluh siswa dibagi menjadi 10 pasang, tiap orang dalam pasangan mempunyai kecerdasan yang hampir sama. Seorang dari tiap pasangan dipilih secara acak dan dimasukkan ke kelompok perlakuan I. Anggota lainnya dimasukkan ke kelompok perlakuan II. Data hasil eksperimen tercatat:

$$(x_1, x_2) : (76, 81)(60, 52)(85, 87)(58, 70)(91, 86)(75, 77)(82, 90)(64, 63)(79, 85)(88, 83)$$

   a) Tentukan taksiran titik.
   b) Tentukan 98% selang kepercayaan selisih mean perlakuan I dan II.

2. Data berikut menyatakan banyaknya ngengat yang tertangkap pada malam hari, masing-masing dengan cara A dan B.

A	42	35	34	37	41	37		
B	53	58	63	56	65	58	57	57

   Tentukan selang kepercayaan 95% untuk selisih kedua rataan dan rasio variansi antara cara A dengan cara B.

3. Seorang peneliti di UCLA menyatakan bahwa umur tikus dapat diperpanjang 25% bila kalori dalam makanannya dikurang sebanyak 40%, sejak tikus itu disapih. Sampai tahap normal makanan tersebut kemudian diperkaya dengan vitamin dan protein. Dari penelitian terdahulu diketahui bahwa simpangan baku umur tikus dengan pola makan seperti itu adalah 5.8 bulan. Peneliti lain melakukan percobaan terhadap beberapa ekor tikus yang diambil secara acak dari populasi tikus tersebut dan ia ingin rataan umur tikus-tikus tersebut lebih lama 2 bulan dari rataan populasi. Jika tingkat keyakinan yang ia gunakan sebesar 99%, berapa ekor tikus yang harus dijadikan sampel?

Bab 8

Statistika Inferensi: Uji Hipotesis

Setelah kita mempelajari mengenai penaksiran, pada bab ini kita akan mempelajari mengenai pengecekan suatu pernyataan. Dalam penaksiran selang, kita mendapatkan nilai suatu parameter dalam sebuah selang. Dalam uji hipotesis, kita akan mengecek apakah pernyataan yang diberikan berada pada selang penolakan atau tidak dalam tingkat signifikansi yang diberikan.

8.1 Konsep Dasar dan Langkah Uji Hipotesis

Kita mulai dengan mendefinisikan istilah-istilah yang digunakan dalam uji hipotesis. Hipotesis adalah suatu anggapan yang mungkin benar atau salah mengenai satu populasi atau lebih yang perlu diuji kebenarannya. Dalam uji hipotesis dikenal hipotesis nol dan hipotesis tandingan.

- Hipotesis nol (H_0) adalah pernyataan yang mengandung tanda kesamaan. H_0 berupa hipotesis yang ingin diuji seperti hasil penelitian sebelumnya, informasi dari buku, atau hasil percobaan orang lain.
- Hipotesis tandingan (H_1) adalah tandingan dari hipotesis nol yang mengandung tanda ketaksamaan. Hipotesis tandingan merupakan hipotesis yang ingin dibuktikan kebenarannya.

Menurut daerah penolakannya, H_0 dan H_1 dibedakan menjadi satu arah dan dua arah. Misal kita ingin menguji suatu parameter θ bernilai θ_0 maka hipotesis untuk uji dua arah adalah sebagai berikut:

$$H_0 : \theta = \theta_0 \qquad\qquad H_1 : \theta \neq \theta_0$$

sedangkan uji satu arah adalah sebagai berikut:

$$H_0 : \theta = \theta_0 \qquad\qquad H_1 : \theta < \theta_0 \text{ atau } \theta > \theta_0$$

Selanjutnya kita akan pelajari konsep galat yang dapat terjadi dalam uji hipotesis. Perhatikan tabel berikut:

	H_0 benar	H_0 salah
H_0 tidak ditolak	Keputusan tepat	Galat tipe II (β)
H_0 ditolak	Galat tipe I (α)	Keputusan tepat

Galat tipe I terjadi ketika H_0 ditolak padahal H_0 benar. Peluang terjadinya hal tersebut dinotasikan dengan α (tingkat signifikansi). Galat tipe II terjadi ketika H_0 tidak ditolak padahal H_0 salah dengan notasi untuk peluangnya β. Dalam bab ini kita lebih fokus kepada galat tipe I.

Catatan: dalam bab ini, tidak ada istilah bahwa H_0 diterima karena kita tidak memperhitungkan galat tipe II pada saat uji hipotesis.

Contoh Soal 1. Suatu vaksin diketahui 25% efektif setelah 2 tahun. Untuk menentukan apakah vaksin itu efektif, dipilih 20 orang secara acak dan diberikan vaksin tersebut. Jika lebih dari 8 orang tidak terkena penyakit selama lebih dari 2 tahun, maka vaksin dinyatakan efektif. Tentukan hipotesis nol dan tandingan dari kasus tersebut. Tentukan pula peluang dari galat tipe I dan galat tipe II.

Solusi. Misal X peubah acak yang menyatakan banyak orang yang tidak terkena penyakit. Hipotesis yang digunakan adalah proses binomial dengan kemungkinan sukses 0.25 sebagai hipotesis nol dan kemungkinan sukses lebih dari 0.25 untuk hipotesis tandingan (karena diinginkan 8 orang tidak terkena penyakit sedangkan seperempat dari 20 orang adalah 5 orang, sehingga 8 lebih dari 5 orang).

$$H_0 : p = 0.25 \qquad\qquad H_1 : p > 0.25$$

Galat tipe I memiliki peluang sebesar:

$$\alpha = P\left(X > 8 \text{ saat } p = \frac{1}{4}\right) = \sum_{x=9}^{20} b(x; 20, 0.25) = 0.0409$$

Asumsikan $p > 0.25$ dalam galat tipe II berarti $p = 0.5$, maka peluang bagi galat tipe II adalah:

$$\beta = P\left(X \leq 8 \text{ saat } p = \frac{1}{2}\right) = \sum_{x=0}^{8} b(x; 20, 0.5) = 0.2517$$

■

Selanjutnya, dalam melakukan uji hipotesis, sebenarnya kita mengecek apakah klaim yang diberikan berada dalam selang kepercayaan atau tidak (tanpa mencari selang kepercayaannya terlebih dahulu). Maka dari itu, kita definisikan statistik uji sebagai penentu keputusan dalam uji hipotesis. Secara umum, kita definisikan pula daerah kritis sebagai daerah penolakan H_0. H_0 ditolak apabila statistik uji berada di dalam daerah kritis. Perbatasan antara daerah kritis dengan daerah penerimaan disebut dengan titik kritis.

Secara umum, langkah-langkah uji hipotesis adalah sebagai berikut:

1. Tentukan H_0 dan H_1.
2. Tentukan α dan jenis populasi.
3. Tentukan statistik uji.
4. Tentukan daerah kritis.
5. Tentukan apakah statistik uji berada dalam daerah penolakan atau tidak.
6. Berikan kesimpulan non matematis.

Di sisi lain, pengambilan keputusan H_0 dapat menggunakan cara lain, yaitu dengan menghitung *P-Value*. *P-Value* adalah nilai batas penolakan H_0. H_0 ditolak jika *P-Value* lebih besar dari α. Metode ini memudahkan kita karena kita tidak perlu mencari statistik uji dan daerah kritisnya. Namun, untuk mencari *P-Value* kita membutuhkan program-program statistika. Fokus dari bab ini adalah metode uji hipotesis secara umum.

8.2 Statistik Uji dan Daerah Kritis untuk Uji Hipotesis

Secara umum, pembagian kategori untuk uji hipotesis sama seperti pada penaksiran.

8.2.1 Pengujian Rataan

8.2.1.1 Satu Populasi

Untuk rataan satu populasi, kita bagi menjadi dua kategori yaitu variansi populasi diketahui dan tidak diketahui.

- **Variansi populasi diketahui**. Untuk $H_0 : \mu = \mu_0$ kita gunakan Distribusi Normal dengan statistik uji:
$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$
dan daerah kritis untuk uji dua arah: $Z < -z_{\alpha/2}$ atau $Z > z_{\alpha/2}$ serta untuk satu arah: $Z < -z_\alpha$ untuk $H_1 : \mu < \mu_0$ dan $Z > z_\alpha$ untuk $H_1 : \mu > \mu_0$.

- **Variansi populasi tidak diketahui**. Untuk $H_0 : \mu = \mu_0$ kita gunakan Distribusi t-Student dengan statistik uji:
$$T = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$
dan daerah kritis untuk uji dua arah: $T < -t_{\frac{\alpha}{2},\nu}$ atau $T > t_{\frac{\alpha}{2},\nu}$ serta untuk satu arah: $T < -t_{\alpha,\nu}$ untuk $H_1 : \mu < \mu_0$ dan $T > t_{\alpha,\nu}$ untuk $H_1 : \mu > \mu_0$ dengan derajat kebebasan $\nu = n - 1$.

Rangkumannya dapat dilihat pada tabel berikut:

	σ^2 diketahui	σ^2 tidak diketahui
Statistik Uji	$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$	$T = \frac{\bar{x} - \mu}{s/\sqrt{n}}$
$H_1 : \mu \neq \mu_0$	$Z < -z_{\alpha/2}$ atau $Z > z_{\alpha/2}$	$T < -t_{\frac{\alpha}{2},n-1}$ atau $T > t_{\frac{\alpha}{2},n-1}$
$H_1 : \mu > \mu_0$	$Z < -z_{\alpha}$	$T < t_{\alpha,n-1}$
$H_1 : \mu < \mu_0$	$Z > z_{\alpha}$	$T < -t_{\alpha,n-1}$

Contoh Soal 2.

Berdasarkan 100 laporan kematian di AS yang diambil secara acak, diperoleh bahwa rata-rata usia saat meninggal adalah 71.8 tahun dengan simpangan baku 8.9 tahun. Berdasarkan literatur diduga bahwa rata-rata usia meninggal di AS lebih dari 70 tahun. Untuk tingkat signifikansi 5%, benarkah dugaan tersebut?

Solusi. Tentukan hipotesisnya terlebih dahulu:
$$H_0 : \mu = 70 \qquad\qquad H_1 = \mu > 70$$
dengan $\alpha = 0.05$ dan data satu populasi dengan variansi tidak diketahui, diketahui titik kritisnya $t_{0.05,99} = 1.645$. Statistik ujinya sebagai berikut:
$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = 2.02$$
Karena $t > t_{0.05,99}$ maka t berada dalam daerah penolakan. Jadi, dugaan tersebut benar bahwa rata-rata usia meninggal di AS lebih dari 70 tahun. ■

8.2.1.2 Dua Populasi Tidak Berpasangan

Sama seperti pada penaksiran, dua populasi tidak berpasangan dibagi menjadi 3 kategori:

- **Variansi populasi diketahui.** Untuk $H_0 : \mu_1 - \mu_2 = \mu_0$ kita gunakan Distribusi Normal dengan statistik uji:
$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - \mu_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$
dan daerah kritis untuk uji dua arah: $Z < -z_{\alpha/2}$ atau $Z > z_{\alpha/2}$ serta untuk satu arah: $Z < -z_{\alpha}$ untuk $H_1 : \mu_1 - \mu_2 < \mu_0$ dan $Z > z_{\alpha}$ untuk $H_1 : \mu_1 - \mu_2 > \mu_0$.

- **Variansi populasi tidak diketahui.**
  - *Asumsi variansi sama.* Untuk $H_0 : \mu_1 - \mu_2 = \mu_0$ kita gunakan Distribusi t-Student dengan statistik uji:
$$T = \frac{(\bar{x}_1 - \bar{x}_2) - \mu_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \text{ dengan } s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$
dan daerah kritis untuk uji dua arah: $T < -t_{\frac{\alpha}{2},\nu}$ atau $T > t_{\frac{\alpha}{2},\nu}$ serta untuk satu arah: $T < -t_{\alpha,\nu}$ untuk $H_1 : \mu_1 - \mu_2 < \mu_0$ dan $T > t_{\alpha,\nu}$ untuk $H_1 : \mu_1 - \mu_2 > \mu_0$ dengan derajat kebebasan $\nu = n_1 + n_2 - 2$.

– *Asumsi variansi tidak sama.* Untuk $H_0 : \mu_1 - \mu_2 = \mu_0$ kita gunakan Distribusi t-Student dengan statistik uji:

$$T = \frac{(\bar{x}_1 - \bar{x}_2) - \mu_0}{s_p\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

dan daerah kritis untuk uji dua arah: $T < -t_{\frac{\alpha}{2},\nu}$ atau $T > t_{\frac{\alpha}{2},\nu}$ serta untuk satu arah: $T < -t_{\alpha,\nu}$ untuk $H_1 : \mu_1 - \mu_2 < \mu_0$ dan $T > t_{\alpha,\nu}$ untuk $H_1 : \mu_1 - \mu_2 > \mu_0$ dengan derajat kebebasan:

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{(s_1^2/n_1)^2}{n_1-1}\right) + \left(\frac{(s_2^2/n_2)^2}{n_2-1}\right)}$$

Rangkuman tersaji dalam tabel berikut:

Kondisi	σ_1^2 dan σ_2^2 diketahui	σ_1^2 dan σ_2^2 tidak diketahui	
Asumsi	-	$\sigma_1^2 = \sigma_2^2 = \sigma^2$	$\sigma_1^2 \neq \sigma_2^2$
Statistik uji	$Z = \frac{(\bar{x}_1 - \bar{x}_2) - \mu_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	$T = \frac{(\bar{x}_1 - \bar{x}_2) - \mu_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	$T = \frac{(\bar{x}_1 - \bar{x}_2) - \mu_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
$H_1 : \mu \neq \mu_0$	$Z < -z_{\alpha/2}$ atau $Z > z_{\alpha/2}$	$T < -t_{\frac{\alpha}{2},\nu}$ atau $T > t_{\frac{\alpha}{2},\nu}$	$T < -t_{\frac{\alpha}{2},\nu}$ atau $T > t_{\frac{\alpha}{2},\nu}$
$H_1 : \mu > \mu_0$	$Z < -z_\alpha$	$T < t_{\alpha,\nu}$	$T < t_{\alpha,\nu}$
$H_1 : \mu < \mu_0$	$Z > z_\alpha$	$T < -t_{\alpha,\nu}$	$T < -t_{\alpha,\nu}$
Derajat kebebasan	-	$\nu = n_1 + n_2 - 2$	$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{(s_1^2/n_1)^2}{n_1-1}\right) + \left(\frac{(s_2^2/n_2)^2}{n_2-1}\right)}$
Keterangan	$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$		

Contoh Soal 3. Suatu percobaan dilakukan untuk membandingkan keausan yang diakibatkan oleh gosokan, dari dua bahan yang dilapisi. Dua belas potong bahan 1 diuji dengan memasukan tiap potong bahan kedalam mesin pengukur aus. Sepuluh potong bahan 2 diuji dengan cara yang sama. Dalam tiap hal, diamati dalamnya keausan. Sampel bahan 1 memberikan rata-rata keausan (sesudah disandi) sebanyak 85 satuan dengan simpangan baku sampel 4, sedangkan sampel bahan 2 memberikan rata-rata keausan sebanyak 81 dengan simpangan baku sampel 5. Dapatkah disimpulkan, pada taraf keberartian 5%, bahwa rata-rata keausan bahan 1 melampaui rata-rata keausan bahan 2 lebih dari dua satuan? Anggaplah kedua populasi berdistribusi hampir normal dengan variansi yang sama.

Solusi. Misal μ_1 dan μ_2 masing-masing menyatakan rataan bagi populasi 1 dan 2. Hipotesis yang diuji adalah:

$$H_0 : \mu_1 - \mu_2 = 2 \qquad\qquad H_1 : \mu_1 - \mu_2 > 2$$

dengan $\alpha = 5\%$ dan data dua populasi dengan variansi tidak diketahui namun asumsikan variansi kedua populasi sama. Statistik ujinya adalah:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \mu_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = 1.04 \text{ dengan } s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = 4.478$$

Didapat titik kritis $t_{0.05,20} = 1.725$. Karena $t < 1.725$ maka H_0 tidak ditolak. Jadi, belum cukup bukti bahwa rata-rata keausan 1 melampaui rata-rata keausan 2 lebih dari 2 satuan. ■

8.2.1.3 Dua Populasi Berpasangan

Seperti biasa dalam populasi yang berpasangan, kita kurangkan setiap elemen data pada masing-masing sampel dan bentuk sampel d yang kita hitung $\bar{d}$ dan s_d nya. Setelah itu, gunakan statistik uji:

$$T = \frac{\bar{d} - \mu_0}{s_d/\sqrt{n}}$$

dan daerah kritis untuk uji dua arah: $T < -t_{\frac{\alpha}{2},\nu}$ atau $T > t_{\frac{\alpha}{2},\nu}$ serta untuk satu arah: $T < -t_{\alpha,\nu}$ untuk $H_1 : \mu_D < \mu_0$ dan $T > t_{\alpha,\nu}$ untuk $H_1 : \mu_D > \mu_0$ dengan derajat kebebasan $\nu = n - 1$.

Contoh Soal 4. Pada tahun 1976, J.A. Weson memeriksa pengaruh obat *succinylcholine* terhadap kadar peredaran hormon androgen dalam darah. Sampel darah dari rusa liar yang hidup bebas diambil melalui urat nadi leher segera setelah *succinylcholine* disuntikkan pada otot rusa. Rusa kemudian diambil lagi darahnya kira-kira 30 menit setelah suntikan dan kemudian rusa tersebut dilepaskan. Kadar androgen pada waktu ditangkap dan 30 menit kemudian diukur dalam nanogram per ml (ng/ml) untuk 15 rusa. Data terdapat pada tabel berikut:

Sesaat	2.76	5.18	2.68	3.05	4.10	7.05	6.60	4.79	7.39	7.30	11.78	3.90	26.00	67.48	17.04
30 menit	7.02	3.10	5.44	3.99	5.21	10.26	13.91	18.53	7.91	4.85	11.10	3.74	94.03	94.03	41.70

Anggap populasi androgen sesaat setelah suntikan dan 30 menit kemudian berdistribusi normal. Ujilah, pada tingkat keberartian 5%, apakah konsentrasi androgen berubah setelah ditunggu 30 menit.

Solusi. Misal μ_1 dan μ_2 masing-masing menyatakan rataan konsentrasi anrogen sesaat setelah suntikan dan 30 menit kemudian. Rumusan hipotesis yang diuji adalah:

$$H_0 : \mu_D = 0 \qquad\qquad H_1 : \mu_D \neq 0$$

dengan $\alpha = 0.05$ dan jenis data berpasangan. Kita dapatkan selisih dari tiap data sebagai berikut:

4.26	-2.08	2.76	0.94	1.11	3.21	7.31	13.74	0.52	-2.45	-0.68	-0.16	68.03	26.55	24.66

Statistik ujinya adalah:

$$t = \frac{\bar{d} - \mu_0}{s_d/\sqrt{n}} = 2.06$$

Karena t tidak berada pada daerah penolakan, yaitu $t < -t_{0.025,14} = -2.145$ atau $t > t_{0.025,14} = 2.145$, maka H_0 tidak ditolak. Jadi, belum cukup bukti untuk menolak bahwa rata-rata kadar peredaran androgen bisa diabaikan. ■

8.2.2 Pengujian Variansi

8.2.2.1 Satu Populasi

Untuk satu populasi, kita gunakan statistik uji Khi Kuadrat sebagai berikut:

$$X^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

dan daerah kritis untuk uji dua arah: $X^2 < \chi^2_{1-\frac{\alpha}{2},\nu}$ atau $X^2 > \chi^2_{\frac{\alpha}{2},\nu}$ serta untuk satu arah: $X^2 < \chi^2_{1-\alpha,\nu}$ untuk $H_1 : \sigma < \sigma_0$ dan $X^2 > \chi^2_{1-\alpha,\nu}$ untuk $H_1 : \sigma > \sigma_0$ dengan derajat kebebasan $\nu = n - 1$.

Contoh Soal 5. Suatu perusahaan baterai mobil menyatakan bahwa umur baterainya berdistribusi hampir normal dengan simpangan baku 0.9 tahun. Bila sampel acak 10 baterai tersebut menghasilkan simpangan baku 1.2 tahun, apakah anda setuju bahwa $\sigma > 0.9$ tahun? Gunakan taraf kebartian 5%.

Solusi.

$$H_0 : \sigma^2 = 0.81 \qquad\qquad H_1 : \sigma^2 > 0.81$$

dengan $\alpha = 0.05$. Statistik ujinya:

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = 16$$

Titik kritisnya $\chi^2_{0.05,9} = 16.919$. Karena $\chi^2 < \chi^2_{0.05,9}$, maka H_0 tidak ditolak. Jadi simpangan baku umur baterai tidak melebihi 0.9. ∎

8.2.2.2 Dua Populasi

Untuk dua populasi kita gunakan statistik uji Fisher sebagai berikut:

$$F = \frac{\sigma_1^2/s_1^2}{\sigma_2^2/s_2^2}$$

dan daerah kritis untuk uji dua arah: $F < f_{1-\frac{\alpha}{2},\nu_1,\nu_2}$ atau $F > f_{\frac{\alpha}{2},\nu_1,\nu_2}$ serta untuk satu arah: $F < f_{1-\alpha,\nu_1,\nu_2}$ untuk $H_1 : \frac{\sigma_1^2}{\sigma_2^2} < \sigma_0$ dan $F > f_{1-\alpha,\nu_1,\nu_2}$ untuk $H_1 : \frac{\sigma_1^2}{\sigma_2^2} > \sigma_0$ dengan derajat kebebasan $\nu_1 = n_1 - 1$ dan $\nu_2 = n_2 - 1$.

Contoh Soal 6. Dalam pengujian keausan kedua bahan di contoh soal 3, dianggap bahwa kedua variansi yang tidak diketahui sama besarnya. Ujilah anggapan ini. Gunakan taraf keberartian 0.10.
Solusi. Misalkan σ_1^2 dan σ_2^2 adalah variansi populasi dari masing-masing keausan bahan 1 dan bahan 2. Rumusan hipotesis yang akan diuji adalah:

$$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1 \qquad\qquad H_1 : \frac{\sigma_1^2}{\sigma_2^2} \neq 1$$

dengan $\alpha = 0.1$. Statistik ujinya adalah:

$$f = \frac{s_1^2}{s_2^2} = 0.64$$

Diketahui daerah kritisnya adalah $f < f_{0.95,11,9} = 0.34$ atau $f > f_{0.05,11,9} = 3.11$. Karena f tidak berada dalam daerah penolakan, maka H_0 tidak ditolak. Jadi belum cukup bukti untuk menyatakan bahwa variansinya berbeda. ∎

8.3 Soal Latihan

1. Data berikut, dalam hari, menyatakan waktu yang diperlukan penderita sampai sembuh. Penderita dipilih secara acak untuk mendapat salah satu dari dua obat yang dapat menyembuhkan infeksi berat pada saluran kencing.

Obat 1	$n = 14$	$\bar{x}_1 = 17$	$s_1^2 = 1.5$
Obat 2	$n = 16$	$\bar{x}_2 = 19$	$s_2^2 = 1.8$

   Apakah obat 2 memberikan waktu yang lebih lama dari obat 1?

2. Data berikut menyatakan waktu belajar (satuan: menit) peserta kuliah Analisis Data menjelang UTS.

Kelas 01	102.4	96.6	85.3	120.4	114.5	198.8	86.8
Kelas 02	92.3	86.5	88.5	100.1	110.7	94.6	

   a) Apakah rasio keragaman (variansi) waktu belajar mahasiswa kelas 01 dan mahasiswa kelas 02 lebih dari 01? (petunjuk: gunakan $\alpha = 5\%$)
   b) Dengan menggunakan hasil (a), ujilah hipotesis yang mengatakan bahwa tidak ada perbedaan rataan waktu belajar mahasiswa kelas 01 dan mahasiswa kelas 02. Distribusi waktu belajar diasumsikan normal. (petunjuk: gunakan $\alpha = 5\%$)

3. Suatu percobaan untuk menentukan apakah atmosfer yang bercampur CO mempengaruhi kemampuan bernafas dilakukan terhadap 9 orang atlet. Seorang petugas kebugaran atlet menghubungkan setiap peserta dengan dua ruang pernafasan, yang salah satunya mengandung CO yang tinggi; dan ia mengukur pernafasan masing-masing peserta di setiap ruang dengan urutan acak. Data berikut adalah banyaknya pernafasan per menit:

Tanpa CO	32	25	27	37	30	51	43	42	32
Dengan CO	32	53	27	36	47	28	48	34	32

   a) Menurut Anda apakah dua kelompok tersebut berasal dari dua kelompok yang berpasangan atau bukan?
   b) Dengan menggunakan $\alpha = 5\%$ selidikit apakah atmosfer yang bercampur CO meningkatkan banyaknya pernafasan sebesar 2 terhadap atmosfer yang bersih tanpa CO.

Catatan: distribusi banyaknya pernafasan per menit diasumsikan hampir normal.

Bab 9

Analisis Variansi (ANOVA)

Sejauh ini kita sudah dapat menentukan apakah dua populasi mempunyai rataan yang sama atau berbeda secara signifikan. Dalam bab ini, kita akan mempelajari tentang uji hipotesis menyangkut rataan dari tiga populasi atau lebih dengan meninjau sisi variansi tiap populasinya dengan tingkat signifikansi yang diberikan. Metode yang akan kita pelajari adalah ANOVA satu arah (*ANOVA One Way*).

9.1 Ilustrasi dan Asumsi Dalam ANOVA Satu Arah

Kita dapat merepresentasikan data dalam populasi dengan menggunakan histogram. Kita sudah mempelajari sebelumnya bahwa populasi akan terlihat jelas rataannya apabila berdistribusi normal. Untuk membandingkan rataan dari populasi satu dan yang lainnya, kita dapat melihat dari variansi data tersebut. Misalkan ada k populasi. Untuk mengecek kesamaan rataannya, kita gabungkan semua populasinya dan bentuk histogramnya kembali. Jika rataannya sama, seharusnya variansinya tidak semakin besar karena sumbu simetris tiap-tiap populasi tidak jauh berbeda, sehingga bentuk histogramnya "tidak menyebar". Berbeda dengan rataan yang berbeda, ketika kita gabungkan dan bentuk histogram gabungannya, histogram akan "menyebar" yang menyebabkan variansinya membesar.

Dari ilustrasi tersebut, kita gunakan asumsi-asumsi dalam ANOVA sebagai berikut:

(a) Populasi ke-i berdistribusi normal dengan $i = 1, 2, \cdots, k$.

(b) $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2 = \sigma^2$.

(c) Populasi-populasi saling bebas.

Selanjutnya kita lihat susunan data pada ANOVA.

Data	Angkatan (Populasi)			
	1	2	$\cdots$	k
1	y_{11}	y_{21}	$\cdots$	y_{k1}
2	y_{12}	y_{22}	$\cdots$	y_{k2}
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
	$\vdots$	$\vdots$	$\cdots$	y_{kn_k}
	y_{1n_1}	$\vdots$	$\cdots$	
		y_{2n_2}	$\cdots$	
Jumlah	$\sum_{j=1}^{n_1} y_{1j}$	$\sum_{j=1}^{n_2} y_{2j}$	$\cdots$	$\sum_{j=1}^{n_k} y_{kj}$

Keterangan:

- Jumlah semua data adalah $\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}$.

- y_{ij} menyatakan pengamatan ke-j dalam perlakuan (populasi) ke-i.
- N menyatakan total banyak pengamatan ($N = n_1 + n_2 + \cdots + n_k$).

9.2 *Prosedur ANOVA Satu Arah*

Dimulai dengan mendefinisikan hipotesis nol dan tandingan

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k \qquad\qquad H_1 : \mu_i \neq \mu_j, i \neq j$$

Setelah menentukan tingkat signifikansi α, kita tentukan besaran-besaran yang dapat membantu dalam ANOVA sebagai berikut:

$$a = \frac{\left(\sum_{i=1}^{k} \sum_{j=1}^{n_i} y_{ij}\right)^2}{N} \qquad b = \sum_{i=1}^{k} \sum_{j=1}^{n_i} y_{ij}^2 \qquad c = \sum_{i=1}^{k} \frac{\left(\sum_{j=1}^{n_i} y_{ij}\right)^2}{n_i}$$

Setelah itu kita masuk ke tabel ANOVA:

Sumber Variansi	Jumlah Kuadrat	Derajat Kebebasan	Rata-rata Kuadrat	f_{hitung}
Perlakuan	$JK_P = c - a$	$k - 1$	$RK_P = \frac{JK_P}{k-1}$	$f_{\text{hitung}} = \frac{RKP}{RKG}$
Galat	$JK_G = b - c$	$N - k$	$RK_G = \frac{JK_G}{N-k}$	
Total	$JK_T = b - a$	$N - 1$		

Selanjutnya kita masuk ke daerah penolakan. H_0 ditolak jika $f_{\text{hitung}} > f_{\alpha(k-1,N-k)}$ (gunakan tabel lampiran 4 dan 5 untuk mencari nilai $f_{\alpha(k-1,N-k)}$). Jangan lupa untuk memberikan kesimpulan non matematis dari kasus yang diberikan.

Contoh Soal 1. Tabel berikut menyatakan waktu kesembuhan (jam) yang diakibatkan tiga merek obat sakit kepala yang berlainan yang diberikan pada 25 penderita.

Obat	Waktu (jam)								
A	5	4	8	6	3	3	5	2	
B	9	7	8	6	9	3	7	4	1
C	7	6	9	4	7	2	3	4	

Ujilah hipotesis yang menyatakan bahwa rata-rata waktu kesembuhan yang diakibatkan oleh ketiga merek obat tersebut tidak berbeda (gunakan tingkat signifikansi 5%).

Solusi.

$$H_0 : \mu_A = \mu_B = \mu_C \qquad\qquad H_1 : \mu_i \neq \mu_j, i \neq j$$

Dengan $\alpha = 5\%$. Kita tentukan besaran-besaran ANOVA:

$$a = \frac{\left(\sum_{i=1}^{3} \sum_{j=1}^{n_i} y_{ij}\right)^2}{25} = 696.96 \qquad b = \sum_{i=1}^{3} \sum_{j=1}^{n_i} y_{ij}^2 = 834 \qquad c = \sum_{i=1}^{3} \frac{\left(\sum_{j=1}^{n_i} y_{ij}\right)^2}{n_i} = 706.5$$

Setelah itu kita masuk ke tabel ANOVA:

Sumber Variansi	Jumlah Kuadrat	Derajat Kebebasan	Rata-rata Kuadrat	f_{hitung}
Perlakuan	9.54	2	4.77	0.8238
Galat	127.5	22	5.79	
Total	137.04	24		

Didapat titik kritis $f_{0.05(2,22)} = 3.443$. Karena $f_{\text{hitung}} < f_{0.05(2,22)}$ maka H_0 tidak ditolak. Dapat disimpulkan bahwa rata-rata waktu kesembuhan yang diakibatkan oleh ketiga merek obat tersebut tidak berbeda secara signifikan. ■

9.3 Soal Latihan

1. Tiga angkatan Program Studi Matematika mendapatkan nilai Bahasa Inggris dengan silabus pembelajaran yang sama. Data mengenai rata-rata nilai akhir tercatat sebagai berikut:

2007	561	530	397	464	430		
2008	461	464	464	364	697		
2009	530	430	430	430	530	457	457

   Asumsikan distribusi nilai adalah hampir normal, dan tiap angkatan saling bebas. Dengan menggunakan tingkat keberartian 0.01, selidiki apakah terdapat perbedaan yang berarti dalam hasil pembelajaran Bahasa Inggris di 3 angkatan di atas. Berikan tabel ANOVA nya.

2. Seorang ahli biologi ingin melihat pengaruh vaksin dengan dosis berbeda pada seekor tikus (sebut dosis A, dosis B, dan dosis C). Data di bawah ini adalah nilai akhir yang didapat

A	93	97	91	94	
B	80	75	79		
C	68	65	72	75	70

   Diasumsikan bahwa nilai pengaruh vaksin berdistribusi normal dan percobaan dilakukan saling bebas. Gunakan taraf keberartian 0.05 untuk menguji pendapat peneliti lain bahwa dosis vaksin tidak mempunyai pengaruh yang berbeda.

3. Lima metode pemberian pakan diterapkan kepada 16 ekor anak kambing yang dikelompokkan menjadi lima kelompok. Ujilah apakah terdapat perbedaan berat anak kambing yang dihasilkan oleh metode pemberian pakan yang berbeda jika berat anak kambing (dalam kg) yang dihasilkan setelah 2 bulan sejak penerapan metode pemberian pakan tersaji dalam tabel berikut. Gunakan taraf signifikansi 1% dan 5% dan asumsikan bahwa sebelum penerapan metode pemberian pakan yang berbeda, kambing tersebut memiliki berat yang sama.

Perlakuan	Berat (kg)			
A	5.21	4.65		
B	5.59	2.69	7.57	5.16
C	6.24	5.94	6.41	
D	6.85	9.18	4.94	
E	4.04	3.29	4.52	3.75

Bab 10

Regresi Linier Sederhana

Hubungan antara dua variabel dapat membantu kita dalam menentukan kecenderungan terbentuk. Dalam bab ini, kita akan mempelajari tentang bagaimana menentukan hubungan antara dua variabel secara linier melalui metode kuadrat terkecil serta menentukan tingkat kecocokan dari dua variabel tersebut.

10.1 Model Regresi Linier

Kita tentu saja mengenal persamaan garis lurus, $y = ax+b$ dengan a menyatakan gradien dan b menyatakan titik potong garis terhadap sumbu-y. Dalam statistika, kita dpat mencari hubungan antara satu populasi terhadap populasi lainnya melalui bentuk hubungan linier. Pada garis linier $y = ax + b$ kita tahu bahwa nilai x dapat diubah-ubah dan nilai y bergantung pada nilai x. Sama seperti pada model linier dalam statistika, kita tuliskan model liniernya adalah sebagai berikut:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon$$

Dengan $i = 1, 2, \cdots, n$ menyatakan banyak data dari kedua populasi. Variabel x disebut variabel bebas (prediktor) dan variabel y disebut variabel tak bebas (respons) karena bergantung dari variabel x. Karena x variabel bebas, maka x tidak berdistribusi, namun y berdistribusi normal. Nilai ϵ dalam persamaan regresi linier.

Mengapa muncul ϵ dalam model linier? Galat dari model linier berasal dari:

- Ketidakmampuan model regresi dalam memodelkan hubungan prediktor dan respons dengan tepat.
- Ketidakmampuan peneliti dalam melakukan pengukuran dengan tepat.
- Ketidakmampuan model untuk melibatkan semua variabel prediktor.

Namun terkadang kita bingung menentukan siapa yang menjadi respons dan siapa yang menjadi prediktor. Secara umum, populasi yang menjadi prediktor adalah: variabel yang nilainya mempengaruhi variabel yang lainnya, variabel yang kejadiannya lebihi dahulu terjadi, atau variabel yang variansinya terkecil.

10.2 Penaksir Kuadrat Terkecil

Kita tidak dapat menentukan model linier sederhana. Seperti biasa, kita dapat menaksir koefisien-koefisiennya melalui sebuah metode yaitu penaksir kuadrat terkecil (*least square*). Sebelumnya, kita gunakan asumsi-asumsi berikut:

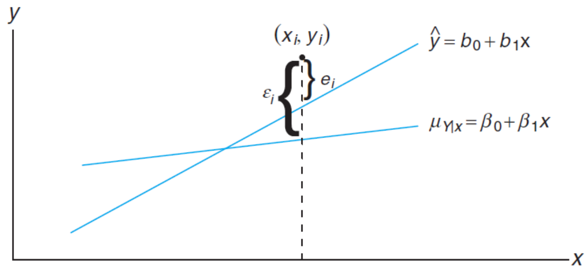
1. Ada pengaruh X terhadap Y.
2. $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ untuk $i \in \mathbb{N}$.
3. $E(\epsilon_i) = 0$.
4. $\mathrm{Var}(\epsilon_i)$ sama untuk semua $i \in \mathbb{N}$.
5. ϵ_i berdistribusi normal untuk semua $i \in \mathbb{N}$.
6. $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ saling bebas.

Prinsip dari metode ini adalah meminimalkan "jarak" antara nilai hasil regresi dengan nilai respons. Kita definisikan taksiran model linier yang akan dibentuk adalah:

$$\hat{y} = b_0 + b_1 x$$

dengan b_0 dan b_1 masing-masing sebagai penaksir bagi β_0 dan β_1.

Kita definisikan galat bagi taksiran model linier sebagai $e_i = y_i - \hat{y}_i$. Ilustrasi galat dapat dilihat pada gambar 10.1.



Gambar 10.1: Ilustrasi Galat Model Linier

Tugas kita sekarang adalah mencari nilai b_0 dan b_1 sehingga jumlah kuadrat galat tersebut minimal. Persamaan bagi jumlah kuadrat galat untuk n pasangan data adalah:

$$JKG = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2$$

Ingat kembali pada kalkulus, untuk mencari nilai maksimum atau minimum dari suatu fungsi, kita buat turunannya menjadi nol. Dalam kasus ini kita lakukan turunan parsial untuk jumlah kuadrat galat masing-masing terhadap b_0 dan b_1. Pertama-tama kita turunkan terhadap b_0 menghasilkan:

$$\begin{aligned} \frac{\partial(JK_G)}{\partial b_0} = 0 &= -2 \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i) \\ &= -2 \left(\sum_{i=1}^{n} y_i - n \cdot b_0 - b_1 \sum_{i=1}^{n} x_i \right) \end{aligned}$$

Selanjutnya turunan parsial terhadap b_1 menghasilkan:

$$\begin{aligned} \frac{\partial(JK_G)}{\partial b_1} = 0 &= -2 \sum_{i=1}^{n} (y_i - b_0 - b_i x_i) x_i \\ &= -2 \left(\sum_{i=1}^{n} x_i y_i - b_0 \sum_{i=1}^{n} x_i - b_1 \sum_{i=1}^{n} x_i^2 \right) \end{aligned}$$

Dengan menyelesaikan kedua persamaan tersebut, kita dapatkan b_0 dan b_1 sesuai dengan teorema berikut.

Teorema 1 Taksiran Koefisien Regresi Linier Sederhana

Diberikan sampel $\{(x_i, y_i); i = 1, 2, \cdots, n\}$, maka taksiran untuk koefisien regresi linier sederhananya adalah:

$$b_1 = \frac{n \sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i \right) \left(\sum_{i=1}^{n} y_i \right)}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2} = \frac{JK_{XY}}{JK_{XX}}$$

$$b_0 = \frac{\sum_{i=1}^{n} y_i - b_1 \sum_{i=1}^{n} x_i}{n} = \bar{y} - b_1\bar{x}$$

dengan besaran $JK_{XY} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$ dan $JK_{XX} = \sum_{i=1}^{n}(x_i - \bar{x})^2$.

Contoh Soal 1. Diberikan tabel reduksi sebuah senyawa kimia dan Oksigen yang dibutuhkan, keduanya dalam satuan persen.

Reduksi (x)	Oksigen (y)
3	5
7	11
11	21
15	16
18	16
27	28
29	27
30	25
30	35
31	30
31	40
32	32
33	34
33	32
34	34
36	37
36	38

Reduksi (x)	Oksigen (y)
36	34
37	36
38	38
39	37
39	36
39	45
40	39
41	41
42	40
42	44
43	37
44	44
45	46
46	46
47	49
50	51

Tentukan taksiran model linier sederhana dari data yang diberikan.

Solusi. Didapat besaran-besaran regresi sebagai berikut:

$$\sum_{i=1}^{33} x_i = 1104 \qquad \sum_{i=1}^{33} y_i = 1124 \qquad \sum_{i=1}^{33} x_i y_i = 41355 \qquad \sum_{i=1}^{33} x_i^2 = 41086$$

Sehingga:

$$b_1 = \frac{(33)(41355) - (1104)(1124)}{(33)(41086) - (1104)^2} = 0.9036 \qquad b_0 = \frac{1124 - (0.9036)(1104)}{33} = 3.8296$$

Taksiran model linier yang didapat adalah $\hat{y} = 3.8296 + 0.9036x$. ■

Setelah kita menemukan taksiran model liniernya, kita bisa menentukan taksiran titik nilai $\hat{y}$ jika diberikan suatu nilai x yang berada pada *range* data tersebut. Hal ini dinamakan *interpolasi*.

Selanjutnya kita akan lihat teorema mengenai taksiran variansi bagi galat, yang akan digunakan untuk inferensi bagi regresi linier.

Teorema 2 Taksiran Variansi Galat

Taksiran variansi bagi galat adalah:

$$s^2 = \frac{JK_G}{n-2} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2} = \frac{JK_{YY} - b_1 JK_{XY}}{n-2}$$

dengan besaran $JK_{XY} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$ dan $JK_{YY} = \sum_{i=1}^{n}(y_i - \bar{y})^2$.

Contoh Soal 2. Tentukan variansi dan simpangan baku galat untuk data pada tabel di contoh soal 1.

Solusi.

$$s^2 = \frac{JK_{YY} - b_1 JK_{XY}}{n-2} = \frac{3713.88 - (0.9036)(3752.09)}{31} = 10.4299$$

$$s = \sqrt{s^2} = \sqrt{10.4299} = 3.2295$$

■

10.3 *Inferensi Parameter Regresi*

Karena kita melakukan penaksiran terhadap koefisien-koefisien pada persamaan regresi, kita tentu dapat menentukan selang kepercayaan bagi gradien maupun titik potong regresi linier sederhana.

10.3.1 *Inferensi Gradien*

Untuk gradien, kita gunakan statistik uji menurut distribusi t-Student sebagai berikut:

$$T = \frac{b_1 - \beta_1}{s/\sqrt{JK_{XX}}}$$

Sehingga menghasilkan selang kepercayaan bagi β_1 adalah:

$$b_1 - t_{\frac{\alpha}{2},\nu}\frac{s}{\sqrt{JK_{XX}}} < \beta_1 < b_1 + t_{\frac{\alpha}{2},\nu}\frac{s}{\sqrt{JK_{XX}}}$$

dengan $\nu = n - 2$ menyatakan derajat kebebasan.

Contoh Soal 3. Tentukan selang kepercayaan 95% bagi gradien pada contoh soal 1.

Solusi. Diketahui melalui lampiran 2 bahwa $t_{0.025,31} \approx 2.045$. Maka selang kepercayaannya adalah:

$$0.9036 - \frac{(2.045)(3.2295)}{\sqrt{4152.18}} < \beta_1 < 0.9036 + \frac{(2.045)(3.2295)}{\sqrt{4152.18}} \Leftrightarrow 0.8012 < \beta_1 < 1.0061$$

■

10.3.2 *Inferensi Titik Potong*

Selanjunya untuk titik potong kita gunakan statistik uji:

$$T = \frac{(b_0 - \beta_0)\sqrt{nJK_{XX}}}{s\sqrt{\sum_{i=1}^{n} x_i^2}}$$

Selang kepercayaan bagi β_0 adalah:

$$b_0 - t_{\frac{\alpha}{2},\nu}\frac{s}{\sqrt{nJK_{xx}}}\sqrt{\sum_{i=1}^{n} x_i^2} < \beta_0 < b_0 + t_{\frac{\alpha}{2},\nu}\frac{s}{\sqrt{nJK_{xx}}}\sqrt{\sum_{i=1}^{n} x_i^2}$$

dengan $\nu = n - 2$ menyatakan derajat kebebasan.

Contoh Soal 4. Uji apakah persamaan regresi pada contoh soal 1 melewati titik pangkal (baca: titik $(0, 0)$) dengan tingkat signifikansi 5%.
Solusi. Soal tersebut bermaksud bahwa kita harus menguji hipotesis sebagai berikut:

$$H_0 : \beta_0 = 0 \qquad\qquad H_1 : \beta_0 \neq 0$$

Pernyataan yang serupa untuk β_0: titik potong bisa diabaikan atau tidak, persamaan regresi mempunyai nilai awal, dan sebagainya.

Pernyataan yang serupa untuk β_1: gradien bisa diabaikan atau tidak, ada hubungan respons dengan prediktor atau tidak, dan sebagainya.

Dengan $\alpha = 5\%$ dan $\nu = 33 - 2 = 31$ kita dapatkan daerah penerimaan $-2.042 < t < 2.042$. Statistik uji yang didapat adalah:

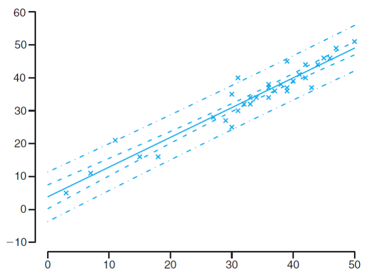
$$t = \frac{(3.8296 - 0)\sqrt{(33)(4152.18)}}{3.2295\sqrt{41086}} = 2.17$$

Karena t berada dalam daerah penolakan, maka kita simpulkan bahwa persamaan regresi tidak melewati titik pangkal. ■

10.4 Prediksi Nilai Respons

Dengan regresi linier sederhana, kita dapat menentukan taksiran dari suatu prediktor ketika diberikan suatu respons yang tidak ada di dalam data (interpolasi). Namun hal tersebut masih berupa suatu taksiran. Seperti pada hal sebelumnya, kita akan melihat taksiran untuk nilai respons. Taksiran berikut dibagi menjadi dua, yaitu prediksi rataan respons dan prediksi respons satuan.

Catatan: Pembuktian statistik uji bagi prediksi respons satuan maupun rataan dapat dilihat di buku Walpole.



Gambar 10.2: Ilustrasi Prediksi Rataan Respons (Ruas Garis Dalam) dan Prediksi Respons Satuan (Ruas Garis Luar)

10.4.1 Prediksi Rataan Respons

Persamaan regresi $\hat{y} = b_0 + b_1 x$ dapat digunakan untuk memprediksi respons rataan $\mu_{Y|x_0}$ di titik $x = x_0$. Maksudnya, karena Y merupakan peubah acak dan X bukan merupakan peubah acak, maka ketika kita memilih suatu nilai x, sebutlah x_0, nilai Y masih beragam. Kita ambil rataannya pada saat $x = x_0$ yang digunakan sebagai suatu nilai prediksi regresi linier. Statistik uji bagi prediksi rataan respons adalah:

$$T = \frac{\hat{Y}_0 - \mu_{Y|x_0}}{s\sqrt{1/n + (x_0 - \bar{x})^2/JK_{XX}}}$$

Sehingga selang kepercayaan yang terbentuk adalah:

$$\hat{y}_0 - t_{\frac{\alpha}{2},\nu} s\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{JK_{XX}}} < \mu_{Y|x_0} < \hat{y}_0 + t_{\frac{\alpha}{2},\nu} s\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{JK_{XX}}}$$

dengan $\nu = n - 2$ merupakan derajat kebebasan bagi t.

10.4.2 Prediksi Respons Satuan

Berbeda dengan prediksi rataan respons, prediksi respons satuan mengambil suatu nilai tunggal pada Y. Karena pengambilan nilai tunggal, maka selang kepercayaan yang terbentuk akan lebih lebar dibandingkan prediksi rataan respons (lihat gambar 10.2). Statistik uji bagi prediksi respons satuan adalah sebagai berikut:

$$T = \frac{\hat{Y}_0 - Y_0}{s\sqrt{1 + 1/n + (x_0 - \bar{x})^2/JK_{XX}}}$$

Sehingga selang kepercayaan yang terbentuk adalah:

$$\hat{y}_0 - t_{\frac{\alpha}{2},\nu} s\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{JK_{XX}}} < y_0 < \hat{y}_0 + t_{\frac{\alpha}{2},\nu} s\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{JK_{XX}}}$$

dengan $\nu = n - 2$ merupakan derajat kebebasan bagi t.

10.5 Kecocokan Model Regresi

10.5.1 Koefisien Determinasi

Alat ukur pertama untuk melihat kecocokan bagi model regresi adalah koefisien determinasi (R^2) yang disajikan melalui persamaan berikut:

$$R^2 = \frac{JK_R}{JK_T} = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

dengan $0 \leq R^2 \leq 1$.

Keterangan: JK_R = jumlah kuadrat regresi dan JK_T = jumlah kuadrat total.

Besaran R^2 menyatakan proporsi variasi total dalam respons Y yang diterangkan oleh model regresi yang diperoleh. Semakin besar R^2, semakin cocok model regresi tersebut dengan data yang diobservasi.

10.5.2 Uji Kebaikan Model

Uji kebaikan model didapat melalui pendekatan inferensi dari gradien melalui ANOVA. Kita tahu bahwa jika $\beta_1 = 0$ maka model regresi tidak memadai (dengan kata lain, tidak terdapat hubungan antara dua variabel). Maka kita rumuskan hipotesis sebagai berikut:

H_0 : model tidak memadai $\qquad$ H_1 : model memadai

Kita gunakan tabel ANOVA berikut:

Sumber Variansi	Jumlah Kuadrat	Derajat Kebebasan	Rata-rata Kuadrat	f_{hitung}
Regresi	JK_R	$2 - 1 = 1$	JK_R	$f_{\text{hitung}} = \frac{JK_R}{s^2}$
Galat	JK_G	$n - 2$	s^2	
Total	JK_T	$n - 1$		

dengan $JK_R = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$. Kita dapatkan statistik uji:

$$f_{\text{hitung}} = \frac{JK_R}{s^2}$$

H_0 ditolak apabila $f_{\text{hitung}} \geq f_{\alpha,1,n-2}$.

10.5.3 Koefisien Korelasi Pearson

Lihat kembali pada bab mengenai ekspektasi bahwa kita sudah mempelajari korelasi dengan persamaan sebagai berikut:

$$\rho = \frac{E((X - \mu_X)(Y - \mu_Y))}{\sigma_X \sigma_Y}$$

ρ disebut sebagai koefisien korelasi Pearson. Besaran ρ digunakan untuk menentukan **hubungan linier** dari dua peubah.

Ingat kembali bahwa nilai korelasi berkisar dari interval $-1 \leq \rho \leq 1$. Semakin dekat nilai ρ ke -1 maka semakin kuat hubungan linier kedua peubah tersebut secara berlawanan arah, dan sebaliknya. Jika nilai ρ mendekati 0, maka tidak terdapat hubungan linier antara dua peubah acak, mungkin saja terdapat hubungan kuadratik, eksponensial, dsb. (belum tentu dua peubah tersebut tidak saling berhubungan).

Taksiran dari korelasi Pearson untuk sampel adalah sebagai berikut:

$$r = \frac{JK_{XY}}{\sqrt{JK_{XX} JK_{XY}}}$$

Sedangkan statistik uji bagi inferensi korelasi Pearson adalah:

$$z = \frac{\sqrt{n-3}}{2} \ln \left(\frac{(1+r)(1-\rho_0)}{(1-r)(1+\rho_0)} \right)$$

Dari statistik uji yang diberikan, Anda bisa menurunkan sendiri selang kepercayaan bagi korelasi Pearson.

10.6 Linearisasi Data

Hubungan antara dua variabel tidak selalu linier. Maka dari itu kita perlu mentransformasikan data terlebih dahulu sehingga membentuk suatu hubungan yang linier. Metode ini disebut dengan *linearisasi data*. Beberapa hubungan dan transformasi data yang sesuai tersaji dalam tabel berikut:

Hubungan	Persamaan	Transformasi
Eksponensial	$y = \beta_0 e^{\beta_1 x}$	$y^* = \ln y$
Perpangkatan	$y = \beta_0 x^{\beta_1}$	$y^* = \log y, x^* = \log x$
Resiprokal	$y = \beta_0 + \frac{\beta_1}{x}$	$y^* = \frac{1}{x}$
Hiperbolik	$y = \frac{x}{\beta_0 + \beta_1 x}$	$y^* = \frac{1}{x}, x^* = \frac{1}{y}$

Apa akibat dari linearisasi data? Galat yang kita dapati dalam persamaan regresi tidak lagi linier, namun mengikuti galat dari transformasi data. Sebagai contoh persamaan eksponensial. Regresi linier yang kita lakukan adalah:

$$\ln y_i = \ln_{\beta_0} + \beta_1 x_i + \epsilon^*$$

dengan mengembalikan ke bentuk eksponensial, maka didapat:

$$y_i = \beta_0 e^{\beta_1 x_i} e^{\epsilon}$$

Perhatikan bahwa galat yang didapat berbentuk eksponensial maka perhitungan bagi variansi galat s^2 dan koefisien determinasi R^2 serta inferensi lainnya menjadi menyesuaikan.

Contoh Soal 5. Data tekanan terhadap volume tersaji dalam tabel berikut:

V (cm^3)	50	60	70	90	100
P (kg/cm^2)	64.7	51.3	40.5	25.9	7.8

Teori gas ideal menyatakan bahwa hubungan tekanan dan volume memenuhi persamaan:

$$PV^\gamma = C$$

dengan γ dan C suatu konstanta. Taksir kedua konstanta tersebut.

Solusi. Model yang diketahui adalah:

$$P_i V_i^\gamma = C \cdot \epsilon_i \qquad\qquad i = 1, 2, 3, 4, 5$$

Gunakan logaritma natural kepada dua sisi sehingga didapat:

$$\ln P_i = \ln C - \gamma \ln V_i + \ln \epsilon_i$$

Kita gunakan regresi linier degan prediktor $\ln V_i$ dan respons $\ln P_i$ untuk mendapatkan gradien $-\gamma$. Transformasi data yang didapat adalah sebagai berikut:

P_i	V_i	$\ln P_i$	$\ln V_i$
64.7	50	4.1697	3.9120
51.3	60	3.9376	4.0943
40.5	70	3.7013	3.2485
25.9	90	3.2542	4.4998
7.8	100	2.0541	4.6051

Dengan menggunakan rumus yang diberikan sebelumnya maka didapat $\ln C = 14.7589$ dan $\gamma = 2.653$. ■

10.7 Soal Latihan

1. Dalam pengujian suatu bahan, tekanan normal yang diberikan terhadap bahan dianggap berkaitan secara fungsional dengan ketahanannya. Berikut hasil percobaan mengenai ketahanan bahan tersebut:

Tekanan	26.8	25.4	28.9	23.6	27.7	23.9	24.7	28.1	26.9
Ketahanan	26.5	27.3	24.2	27.1	23.6	25.9	26.3	22.5	21.7

   a) Taksir model regresi liniernya.
   b) Hitung taksiran simpangan baku galatnya.
   c) Tentukan taksiran selang untuk kemiringannya.
   d) Bila tekanan normal 24,5, apakah benar bahwa perkiraan untuk ketahanannya kurang dari dari 26 untuk taraf keberartian 1%?

2. Seorang pegawai perhutanan ingin menaksir volume (dalam feet^3) pohon-pohon pada suatu penjualan kayu. Dalam penelitiannya ia mengambil seorang asisten dan diminta untuk menaksir volume kayu. Ia memilih secara acak 9 kayu dan mengukur volume sebenarnya, sedangkan asistennya diminta untuk menaksir. Hasilnya sebagai berikut

Taksiran	12	14	8	12	17	16	14	14	15
Aktual	13	14	9	15	19	20	16	15	17

   Jika diambil taksiran sebagai peubah bebas model regresi linier sederhana dan aktual sebagai peubah tak bebas, hitung persamaan regresi dan juga korelasinya.

3. Data berikut diperoleh dalam penelitian hubungan antara berat dan panjang lingkar kepala bayi yang baru lahir.

Berat (kg)	2.75	2.15	4.41	5.52	3.21	4.32	2.31	4.30	3.71
Panjang lingkar kepala (cm)	29.5	26.3	32.3	36.5	27.2	27.7	28.3	30.3	28.7

a) Tentukan titik potong dan gradien (slope) dari taksiran garis regresinya.

b) Hitung taksiran untuk variansi galat.

c) Tentukan selang kepercayaan 95% untuk gradien garis regresinya.

d) Apakah hasil penelitian tersebut sesuai dengan pernyataan "Bila berat bayi tersebut tepat 3 kg, maka panjang lingkar kepalanya diperkirakan sama dengan 29 cm"?

e) Apakah benar bahwa terdapat hubungan antara berat dan panjang lingkar kepala bayi yang baru lahir?

Bab 11

Pengantar Analisis Deret Waktu

Data-data yang kita ambil dari waktu ke waktu dapat memberikan informasi untuk menentukan data yang akan muncul di waktu yang selanjutnya. Dalam bab ini, kita akan mempelajari cara menentukan data untuk waktu yang selanjutnya (*forecasting*) dengan menggunakan analisis deret waktu, berikut data yang dapat kita kenakan analisis deret waktu.

11.1 Terminologi Deret Waktu

11.1.1 Proses Stokastik

Proses stokastik adalah barisan peubah acak $\{Y_t, t \in T\}$. Setiap proses stokastik memuat ruang keadaan S (semua nilai yang mungkin dari Y_t) dan indeks parameter T. Misalkan y_t nilai dari Y_t maka barisan nilai $\{y_t, t \in T\}$ disebut realisasi dari $\{Y_t, t \in T\}$.

Deret waktu adalah suatu proses stokastik. Jika pada proses stokastik $\{Y_t, t \in T\}$ dengan T menyatakan waktu, maka proses tersebut disebut deret waktu. Realisasinya disebut dengan data deret waktu. Bentuk umum model deret waktu adalah:

$$Y_t = T_t \times C_t \times S_t \times I_t$$

Dengan Y_t variabel respon pada waktu t, T_t gerakan umum plot data dalam jangka panjang (trend sekuler), C_t pola data deret waktu yang terjadi dan mengalami perulangan setelah periode waktu tertentu (pergerakan siklus), S_t pola teratur tahunan yang berulang pada tiap tahun (fluktuasi musim), dan I_t komponen tak terduga dan bersifat acak (variasi tak beraturan).

Tujuan deret waktu secara umum adalah memodelkan suatu data deret waktu sehingga dapat terlihat pola dan perilaku data lebih lanjut serta melakukann prediksi jangka pendek.

11.1.2 Kestasioneran Deret Waktu

Suatu deret waktu haruslah memiliki sifat stasioner agar dapat dimodelkan. Deret waktu $\{Y_t, t \in T\}$ dikatakan stasioner jika untuk setiap t:

$$E(Y_t) = \mu \text{ (konstan)} \qquad \text{dan} \qquad Cov(Y_t, Y_{t-k}) = \gamma_k \text{ (tidak bergantung } t)$$

Jika data deret waktu diplot, secara visual akan terlihat bahwa deret waktu yang stasioner akan berfluktuasi di sekitar rataannya dengan variansi konstan.

Satu kasus lain dari sifat stasioner adalah stasioner lemah. Suatu deret waktu dikatakan stasioner lemah apabila:

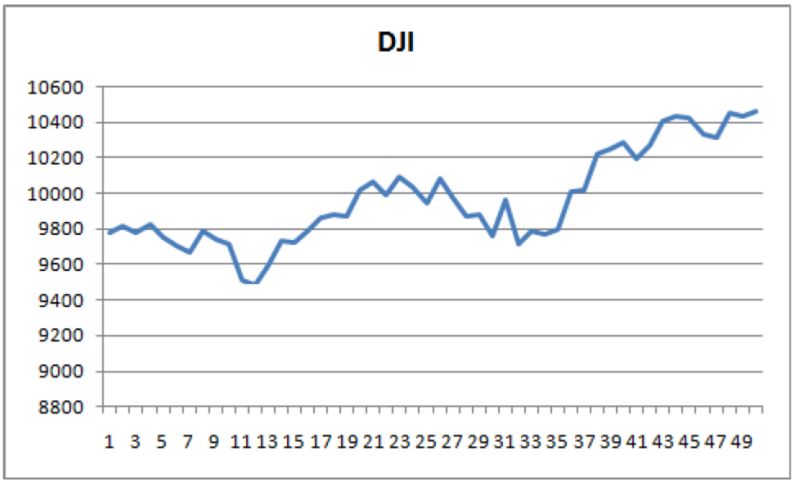
$$E(Y_t) = \mu \text{ (konstan)} \qquad \text{dan} \qquad Cov(Y_t, Y_{t-k}) = \gamma_{0,k}$$

11.2 Fungsi Autokorelasi

Langkah kedua untuk memodelkan deret waktu setelah mengecek kestasionerannya adalah mencari fungsi autokorelasi dan membuat plot tiap jeda waktu (lag) k. Kita akan mencoba menentukan Fungsi Autokorelasi dan Fungsi Autokorelasi Parsial menggunakan suatu ilustrasi data deret waktu. Data deret waktu berikut adalah nilai tutup saham harian (*daily closing stock*) DJI dari tanggal 17 September 2009 sampai 25 November 2009 dalam USD.

Tanggal	DJI	Tanggal	DJI	Tanggal	DJI	Tanggal	DJI	Tanggal	DJI
17/9/09	9783.92	1/10/09	9509.28	15/10/09	10062.94	29/10/09	9962.58	12/11/09	10197.47
18/9/09	9820.20	2/10/09	9487.67	16/10/09	9995.91	30/10/09	9712.73	13/11/09	10270.47
21/9/09	9778.86	5/10/09	9599.75	19/10/09	10092.19	2/11/09	9789.44	16/11/09	10406.96
22/9/09	9829.87	6/10/09	9731.25	20/10/09	10041.18	3/11/09	9771.91	17/11/09	10437.42
23/9/09	9748.55	7/10/09	9725.58	21/10/09	9949.36	4/11/09	9802.14	18/11/09	10426.31
24/9/09	9707.44	8/10/09	9786.87	22/10/09	10081.31	5/11/09	10005.96	19/11/09	10322.44
25/9/09	9665.19	9/10/09	9864.94	23/10/09	9972.18	6/11/09	10023.42	20/11/09	10318.16
28/9/09	9789.36	12/10/09	9885.80	26/10/09	9867.96	9/11/09	10226.94	23/11/09	10450.15
29/9/09	9742.20	13/10/09	9871.06	27/10/09	9882.17	10/11/09	10246.97	24/11/09	10433.71
30/9/09	9712.28	14/10/09	10015.86	28/10/09	9762.69	11/11/09	10291.26	25/11/09	10464.40

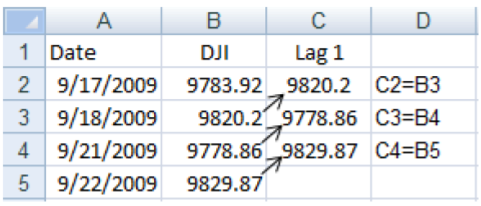
Sebelumnya, semua perhitungan dalam ACF dan PACF menggunakan program Microsoft Excel. Kita plot data tersebut menjadi diagram garis sebagai berikut:



Gambar 11.1: Diagram Garis Data Nilai Saham DJI

11.2.1 Auto Correlation Function (ACF)

Untuk menghitung ACF, data DJI dalam Excel kita "geser" satu demi satu. "Pergeseran" yang dilakukan (sebanyak k kali) menandakan *lag* atau jeda waktu sebanyak k satuan waktu. Kita lakukan hal yang terlihat pada gambar 11.2.



	A	B	C	D
1	Date	DJI	Lag 1	
2	9/17/2009	9783.92	9820.2	C2=B3
3	9/18/2009	9820.2	9778.86	C3=B4
4	9/21/2009	9778.86	9829.87	C4=B5
5	9/22/2009	9829.87		

Gambar 11.2: Pergeseran Data Menjadi Lag 1

Perlu diperhatikan bahwa jumlah data yang dimiliki saat ini adalah $n = 50$. Standarnya, kita lakukan prosedur ini hingga lag ke-k dengan $k \approx 40\% \cdot n$. Pada data ini, maka kita lakukan hingga lag ke-20.

Perhitungan ACF pada dasarnya adalah perhitungan korelasi data deret waktu dengan data lag ke-k. ACF untuk lag ke-k ditentukan melalui persamaan berikut:

$$r_k = \frac{\displaystyle\sum_{t=1}^{n-k}(y_t - \bar{y})(y_{t+k} - \bar{y})}{\displaystyle\sum_{t=1}^{n}(y_t - \bar{y})^2}$$

dengan $k = 1, 2, \cdots, \frac{2n}{5}$ dan $\bar{y}$ rataan dari y_t.

Dalam Excel, siapkan terlebih dahulu kolom yang berisi y_t dan tanggalnya. Kita bentuk tabel seperti gambar 11.3.

	A	B	C	D	E
1					M1
2	PERIOD	DATE	DJI	Deviation	ACF
3	1	17/09/2009	9783.92	-182.80	0.8906
4	2	18/09/2009	9820.2	-146.52	0.8119
5	3	21/09/2009	9778.86	-187.86	0.7098
6	4	22/09/2009	9829.87	-136.85	0.6358
7	5	23/09/2009	9748.55	-218.17	0.5649
8	6	24/09/2009	9707.44	-259.28	0.4417
9	7	25/09/2009	9665.19	-301.53	0.3254
10	8	28/09/2009	9789.36	-177.36	0.2090
11	9	29/09/2009	9742.2	-224.52	0.1278
12	10	30/09/2009	9712.28	-254.44	0.0729
13	11	01/10/2009	9509.28	-457.44	-0.0013
14	12	02/10/2009	9487.67	-479.05	-0.0460
15	13	05/10/2009	9599.75	-366.97	-0.0878
16	14	06/10/2009	9731.25	-235.47	-0.0981
17	15	07/10/2009	9725.58	-241.14	-0.1001
18	16	08/10/2009	9786.87	-179.85	-0.0577
19	17	09/10/2009	9864.94	-101.78	-0.0029
20	18	12/10/2009	9885.8	-80.92	0.0262
21	19	13/10/2009	9871.06	-95.66	0.0637
22	20	14/10/2009	10015.86	49.14	0.0587
23	21	15/10/2009	10062.94	96.22	
24	22	16/10/2009	9995.91	29.19	
25	23	19/10/2009	10092.19	125.47	

Gambar 11.3: Ilustrasi Perhitungan ACF

Keterangan: kolom A menyatakan t, kolom B menyatakan tanggal, kolom C menyatakan y_t, dan kolom D dan E tersaji dalam gambar 11.4.

	A	B	C	D	E
1					M1
2	PERIOD	DATE	DJI	Deviation	ACF
3	1	17/09/2009	9783.92	=C3-AVERAGE($C$3:$C$52)	=(SUMPRODUCT($D$3:D51,$D4:D$52)/DEVSQ($D$3:$D$52))
4	2	18/09/2009	9820.2	=C4-AVERAGE($C$3:$C$52)	=(SUMPRODUCT($D$3:D50,$D5:D$52)/DEVSQ($D$3:$D$52))
5	3	21/09/2009	9778.86	=C5-AVERAGE($C$3:$C$52)	=(SUMPRODUCT($D$3:D49,$D6:D$52)/DEVSQ($D$3:$D$52))
6	4	22/09/2009	9829.87	=C6-AVERAGE($C$3:$C$52)	=(SUMPRODUCT($D$3:D48,$D7:D$52)/DEVSQ($D$3:$D$52))
7	5	23/09/2009	9748.55	=C7-AVERAGE($C$3:$C$52)	=(SUMPRODUCT($D$3:D47,$D8:D$52)/DEVSQ($D$3:$D$52))

Gambar 11.4: *Syntax* Kolom D dan E

Sekarang, mari kita lihat *syntax* yang digunakan pada kolom D terlebih dahulu. Nilai pada kolom D disebut deviasi yang didefinisikan sebagai nilai dari $y_t - \bar{y}$.

Selanjutnya pada kolom E, rumus pada penyebut ACF dibentuk dari *syntax* DEVSQ($D$3:$D$52). Kolom D sebelumnya mengandung nilai deviasi dari rataan, namun punyebut pada persamaan ACF haruslah berbentuk kuadrat. *Syntax* DEVSQ membantu kita untuk menguadratkan semua *cell* pada kolom D lalu menjumlahkannya.

Untuk pembilang pada kolom E, kita gunakan *syntax* SUMPRODUCT. Kita ambil contoh ilustrasi pada ACF lag-1 dan lag-2. Perhatikan bahwa pembilang pada persamaan ACF lag-1 dan lag-2 berturut-turut adalah:

$$r_1 : (y_1 - \bar{y})(y_2 - \bar{y}) + (y_2 - \bar{y})(y_3 - \bar{y}) + \cdots + (y_{49} - \bar{y})(y_{50} - \bar{y})$$

$$r_2 : (y_1 - \bar{y})(y_3 - \bar{y}) + (y_2 - \bar{y})(y_4 - \bar{y}) + \cdots + (y_{48} - \bar{y})(y_{50} - \bar{y})$$

Bentuk umumnya adalah semua penjumlahan dari suku $(y_t - \bar{y})(y_{t+k} - \bar{y})$. Perhatikan kembali gambar 11.3. Perhitungan pembilang ACF lag-1 didapat dengan melakukan operasi berikut:

$$= D3 * D4 + D4 * D5 + \cdots + D49 * D50$$

atau dapat disederhanakan menjadi:

$$= \text{SUMPRODUCT}(\$D\$3 : D51, D4 : \$D\$52)$$

untuk lag ke-2, 3, hingga 20 rumusnya menjadi:

$$r_2 : \text{SUMPRODUCT}(\$D\$3 : D50, D5 : \$D\$52)$$

$$r_3 : \text{SUMPRODUCT}(\$D\$3 : D49, D6 : \$D\$52)$$

$$\vdots$$

$$r_{20} : \text{SUMPRODUCT}(\$D\$3 : D34, D21 : \$D\$52)$$

Cara manual Akan lebih mudah bila kita bentuk tabel terlebih dahulu. Bentuk tabel dengan kolom-kolom sebagai berikut: t, y_t, y_{t+1}, y_{t+2}, $\cdots$, $y_{t+\frac{2n}{5}}$. Lalu hitung $\bar{y}$ terlebih dahulu. Kurangi semua *cell* dari kolom y_t hingga $y_{t+\frac{2n}{5}}$ dengan $\bar{y}$, dan catat semuanya di suatu tabel yang baru. Jangan lupa untuk menghitung $(y_t - \bar{y})^2$ dan jumlahkan semuanya. Selanjutnya, untuk menghitung r_k kalikan y_t dengan y_{t+k} (hanya yang ada pasangannya saja), lalu jumlahkan. Bagi dengan $(y_t - \bar{y})^2$ dan didapat r_k.

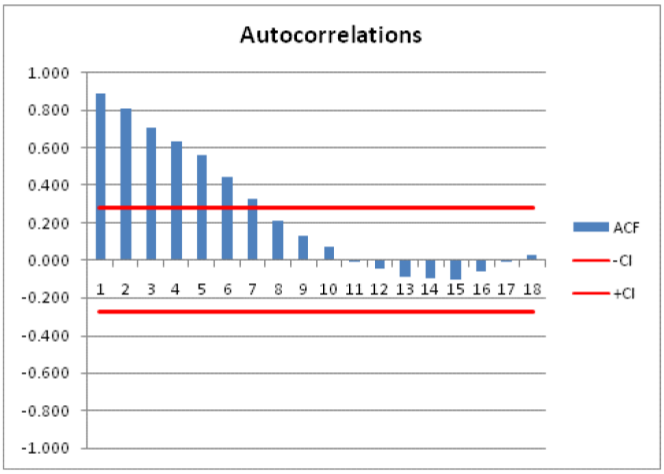
Catatan: $-1 \leq r_k \leq 1$.

Selanjutnya, kita akan melihat kenormalan ACF. ACF lag-k dapat diabaikan ($r_k = 0$) apabila:

$$|r_k| < \frac{z_{\alpha/2}}{\sqrt{n-k}}$$

dengan α standar 5% ($z_{0.025} = 1.96$). Jika r_k masuk ke dalam interval tersebut, maka r_k akan lanjut ke dalam uji Ljung-Box (dijelaskan di subbab selanjutnya).

Secara visual, pengujian ACF untuk data saham DJI dapat dilihat pada gambar 11.5.



Gambar 11.5: Ilustrasi Pengujian ACF dengan Diagram Batang

11.2.2 *Partial Auto Correlation Function (PACF)*

Berbeda dengan ACF, nilai PACF didapat dengan melibatkan nilai-nilai ACF yang telah didapat. Kita definisikan PACF untuk lag $k = 1$ adalah:

$$p_{1.1} = r_1$$

dan untuk lag $k > 1$:

$$p_{k.k} = \frac{r_k \sum_{j=1}^{k-1} p_{k-1.j} r_{k-j}}{1 - \sum_{j=1}^{k-1} p_{k-1.j} r_j}$$

dengan

$$p_{k.j} = p_{k-1.j} - p_{k.k} p_{k-1.k-j}$$

untuk $j = 1, 2, \cdots, k-1$.

Memang bukan perhitungan yang mudah bila dilakukan secara manual, namun kita coba lihat polanya.

$$p_{2.2} = \frac{r_2 - p_{1.1}r_1}{1 - p_{1.1}r_1}$$

$$p_{3.3} = \frac{r_3 - (p_{2.1}r_2 - p_{2.2}r_1)}{1 - (p_{2.1}r_1 - p_{2.2}r_2)}$$

$$p_{4.4} = \frac{r_4 - (p_{3.1}r_3 + p_{3.2}r_2 + p_{3.3}r_1)}{1 - (p_{3.1}r_1 + p_{3.2}r_2 + p_{3.3}r_3)}$$

dengan:

$$p_{2.1} = p_{1.1} - p_{2.2}p_{1.1}$$

$$p_{3.1} = p_{2.1} - p_{3.3}p_{2.2}$$

$$p_{3.2} = p_{2.2} - p_{3.3}p_{2.1}$$

dan seterusnya.

Kembali pada ilustrasi deret waktu saham DJI. Perhitungan PACF dengan Excel dapat dilihat pada gambar 11.6.

	A	B	C	D	E	F	G	H	I	J	K	L
1	**Lag**	**AC**	**PAC**									
2	k	r_k	$r_{k,j}$	k,1	k,2	k,3	k,4	k,5	k,6	k,7	k,8	k,9
3	1	0.891	0.891	**0.891**								
4	2	0.812	0.091	0.810	**0.091**							
5	3	0.710	-0.138	0.823	0.203	**-0.138**						
6	4	0.636	0.053	0.830	0.192	-0.182	**0.053**					
7	5	0.565	0.008	0.830	0.193	-0.184	0.047	**0.008**				
8	6	0.442	-0.321	0.832	0.208	-0.243	0.109	0.273	**-0.321**			
9	7	0.325	-0.107	0.797	0.238	-0.231	0.083	0.296	-0.231	**-0.107**		
10	8	0.209	-0.025	0.795	0.232	-0.224	0.085	0.290	-0.225	-0.088	**-0.025**	
11	9	0.128	0.032	0.796	0.235	-0.217	0.076	0.288	-0.218	-0.095	-0.050	**0.032**

Gambar 11.6: Perhitungan PACF

Beberapa *syntax* yang digunakan tersaji dalam tabel berikut:

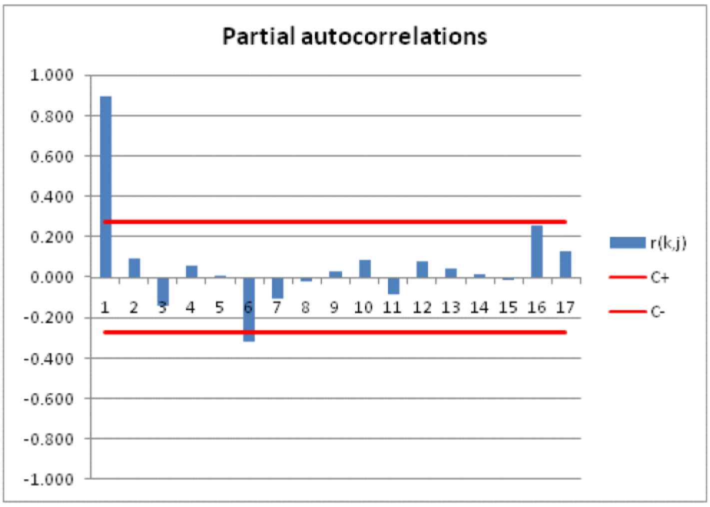
PACF untuk lag k,1	*Cell* D3	D3=B3
	Cell D4:D19	D4=D3-(D3*E4) D5=D4-(E3*F5) dst.
PACF untuk lag k,2	*Cell* E4	E4=(B4-(D3*B3))/(1-(D3*C3))
	Cell E5:E19	E5=E4-(D4*F5) E6=E5-(E5*G6) dst.
PACF untuk lag k,3	*Cell* F5	F5=(B5-(D4*B4+E4*B3))/(1-(D4*B3+E4*B4))
	Cell F6:F19	F6=F5-(D5*G6) F7=E5-(E6*H7) dst.

Pengujian PACF secara signifikan pun serupa dengan ACF. PACF lag-k dapat diabaikan ($p_{k.k} = 0$) apabila:

$$|p_{k.k}| < \frac{z_{\alpha/2}}{\sqrt{n-k}}$$

dengan α standar 5% ($z_{0.025} = 1.96$).

Dengan menggunakan rumus yang diberikan, maka didapat diagram batang sebagai berikut:



Gambar 11.7: Ilustrasi Pengujian PACF dengan Diagram Batang

11.3 *Uji Diagnosis Ljung-Box*

Selanjutnya, kita akan melihat pengujian hipotesis mengenai signifikansi ACF atau PACF. Uji ini dinamakan Ljung-Box.

$$H_0 : \text{korelasi antar sisaan} = 0 \qquad H_1 : \text{korelasi antar sisaan} \neq 0$$

Gunakan statistik uji sebagai berikut:

$$Q^* = n(n+2)\sum_{j=1}^{k} \frac{r_j^2}{n-j}$$

dengan r_j sebagai ACF untuk lag-j.

H_0 ditolak jika $Q^* > \chi^2_{\alpha,k}$.

11.4 *Model Deret Waktu*

11.4.1 *Model Auto Regressive (AR)*

Prinsip dasar dari model AR adalah regresi terhadap data deret waktu yang lalu dan galat sekarang. Bentuk umumnya:

$$AR(1) : y_t = \mu + \phi_1 y_{t-1} + e_t$$

dengan $|\phi_1| < 1$

$$AR(2) : y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + e_t$$

dengan $|\phi_2| < 1$, $\phi_2 + \phi_1 < 1$, $\phi_2 - \phi_1 < 1$

$$\vdots$$

$$AR(p) : y_t = \mu + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + e_t$$

Sejauh ini kita baru bisa memodelkan AR(1) karena kita baru mempelajari regresi linier sederhana saja. Untuk orde selanjutnya, kita bisa menggunakan program statistika.

11.4.2 *Model Moving Average (MA)*

Prinsip dasar dari model MA adalah regresi terhadap galat yang lalu dan galat sekarang. Bentuk umumnya:

$$MA(1) : y_t = \mu + e_t - \theta_1 e_{t-1}$$

dengan $|\theta_1| < 1$

$$MA(2) : y_t = \mu + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}$$

dengan $|\theta_2| < 1, \theta_2 + \theta_1 < 1, \theta_2 - \theta_1 < 1$

$$\vdots$$

$$MA(q) : y_t = \mu + e_t - \theta_1 et - 1 - \cdots - \theta_q e_{t-q}$$

11.4.3 Model ARMA dan ARIMA

Model ARMA adalah gabungan dari model AR dan MA. Orde dari model ARMA adalah (p,q) dengan p melambangkan orde AR dan q melambangkan orde MA. Perhatikan bahwa ARMA(0,q) = MA(q) dan ARMA(p,0) = AR(p).

Prinsip dasar dari model ARMA adalah regresi terhadap data deret waktu yang lalu dan semua galat. Bentuk umumnya adalah sebagai berikut:

$$ARMA(1, 1) : y_t = \mu + \phi_1 y_{t-1} - \theta_1 e_{t-1}$$

$$\vdots$$

$$ARMA(p, q) : y_t = \mu + (\phi_1 y_{t-1} + \cdots + \phi_p y_{t-p}) + (e_t - \theta_1 et - 1 - \cdots - \theta_q e_{t-q})$$

Bila deret waktu tidak stasioner, lakukan diferensi terlebih dahulu, yaitu dengan membentuk deret waktu baru, $Z_t = Y_t - Y_{t-1}$ untuk setiap t. ARMA(p,q) untuk Z_t disebut ARIMA(p,1,q). Jika diferensi dilakukan d kali, maka ditulis ARIMA(p,d,q).

11.4.4 Identifikasi Model Deret Waktu

Parameter ACF dan PACF menjadi penentu model deret waktu yang akan kita buat selanjutnya. Plot ACF dan PACF akan membentuk suatu pola tertentu. Rangkumannya tersaji dalam tabel berikut:

Model	**ACF**	**PACF**
AR(p)	Eksponensial turun atau sinus teredam	Terpotong setelah lag-p.
MA(q)	Terpotong setelah lag-q	Eksponensial turun atau sinus teredam
ARMA(p,q)	Eksponensial turun	Eksponensial turun

Dalam program statistika, kita dapat mengidentifikasi orde bagi model ARMA melalui kriteria Akaike (AIC):

$$AIC = 2k - 2 \ln L$$

dengan k jumlah parameter dan L nilai maksimum dari fungsi likelihood dari estimasi model.

Untuk setiap orde (p, q), hitung nilai AIC. Orde yang dipilih adalah (p, q) dengan nilai AIC yang paling kecil.

Sebagai contoh dalam ilustrasi saham DJI, kita dapat menentukan bahwa terjadi *cut-off* di lag-1 pada PACF dan plot ACF membentuk sinus teredam. Maka kita dapat simpulkan bahwa model yang digunakan adalah model AR(1). Untuk menentukan model deret waktunya, lakukan regresi linier sederhana dengan y_t terhadap y_{t-1}.

11.5 Soal Latihan

1. Diberikan data deret waktu berikut:

t	0	1	2	3	4	5	6	7	8	9
Y_t	1.0000	1.0231	1.0990	1.9896	2.3005	0.7800	-0.3928	0.5442	-0.9379	0.3406

   a) Tentukan ACF dari sampel tersebut.

   b) Plot ACF terhadap lag-k.

   c) Tentukan model deret waktu yang sesuai berdasarkan plot ACF terhadap lag-k yang sudah dibuat.

Bab 12

Pengantar Analisis Spasial

Setelah kita mempelajari sedikit mengenai analisis deret waktu, kita akan melihat peramalan data jangka pendek dari segi lokasi atau daerah. Dalam deret waktu kita dapat menganalisis satu kasus stokastik dari berbagai segi waktu. Dalam bab ini, kita akan menganalisis kasus stokastik dari berbagai lokasi dalam satu waktu.

12.1 *Terminologi Analisis Spasial*

Dalam analisis spasial (atau lebih sering siebut analisis variaogram), kita ingin mengukur hubungan dari suatu parameter di tiap-tiap lokasi, misalnya: kandungan mineral, cuaca, kelembaban, dll. Suatu hubungan tersebut kita bentuk dalam suatu fungsi $Z(s)$. Cara mendapatkan fungsi tersebut bisa dilakukan dalam berbagai jenis. Dalam bab ini, kita akan mempelajari mengenai semivariogram. Semivariogram dibagi menjadi dua, yaitu isotropik (tidak memperhitungkan sudut) dan anisotropik (memperhitungkan sudut). Model semivariogram yang didapat nantinya dapat kita gunakan untuk menaksir parameter di dalam domain model semivariogram tersebut. Teknik tersebut diberi nama *kriging*. Bab ini lebih difokuskan untuk mencari model semivariogram yang sesuai.

12.2 *Semivariogram dan Modelnya*

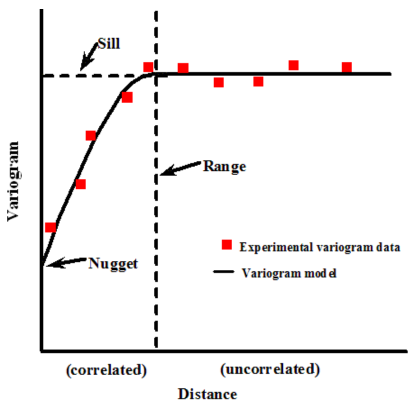
Semivariogram adalah alat ukur kebergantungan antara observasi yang didasarkan pada jarah (h) antar pasangan lokasi. Semivariogram disajikan oleh persamaan berikut:

$$2\gamma(h) = \text{Var}(Z(s) - E(Z(s+h)))$$

Dalam pengolahan observasi, kita gunakan semivariogram eksperimental:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(s_i) - Z(s_i + h)]^2$$

Dari hasil semivariogram eksperimental yang didapat, kita cocokan dengan model semivariogram standar yang diberikan. Sebelumnya, kita lihat dulu beberapa parameter dari model semivariogram:



Gambar 12.1: Ilustrasi Model Semivariogram

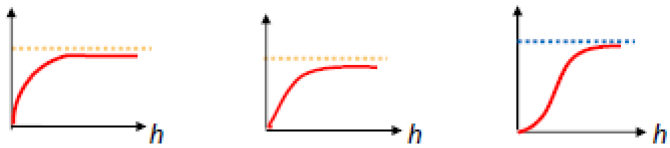
- Range (a)
  Jarak lag hingga nilai semivariogram konstan (jarak maksimum antara lokasi yang masih memiliki korelasi spasial).

- Sill $(C_0 + C)$
  Nilai semivariogram yang konstan untuk h yang tidak terbatas. Umumnya, nilai sill mendekati variansi data.

- Nugget (C_0)
  Kesalahan pengukuran di mana nilai semivariogram pada lag jarak nol nilainya tak nol. Efek nugget akan hilang dengan memperkecil jarak antara titik-titik sampel.

Berikut akan diberikan beberapa model semivariogram standar:

Model Sperikal (untuk pertambangan)	$\gamma(h) = \begin{cases} C\left[\frac{3}{2}\left(\frac{h}{a}\right) - \frac{1}{2}\left(\frac{h}{a}\right)^3\right] & h < a \\ C & h \geq a \end{cases}$
Model Eksponensial (untuk hidrologi)	$\gamma(h) = C\left(1 - \exp\left(-\frac{h}{a}\right)\right)$
Model Gauss (untuk perminyakan)	$\gamma(h) = C\left(1 - \exp\left(-\left(\frac{h}{a}\right)^2\right)\right)$



Gambar 12.2: Grafik Model Semivariogram Sperikal (Kiri), Eksponensial (Tengah), dan Gauss (Kanan)

Untuk mencocokan model semivariogram standar dengan semivariogram eksperimental yang diberikan, kita harus menggunakan program khusus analisis spasial, misalnya: *Surfer*.

Contoh Soal 1. Cari model variogram yang sesuai (secara intuitif) dari grid yang diberikan sebagai berikut:

10	30	10	55
5	20	50	35
40	15	40	30
5	35	10	10

Catatan: tiap grid memiliki jarak 1 satuan.
Solusi. Banyak pasangan data yang terbentuk adalah $\binom{16}{2} = 120$. Kita cari semua nilai variogram yang mungkin. Sebagai contoh untuk koordinat (1,1):

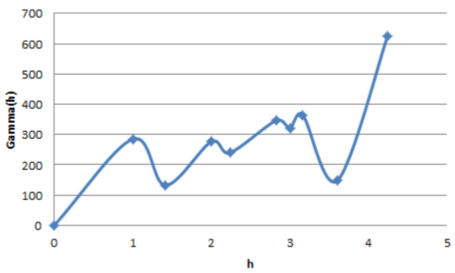
Koordinat Head	Koordinat Tail	Head	Tail	Jarak (h)	$(Z(\text{head}) - Z(\text{tail})^2$
(1,1)	(1,2)	10	30	1	400
(1,1)	(1,3)	10	10	2	0
(1,1)	(1,4)	10	55	3	2025
(1,1)	(2,1)	10	5	1	25
(1,1)	(2,2)	10	20	1.414	100
(1,1)	(2,3)	10	50	2.236	1600
(1,1)	(2,4)	10	35	3.162	625
(1,1)	(3,1)	10	40	2	900
(1,1)	(3,2)	10	15	2.236	25
(1,1)	(3,3)	10	40	2.828	900
(1,1)	(3,4)	10	30	3.6055	400
(1,1)	(4,1)	10	5	3	25
(1,1)	(4,2)	10	35	3.162	625
(1,1)	(4,3)	10	10	3.605	0
(1,1)	(4,4)	10	10	4.242	0

dan seterusnya hingga pasangan terakhir. Lalu kita kelompokkan semua nilai yang jaraknya sama dan kita hitung $\hat{\gamma}(h)$. Hasil perhitungan tersaji pada tabel berikut:

h	$N(h)$	$\hat{\gamma}(h)$
0	0	0
1	27	285.1852
1.414	15	132.5
2	16	276.5625
2.236	21	239.881

h	$N(h)$	$\hat{\gamma}(h)$
2.828	8	346.875
3	8	320.3125
3.1622	15	365
3.6055	8	148.4375
4.242	2	625

Lalu plot $\hat{\gamma}(h)$ terhadap h sehingga didapat seperti pada gamabar 12.3.



Gambar 12.3: Ilustrasi Semivariogram Eksperimental Contoh Soal 1

Secara intuitif, model variogram yang sesuai adalah model Gauss. ■

12.3 *Soal Latihan*

1. Cari model variogram yang sesuai (secara intuitif) dari grid yang diberikan sebagai berikut:

15	35	20	35
15	20	45	25
40	10	30	30
15	35	20	10

Catatan: tiap grid memiliki jarak 1 satuan.

Bab 13

Pengantar Statistika Pengendalian Kualitas

Data-data dalam suatu pabrik produksi tentu tidak terlepas dari statistika. Data-data tersebut dapat menjelaskan apakah hasil dari pabrik tersebut masih berada dalam kontrol atau tidak. Tentu saja, barang yang diinginkan oleh pemilik pabrik adalah barang yang berada dalam kontrol. Hal tersebut dapat kita tentukan melalui statistika pengendalian kualitas.

13.1 Membangun Diagram Kendali

Dalam bab ini kita akan mempelajari 3 (tiga) bagan pengendalian mutu, yaitu *X bar-chart* (pengontrolan variabel) dan *R-chart* serta *S-chart* (pengontrolan variabilitas).

Untuk membangun diagram kendali dari suatu parameter, kita harus menentukan rataan dari parameter tersebut (yang dijadikan sebuah standar), serta selang toleransinya. Perhatikan kembali pada distribusi normal, Luas di bawah kurva distribusi tersebut akan mendekati 1 untuk interval $[x - 3\sigma, x + 3\sigma]$. Hal tersebut yang dijadikan sebuah standar bagi diagram kendali.

Secara umum, langkah-langkah membuat diagram kendali adalah sebagai berikut:

- Tentukan nilai dari suatu parameter tiap observasi masing-masing sampel.
- Tentukan rataan dari parameter yang dimaksud (CL = *center line*).
- Tentukan UCL (*upper control limit*) dan LCL (*lower control limit*). Misal parameter yang digunakan θ maka:

  $$UCL = \bar{\theta} + k\sigma_\theta \qquad\qquad LCL = \bar{\theta} - k\sigma_\theta$$

  dengan $k_{\text{standar}} = 3$ dan σ_θ merupakan standar deviasi dari parameter theta yang dimaksud.
- Plot nilai parameter tiap sampel beserta CL, UCL, dan LCL.

<u>Contoh Soal 1.</u> Diberikan data mengenai ketahanan suatu komponen senjata angin (dalam pon per inci kuadrat). Dua puluh lima sampel dikenakan pengujian sebanyak 5 kali tiap jamnya. Hasilnya tersaji dalam tabel berikut:

Sampel	Observasi					$\bar{x}_i$	R_i
1	1515	1518	1512	1498	1511	1510.8	20
2	1504	1511	1507	1499	1502	1504.6	12
3	1517	1513	1504	1521	1520	1515.0	17
4	1497	1503	1510	1508	1502	1504.0	13
5	1507	1502	1497	1509	1512	1505.4	15
6	1519	1522	1523	1517	1511	1518.4	12
7	1498	1497	1507	1511	1508	1504.2	14
8	1511	1518	1507	1503	1509	1509.6	15
9	1506	1503	1498	1508	1506	1504.2	10
10	1503	1506	1511	1501	1500	1504.2	11
11	1499	1503	1507	1503	1501	1502.6	8
12	1507	1503	1502	1500	1501	1502.6	7
13	1500	1506	1501	1498	1507	1502.4	9
14	1501	1509	1503	1508	1503	1510.8	8
15	1507	1508	1502	1509	1501	1510.4	8
16	1511	1509	1503	1510	1507	1508.0	8
17	1508	1511	1513	1509	1506	1509.4	7
18	1508	1509	1512	1515	1519	1512.6	11
19	1520	1517	1519	1522	1516	1518.8	6
20	1506	1511	1517	1516	1508	1511.6	11
21	1500	1498	1503	1504	1508	1502.6	10
22	1511	1514	1509	1508	1506	1509.6	8
23	1505	1508	1500	1509	1503	1505.0	9
24	1501	1498	1505	1502	1505	1502.2	7
25	1509	1511	1507	1500	1499	1505.2	12

Bentuk diagram kendali *X bar-chart* dan *R-chart* dari data yang diberikan.
Solusi. Tentukan terlebih dahulu rataan dari $\bar{x}$ dan R masing-masing:

$$\bar{\bar{x}} = \frac{1}{25}\sum_{i=1}^{25}\bar{x}_i = 1507.328 \qquad \bar{R} = \frac{1}{25}\sum_{i=1}^{25}R_i = 10.72$$

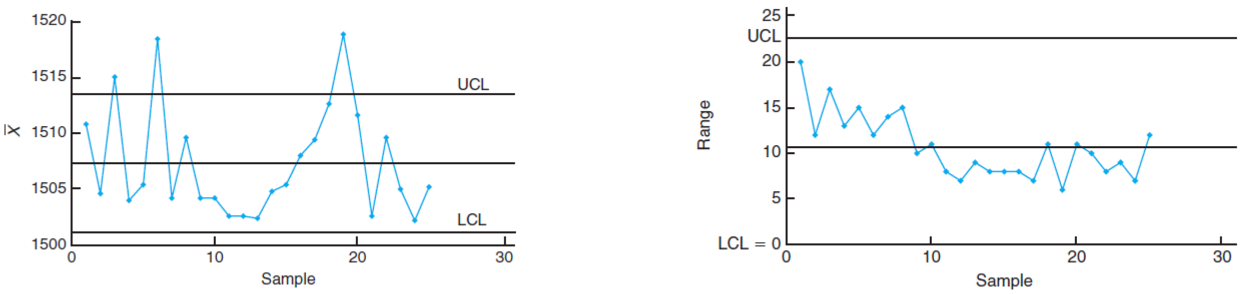
UCL dan LCL bagi *X bar-chart* adalah sebagai berikut:

$$UCL = \bar{\bar{x}} + 3\sigma_{\bar{x}} = 1513.5134 \qquad LCL = \bar{\bar{x}} - 3\sigma_{\bar{x}} = 1501.1426$$

Lalu UCL dan LCL bagi *R-chart* adalah sebagai berikut:

$$UCL = \bar{R} + 3\sigma_R = 22.6621 \qquad LCL = \bar{R} - 3\sigma_R = 0$$

Plot semua $\bar{x}$ dan R sehingga menghasilkan diagram kendali pada gambar 12.1 berikut.



Gambar 12.1: Diagram Kendali *X bar-chart* (kiri) dan *R-chart* (kanan)

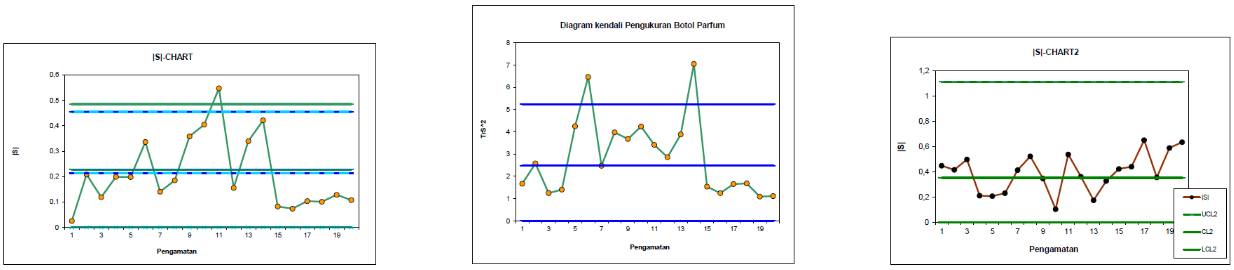
■

13.2 *Deteksi Out of Control (OOC)*

Klasifikasi suatu proses dikatakan *out of control* (OOC) apabila:

- Ada titik di luar batas kedua garis *control limit*.
- **Delapan** titik berturut-turut jatuh di antara CL-UCL atau CL-LCL.
- Terdapat tren naik atau tren turun.

Contoh Soal 2. Diberikan diagram kendali berikut. Tentukan apakah diagram kendali tersebut termasuk ke dalam OOC atau tidak. Berikan penjelasan Anda.



Gambar 12.2: Diagram Kendali Contoh Soal 2.

Solusi. Keterangan: dari kiri ke kanan kita namakan diagram kendali nomor 1, 2, dan 3.

Pada diagram kendali 1 dan 2 terlihat bahwa ada data observasi yang di luar batas garis *control limit* (di atas UCL), maka diagram kendali 1 dan 2 termasuk OOC. Sedangkan pada diagram kendali 3, tidak ada syarat OOC yang memenuhi. Jadi diagram kendali 3 masih dalam kontrol. ■

13.3 *Soal Latihan*

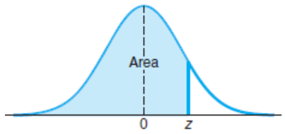
1. Berikut adalah data hasil pengukuran ketebalan dari produksi suatu kepingan logam (dalam mm). Pengamatan dilakukan selama 10 kali dan masing-masing pengamatan mengambil 4 sampel. Bangun bagan kendali yang sesuai.

m	n			
1	459	449	435	450
2	443	440	442	442
3	457	444	449	444
4	469	463	453	438
5	443	457	445	454
6	444	456	456	457
7	445	449	450	445
8	446	455	449	452
9	459	452	457	440
10	441	457	463	443

Apakah data pengamatan termasuk ke dalam klasifikasi *Out Of Control*? Jelaskan jawaban Anda.

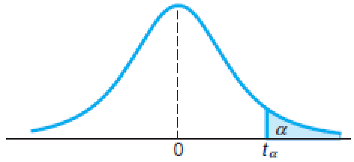
Lampiran

Lampiran 1: Tabel Luas di Bawah Kurva Normal



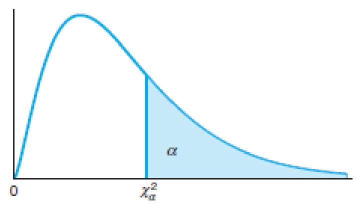
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9985	0.9986	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9989	0.9990	0.9990
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9992	0.9993	0.9993
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998



Lampiran 2: Tabel Nilai Kritis Distribusi *t-Student*

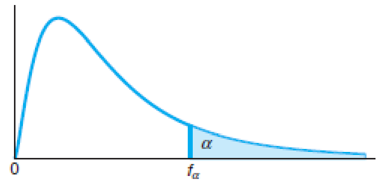
ν	α										
	0.40	**0.30**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.02**	**0.015**	**0.01**	**0.005**
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706	15.894	21.205	31.821	63.656
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303	4.849	5.643	6.965	9.925
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182	3.482	3.896	4.541	5.841
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776	2.999	3.298	3.747	4.604
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571	2.757	3.003	3.365	4.032
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447	2.612	2.829	3.143	3.707
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365	2.517	2.715	2.998	3.499
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306	2.449	2.634	2.896	3.355
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262	2.398	2.574	2.821	3.250
10	0.260	0.542	0.879	1.093	1.372	1.812	2.228	2.359	2.527	2.764	3.169
11	0.260	0.540	0.876	1.088	1.363	1.796	2.201	2.328	2.491	2.718	3.106
12	0.259	0.539	0.873	1.083	1.356	1.782	2.179	2.303	2.461	2.681	3.055
13	0.259	0.538	0.870	1.079	1.350	1.771	2.160	2.282	2.436	2.650	3.012
14	0.258	0.537	0.868	1.076	1.345	1.761	2.145	2.264	2.415	2.624	2.977
15	0.258	0.536	0.866	1.074	1.341	1.753	2.131	2.249	2.397	2.602	2.947
16	0.258	0.535	0.865	1.071	1.337	1.746	2.120	2.235	2.382	2.583	2.921
17	0.257	0.534	0.863	1.069	1.333	1.740	2.110	2.224	2.368	2.567	2.898
18	0.257	0.534	0.862	1.067	1.330	1.734	2.101	2.214	2.356	2.552	2.878
19	0.257	0.533	0.861	1.066	1.328	1.729	2.093	2.205	2.346	2.539	2.861
20	0.257	0.533	0.860	1.064	1.325	1.725	2.086	2.197	2.336	2.528	2.845
21	0.257	0.532	0.859	1.063	1.323	1.721	2.080	2.189	2.328	2.518	2.831
22	0.256	0.532	0.858	1.061	1.321	1.717	2.074	2.183	2.320	2.508	2.819
23	0.256	0.532	0.858	1.060	1.319	1.714	2.069	2.177	2.313	2.500	2.807
24	0.256	0.531	0.857	1.059	1.318	1.711	2.064	2.172	2.307	2.492	2.797
25	0.256	0.531	0.856	1.058	1.316	1.708	2.060	2.167	2.301	2.485	2.787
26	0.256	0.531	0.856	1.058	1.315	1.706	2.056	2.162	2.296	2.479	2.779
27	0.256	0.531	0.855	1.057	1.314	1.703	2.052	2.158	2.291	2.473	2.771
28	0.256	0.530	0.855	1.056	1.313	1.701	2.048	2.154	2.286	2.467	2.763
29	0.256	0.530	0.854	1.055	1.311	1.699	2.045	2.150	2.282	2.462	2.756
30	0.256	0.530	0.854	1.055	1.310	1.697	2.042	2.147	2.278	2.457	2.750
40	0.255	0.529	0.851	1.050	1.303	1.684	2.021	2.123	2.250	2.423	2.704
60	0.254	0.527	0.848	1.045	1.296	1.671	2.000	2.099	2.223	2.390	2.660
120	0.254	0.526	0.845	1.041	1.289	1.658	1.980	2.076	2.196	2.358	2.617
∞	0.253	0.524	0.842	1.036	1.282	1.645	1.960	2.054	2.170	2.326	2.576



Lampiran 3: Tabel Nilai Kritis Distribusi Khi Kuadrat

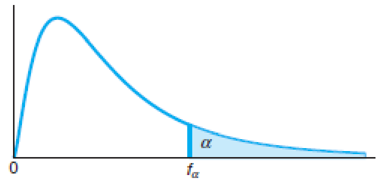
ν	α														
	0.995	**0.99**	**0.98**	**0.975**	**0.95**	**0.90**	**0.75**	**0.50**	**0.25**	**0.10**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**
1	0.0^4393	0.0^3157	0.0^3628	0.0^3982	0.00393	0.0158	0.102	0.455	1.323	2.706	3.841	5.024	5.412	6.635	7.879
2	0.0100	0.0201	0.0404	0.0506	0.103	0.211	0.575	1.386	1.323	4.605	5.991	7.378	7.824	9.210	10.597
3	0.0717	0.115	0.185	0.216	0.352	0.584	1.213	2.366	1.323	6.251	7.815	9.348	9.837	11.345	12.838
4	0.217	0.297	0.429	0.484	0.711	1.064	1.923	3.357	1.323	7.779	9.488	11.143	11.668	13.277	14.860
5	0.412	0.554	0.752	0.831	1.145	1.610	2.675	4.351	1.323	9.236	11.070	12.832	13.388	15.086	16.750
6	0.676	0.872	1.134	1.237	1.635	2.204	3.455	5.348	1.323	10.645	12.592	14.449	15.033	16.812	18.548
7	0.989	1.239	1.564	1.690	2.167	2.833	4.255	6.346	1.323	12.017	14.067	16.013	16.622	18.475	20.278
8	1.344	1.647	2.032	2.180	2.733	3.490	5.071	7.344	1.323	13.362	15.507	17.535	18.168	20.090	21.955
9	1.735	2.088	2.532	2.700	3.325	4.168	5.899	8.343	1.323	14.684	16.919	19.023	19.679	21.666	23.589
10	2.156	2.558	3.059	3.247	3.940	4.865	6.737	9.342	1.323	15.987	18.307	20.483	21.161	23.209	25.188
11	2.603	3.053	3.609	3.816	4.575	5.578	7.584	10.341	1.323	17.275	19.675	21.920	22.618	24.275	26.757
12	3.074	3.571	4.178	4.404	5.226	6.304	8.438	11.340	1.323	18.549	21.026	23.337	24.054	26.217	28.300
13	3.565	4.107	4.765	5.009	5.892	7.041	9.299	12.340	1.323	19.812	22.362	24.736	25.471	27.688	29.819
14	4.075	4.660	5.368	5.629	6.571	7.790	10.165	13.339	1.323	21.064	23.685	26.119	26.873	29.141	31.319
15	4.601	5.229	5.985	6.262	7.261	8.547	11.037	14.339	1.323	22.307	24.996	27.488	28.259	30.578	32.801
16	5.142	5.812	6.614	6.908	7.962	9.312	11.912	15.338	1.323	23.542	26.296	28.845	29.633	32.000	34.267
17	5.697	6.408	7.255	7.564	8.672	10.085	12.792	16.338	1.323	24.769	27.587	30.191	30.995	33.409	35.718
18	6.265	7.015	7.906	8.231	9.390	10.865	13.675	17.338	1.323	25.989	28.869	31.526	32.346	34.805	37.156
19	6.844	7.633	8.567	8.907	10.117	11.651	14.562	18.338	1.323	27.204	30.144	32.852	33.687	36.191	38.582
20	7.434	8.260	9.237	9.591	10.851	12.443	15.452	19.337	1.323	28.412	31.410	34.170	35.020	37.566	39.997
21	8.034	8.897	9.915	10.283	11.591	13.240	16.344	20.337	1.323	29.615	32.671	35.479	36.343	38.932	41.401
22	8.643	9.542	10.600	10.982	12.338	14.041	17.240	21.337	1.323	30.813	33.924	36.781	37.659	40.289	42.796
23	9.260	10.196	11.293	11.689	13.091	14.848	18.137	22.337	1.323	32.007	35.172	38.076	38.968	41.638	44.181
24	9.886	10.856	11.992	12.401	13.848	15.659	19.037	23.337	1.323	33.196	36.415	39.364	40.270	42.980	45.558
25	10.520	11.524	12.697	13.120	14.611	16.473	19.939	24.337	1.323	34.382	37.652	40.646	41.566	44.314	46.928

Lampiran 4: Tabel Nilai Kritis Distribusi Fisher ($\alpha = 0.05$)



ν_2	ν_1														
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**12**	**15**	**20**	**24**	**30**
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91	245.95	248.01	249.05	250.10
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62
4	7.71	6.54	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.07	2.04
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.04	2.01
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.01	1.98
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	1.98	1.96
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.96	1.94
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.94	1.92

Lampiran 5: Tabel Nilai Kritis Distribusi Fisher ($\alpha = 0.01$)



ν_2	ν_1														
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30
1	4052.18	4999.50	5403.35	5624.58	5763.65	5858.99	5928.36	5981.07	6022.47	6055.85	6106.32	6157.28	6208.73	6234.63	6260.65
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.60	26.50
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.93	13.84
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.47	9.38
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54

Daftar Pustaka

Cryer, J. D., et.al. 2008. *Time Series Analysis with Applications in R*, Springer, New York.
Walpole, Ronald E., et.al. 2007. *Statistics for Scientist and Engineering*, 8th *edition.*
http://pecar-uk.com/Autocorrelations.pdf