MODUL 1
PENDAHULUAN DAN DISTRIBUSI FREKUENSI

Pokok Bahasan:

- Arti dan Kegunaan Statistika
- Macam-macam Data: Data Kuantitatif dan Data Kualitatif
- Pengertian tentang Populasi dan Sampel
- Pengertian tentang dan Cara Menyusun Distribusi Frekuensi
- Distribusi Frekuensi Relatif dan Kumulatif
- Gambar dan Grafik Distribusi Frekuensi

1.1 Arti dan Kegunaan Statistik

Statistik memiliki arti yang sangat luas, bisa berbeda antara satu dengan yang lain tergantung dari tujuan penggunaannya. Singkatnya, statistik mengacu pada penggunaan data kuantitatif.

Secara umum statistik dibedakan menjadi dua, yaitu statistik deskriptif dan statistik inferensi. Statistik deskriptif bisa diartikan sebagai penyajian data dalam bentuk yang lebih ringkas agar lebih mudah dipahami. Beberapa topik yang dibahas dalam statistik deskriptif adalah ukuran pemusatan (rata-rata, median, modus) dan ukuran penyebaran (standar deviasi dan koefisien variasi). Sedangkan statistik inferensia merupakan pernyataan yang digeneralisasi untuk populasi berdasarkan sampel random (dari populasi tersebut). Berkaitan dengan statistik inferensi, perlu dipahami konsep sampling, pendugaan, pengujian hipotesis dan sebagainya.

1.2 Macam-macam Data: Data Kuantitatif dan Data Kualitatif

Statistik berkaitan erat dengan pemanfaatan data. Ada berbagai jenis pengelompokan data, salah satunya adalah kategori data kuantitatif dan data kualitatif. Data kuantitatif mengacu pada angka, seperti: jumlah penduduk, Produk Domestik Bruto (PDB), tingkat pertumbuhan ekonomi, dan sebagainya.

Sedangkan data kualitatif mengacu pada yang bukan angka, seperti: agama, jenis kelamin, tingkat pendidikan, tingkat pengalaman, dan sebagainya. Akan tetapi yang perlu dicatat, dalam proses pengolahan data kuantitatif, bisa dilakukan *kuantifikasi* dari data kualitatif tersebut. Salah satunya adalah dengan membuat kode untuk data kualitatif. Sebagai contoh, jenis kelamin. Untuk memudahkan proses pengolahan data, diberikan kode 1 untuk responden laki-laki dan kode 0 untuk responden perempuan. Sehingga, data jenis kelamin yang pada awalnya adalah kualitatif, setelah diberikan kode 1 dan 0, berubah menjadi data kuantitatif.

1.3 Pengertian tentang Populasi dan Sampel

Pengertian tentang populasi dan sampel dapat dipahami melalui ilustrasi berikut. Sebagai contoh, Perusahaan Konsultan ”Selalu Berkarya” ingin melihat hubungan kemampuan bahasa asing TKW Indonesia dengan tingkat pendapatan. Berkaitan dengan studi tersebut, populasinya adalah seluruh TKW Indonesia yang memiliki kemampuan bahasa asing. Tentunya, perusahaan tersebut agak sulit untuk mendata seluruh TKW Indonesia yang memiliki kemampuan bahasa asing. Sehingga diambil sampel random dari TKW Indonesia yang memiliki kemampuan bahasa asing yang bisa mewakili kondisi populasi.

Berdasarkan ilustrasi di atas, dapat dijelaskan pengertian populasi dan sampel. **Populasi** mengacu pada seluruh elemen yang akan diobservasi untuk mendukung kesimpulan yang diambil, yaitu seluruh TKW Indonesia yang memiliki kemampuan bahasa asing. Sedangkan **sampel** adalah elemen observasi yang diambil secara random dari populasi, yaitu individu-individu TKW Indonesia yang memiliki kemampuan bahasa asing. Yang perlu dicatat adalah sampel yang diambil harus **random.** Tujuannya adalah agar kesimpulan yang diambil tidak bias. Suatu sampel dikatakan random apabila setiap elemen dalam populasi memiliki kesempatan yang sama untuk dipilih menjadi sampel.

1.4 Pengertian dan Cara Menyusun Distribusi Frekuensi

Pada umumnya, data tidak langsung tersedia dengan rapi (siap pakai). Dari kumpulan data mentah tersebut, perlu dilakukan pengaturan ataupun pengelompokan agar data menjadi lebih ringkas dan mudah dibaca. Akan tetapi, peringkasan dan pengelompokan data tersebut tidak boleh mengurangi inti informasi. Untuk itu, perlu dibuat tabel distribusi frekuensi. Sederhananya, frekuensi dapat diartikan sebagai banyaknya data yang muncul pada interval tertentu. Sedangkan frekuensi relatif diartikan sebagai kemunculan data dalam setiap interval dibagi dengan jumlah total.

Berdasarkan tabel distribusi frekuensi, secara umum dapat dilihat berapa nilai tengah, nilai yang paling sering muncul (modus), dan pemusatan data. Salah satu acuan dalam pengelompokan data adalah kemiripan karakteristik. Tujuan dari pengelompokan data tersebut adalah jumlah kelompok data relatif kecil, memudahkan penghitungan dan analisa. Tetapi kekurangan dari pengelompokan data tersebut adalah ada beberapa informasi rinci yang tidak tampak lagi dalam susunan data.

Penyusunan distribusi frekuensi dapat dilakukan sebagai berikut. Untuk keperluan pengelompokan, data mentah perlu diurutkan dari data yang paling kecil sampai besar. Sebagai contoh adalah data penjualan donat harian di warung P'Udin periode 1 Juni-20 Juli 2006 (50 hari).

Jumlah penjualan donat per hari (Sebelum Diurutkan)

55	48	22	49	78	59	27	41	68	54
34	80	68	42	75	51	76	45	32	53
66	32	64	47	76	58	75	60	35	57
73	38	30	44	54	57	72	67	51	86
25	37	69	71	52	25	47	63	59	64

Jumlah penjualan donat per hari (Setelah Diurutkan)

22	25	25	27	30	32	32	34	35	37
38	41	42	44	45	47	47	48	49	51
51	52	53	54	54	55	57	57	58	59
59	60	63	64	64	66	67	68	68	69
71	72	73	75	75	76	76	78	80	86

Setelah data diurutkan, dapat dibuat tabel distribusi frekuensi. Beberapa aturan yang harus diperhatikan dalam penyusunan distribusi frekuensi adalah:

1. Tentukan nilai terkecil, nilai terbesar, dan jarak (nilai terbesar dengan nilai terkecil). Berdasarkan data penjualan donat periode 1 Juni-20 Juli

2006 dapat dilihat jumlah penjualan terendah dan tertinggi, yaitu masing-masing sebesar 22 dan 86. Sedangkan jarak antara penjualan terendah dan tertinggi adalah 64.

2. Banyaknya kelas pada umumnya adalah 5-15 kelas. Penentuan banyaknya kelas ini tergantung pada keperluan. Di sini, metode statistik tidak memberikan ukuran tertentu dalam hal penentuan kelas.

Jumlah kelas dari suatu distribusi bisa ditentukan berdasarkan karakteristik data mentah dan tujuan penggunaan data tersebut. Meskipun tidak ada ukuran tertentu di statistik, ada satu rumus rujukan dari ahli statistik yang dapat digunakan untuk menentukan jumlah kelas, yaitu:

$$k = 1 + 3.322 \log N$$

Dimana: k = jumlah kelas

N = jumlah data yang diobservasi

Berdasarkan contoh di atas, jumlah kelas dapat ditentukan sebagai berikut:

$$k = 1 + 5.64 = 6.64 \approx 7$$

Selanjutnya, interval kelas dapat ditentukan dengan membagi jarak dengan jumlah kelas. Jadi berdasarkan contoh di atas, interval kelas adalah:

$$= \frac{64}{7} = 9.14 \approx 10$$

1.5 Distribusi Frekuensi Relatif dan Kumulatif

Distribusi frekuensi bisa diklasifikasikan menjadi distribusi relatif dan kumulatif. Singkatnya, distribusi frekuensi relatif merupakan pembagian masing-masing frekuensi kelas dengan total frekuensi. Sehingga hasil dari distribusi frekuensi relatif berupa proporsi; pada umumnya dinyatakan dalam persen. Sedangkan distribusi frekuensi kumulatif merupakan penjumlahan masing-masing kelas dari distribusi frekuensi relatif; pejumlahan pada kelas terakhir merupakan nilai total kelas.

Sebagai contoh adalah pilihan pekerjaan bagi mahasiswa yang sudah lulus dari Universitas “Berwarna” per Fakultas. Total mahasiswa yang lulus pada periode Agustus 2006 adalah 300 mahasiswa. Berdasarkan pilihan pekerjaan per fakultas yang diminati oleh mahasiswa yang lulus pada periode tersebut, dapat diperoleh distribusi frekuensi relatif dan kumulatif sebagai berikut. Selanjutnya dinyatakan dalam persen.

Tabel 1.1
Pilihan Pekerjaan dan Minat (Pilihan) Mahasiswa yang Lulus
Periode Agustus 2006

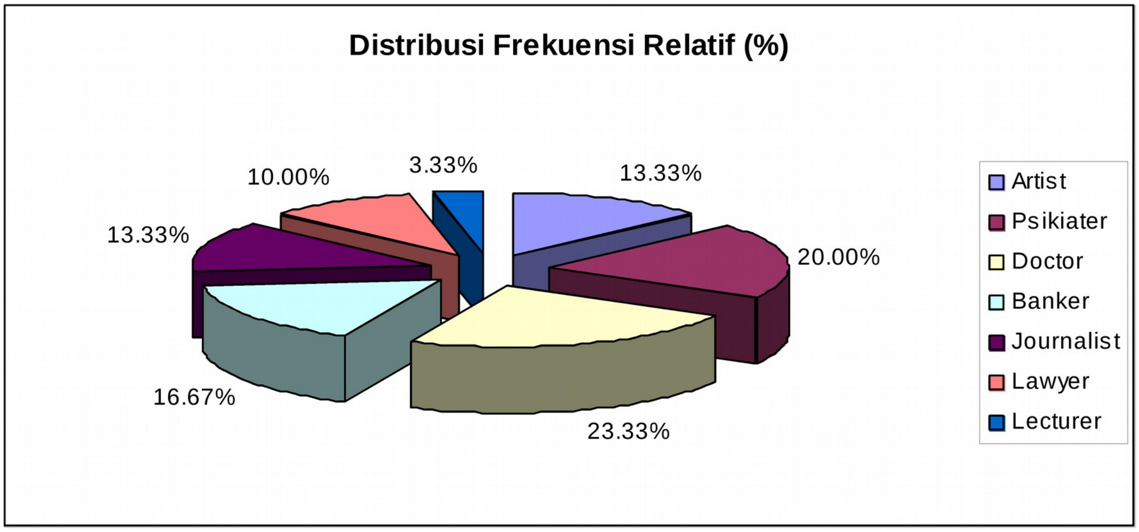
Pilihan Pekerjaan	Fakultas	Jumlah Lulusan Per Fakultas 2006	Distribusi Frekuensi Relatif (%)	**Distribusi Frekuensi Kumulatif (%)**
Artist	Budaya	40	13.3	**13.33**
Psikiater	Psikologi	60	20.0	**33.33**
Doctor	Kedoktera n	70	23.3	**56.67**
Banker	Ekonomi	50	16.7	**73.33**
Journalist	FISIP	40	13.3	**86.67**
Lawyer	Hukum	30	10.0	**96.67**
Lecturer	Keguruan	10	3.3	**100.00**
		300	**100**	

1.6 Gambar dan Grafik Distribusi Frekuensi

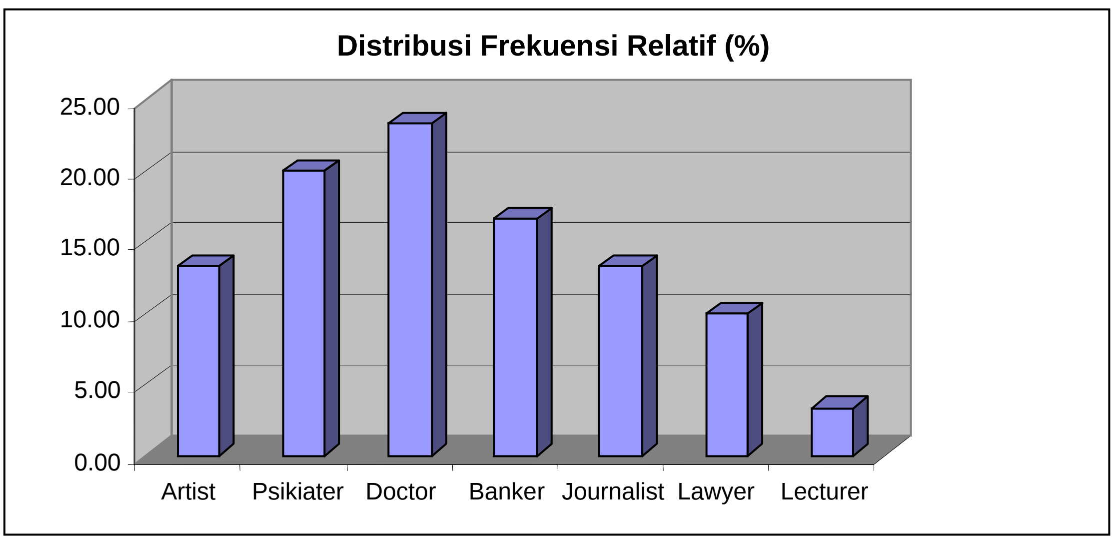
Penyajian data bisa dilakukan dengan berbagai cara. Tujuannya adalah agar data lebih mudah untuk dibaca dan dipahami jika dibandingkan dengan penyajian data dalam bentuk tabel frekuensi. Beberapa bentuk penyajian data berupa histogram, polygon frekuensi dan diagram pie.

Dengan menggunakan contoh di atas, berbagai tampilan data dalam bentuk grafik dapat disajikan sebagai berikut. Pada gambar 1.1.a dan 1.1.b disajikan diagram pie dan diagram batang dari distribusi frekuensi relatif. Sedangkan pada gambar 1.2.a dan 1.2.b disajikan diagram pie dan diagram batang dari distribusi frekuensi kumulatif.

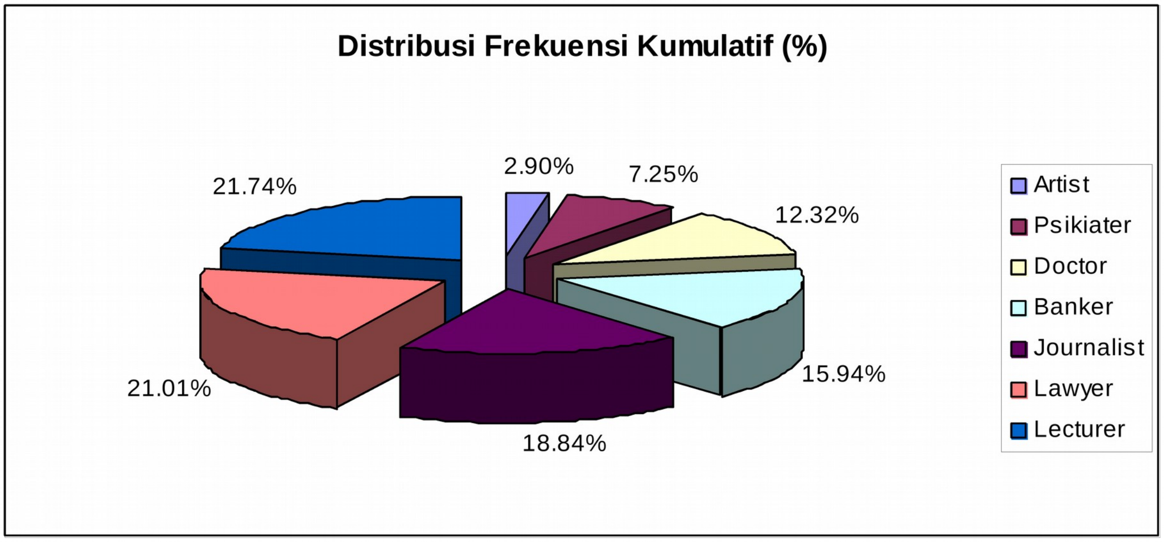
Gambar 1.1.a Diagram Pie



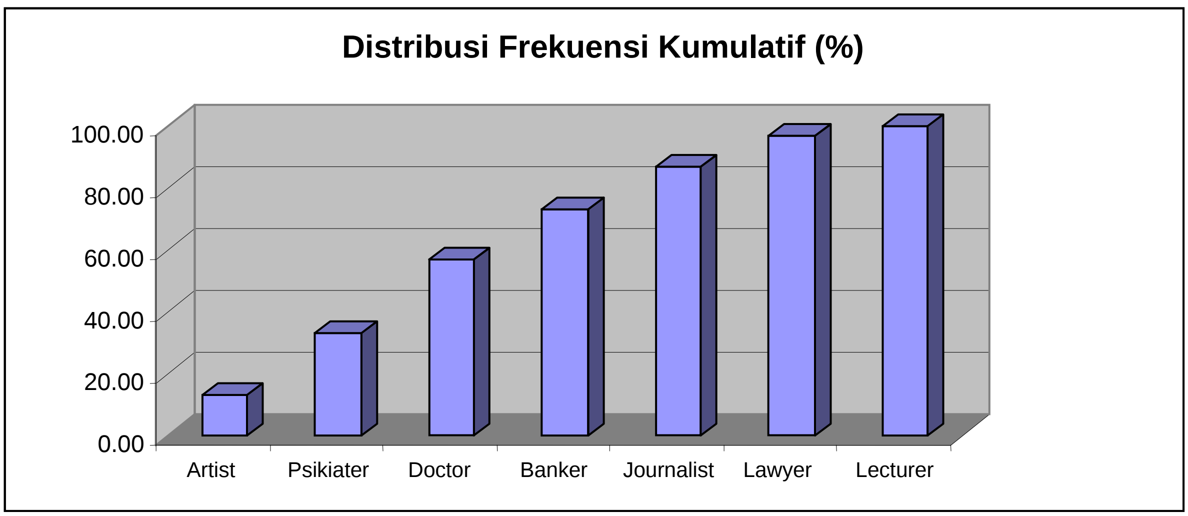
Gambar 1.1.b Diagram Batang



Gambar 1.2.a Diagram Pie



Gambar 1.2.b Diagram Batang



1.7 Latihan: Soal & Jawab

1. Berikut ini adalah data dari usia masyarakat di distrik A yang mendapatkan pelayanan sosial dari pemerintah.

83	51	66	61	82	65	54	56	92	60
65	87	68	64	51	70	75	66	74	68
44	55	78	69	98	67	82	77	79	62
38	88	76	99	84	47	60	42	66	74
91	71	83	80	68	65	51	56	73	55

Berdasarkan data tersebut, buatlah distribusi frekuensi relatif dengan membuat 7 dan 13 kelas.

Jawab:

Tabel Distribusi Frekuensi Relatif (7 kelas)

No.	Kelas	Frek. Relatif
1	30-39	0.02
2	40-49	0.06
3	50-59	0.16
4	60-69	0.32
5	70-79	0.20
6	80-89	0.16
7	90-99	0.08
		1.00

Tabel Distribusi Frekuensi Relatif (13 kelas)

No.	Kelas	Frek. Relatif
1	35-39	0.02
2	40-44	0.04
3	45-49	0.02
4	50-54	0.08
5	55-59	0.08
6	60-64	0.10
7	65-69	0.22
8	70-74	0.10
9	75-79	0.10
10	80-84	0.12
11	85-89	0.04
12	90-94	0.04
13	95-99	0.04
		1.00

2. Berdasarkan soal (1), pihak pemerintah mensyaratkan bahwa 50% dari penerima jasa pelayanan sosial memiliki usia lebih dari 50 tahun. Apakah persyaratan tersebut terpenuhi?

   Jawab:

   Berdasarkan distribusi di atas (baik 7 ataupun 13 kelas) menunjukkan bahwa sekitar 90% penerima jasa pelayanan sosial memiliki usia lebih dari 50 tahun. Jadi target kebijakan pemerintah sesuai dengan yang diharapkan.

3. Berdasarkan soal (1), apakah pembagian kelas menjadi 13 lebih membantu dalam memberikan informasi data dibandingkan dengan pembagian kelas menjadi 7.

Jawab:

Untuk kasus ini, kedua kategori kelas cukup mudah untuk digunakan. Jadi, sama saja apakah menggunakan pembagian kelas menjadi 7 atau 13 kategori.

4. Berdasarkan soal (1), jika pemerintah ingin mengetahui proporsi masyarakat yang mendapatkan pelayanan sosial dengan usia antara 45-50 tahun, manakah yang lebih baik, distribusi frekuensi dengan 7 atau 13 kelas.

   Jawab:

   Dalam hal ini, pembagian kelas menjadi 13 kategori lebih bagus dalam memberikan informasi. Karena pembagian kelas menjadi 7 kategori hanya menyediakan kelas dengan interval 40-49.

5. Berikut adalah data skor anak SMU dan SMK di bawah. Berdasarkan data di bawah, susunlah data mulai dari skor terendah sampai dengan tertinggi dengan berbasis pada skor SMU.

SMU	SMK	SMU	SMK
3.6	2.5	3.4	3.6
2.6	2.7	2.9	3.0
2.7	2.2	3.9	4.0
3.7	3.2	3.2	3.5
4	3.8	2.1	2.5
3.5	3.6	2.2	2.8
3.5	3.8	3.4	3.4
2.2	3.5	3.6	3.0
3.9	3.7	2.6	1.9
4.0	3.9	2.4	3.2

Jawab:

Skoring Data: SMU

SMU	SMK	SMU	SMK
4.0	3.8	3.4	3.4
4.0	3.9	3.2	3.5
3.9	3.7	2.9	3.0
3.9	4.0	2.7	2.2
3.7	3.2	2.6	2.7

3.6	2.5	2.6	1.9
3.6	3.0	2.4	3.2
3.5	3.6	2.2	3.5
3.5	3.8	2.2	2.8
3.4	3.6	2.1	2.5

6. Berdasarkan soal (5), susunlah data mulai dari skor terendah sampai dengan data tertinggi dengan berbasis pada skor SMK.

   Jawab:

   Skoring Data: SMK

SMK	SMU	SMK	SMU
4.0	3.9	3.2	3.7
3.9	4.0	3.2	2.4
3.8	4.0	3.0	2.9
3.8	3.5	3.0	3.6
3.7	3.9	2.8	2.2
3.6	3.5	2.7	2.6
3.6	3.4	2.5	3.6
3.5	2.2	2.5	2.1
3.5	3.2	2.2	2.7
3.4	3.4	1.9	2.6

7. Berdasarkan jawaban no. 5 dan 6, apa yang dapat anda simpulkan dari data tersebut.

   Jawab:

   Skoring siswa SMU dan SMK tidak jauh berbeda, artinya skoring tinggi di SMU beriringan dengan di SMK dan seterusnya sampai skoring terendah meskipun ada beberapa pengecualian

8. Sebuah perusahaan jasa pengantar kue memiliki jadwal untuk mengantar pesanan (jumlah hari dalam sebulan) ke beberapa tempat pada periode tertentu. Buatlah tabel distribusi frekuensi relatif dengan interval kelas sebanyak 6 hari.

4	12	8	14	11	6	7	13	13	11
11	20	5	19	10	15	24	7	29	26

   Jawab:

Kelas	Frekuensi	Frekuensi Relatif
1-6	4	0.20
7-12	8	0.40
13-18	4	0.20

19-24	3	0.15
25-30	1	0.05

9. Berdasarkan soal (8), bagaimana efektifitas dari jasa pengantar kue tersebut berdasarkan hasil dari distribusi frekuensi di atas.

   Jawab:

   Dengan asumsi jasa pengantar kue buka selama 6 hari dalam satu minggu, 80% pemesanan dapat diselesaikan dalam 3 minggu atau kurang.

10. Berdasarkan soal (8), jika perusahaan ingin memastikan bahwa 50% dari pemesanan bisa diselesaikan dalam 10 hari atau kurang, apakah tabel distribusi frekuensi cukup memberikan informasi?

    Jawab:

    Berdasarkan tabel dist

    ribusi frekuensi, hanya digambarkan bahwa 20%-60% pemesanan dapat dipenuhi dalam 10 hari atau kurang. Jadi tabel di atas tidak memberikan cukup informasi.

11. Berdasarkan soal (8), apa manfaat dari perhitungan frekuensi relatif pada tabel di atas.

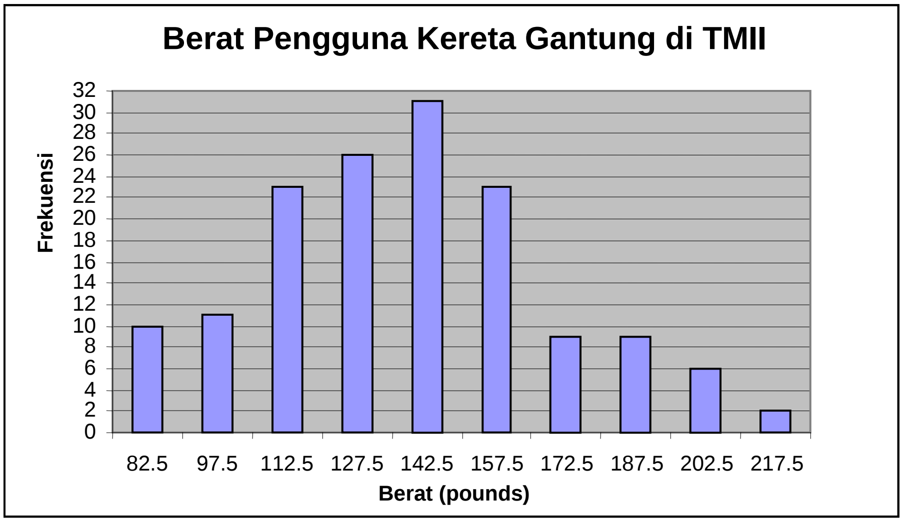
    Jawab:

    Distribusi frekuensi relatif membantu kita dalam menginterpretasikan informasi yang ada pada data dalam bentuk persentase ataupun proporsi.

12. Berikut ini adalah bobot (dalam pounds) dari 150 orang yang naik kereta gantung di TMII. Berdasarkan data di bawah, buatlah kurva histogram.

Kelas	Frekuensi	Kelas	Frekuensi
75-89	10	150-164	23
90-104	11	165-179	9
105-119	23	180-194	9
120-134	26	195-209	6
135-149	31	210-224	2

Jawab:



13. Berdasarkan soal (12), informasi apa yang dapat anda gali dari histogram di atas yang tidak bisa digambarkan melalui sajian data dalam bentuk tabel.

    Jawab:

    Banyak observasi yang berada di bagian bawah *(lower tail)* dibandingkan dengan di bagian atas (*upper tail).*

14. Berdasarkan soal (12), jika pihak manajemen sudah menentukan bahwa berat maksimum untuk setiap kereta gantung adalah 400 pounds, apa yang dapat dilakukan operator untuk memaksimumkan penggunaan kereta gantung.

    Jawab:

    Pihak operator bisa menggabungkan orang yang memiliki berat badan besar dan kecil. Hal ini dapat dibantu melalui pembuatan tabel di atas.

15. Dari rumah sakit bayi dan anak "Sangat Sehat" diketahui data berat badan bayi (dalam pounds). Buatlah tabel frekuensi kumulatif dari data berikut.

Kelas	Frekuensi	Kelas	Frekuensi
0.5-0.9	10	2.5-2.9	31
1.0-1.4	11	3.0-3.4	23
1.5-1.9	23	3.5-3.9	9
2.0-2.4	26	4.0-4.4	9

Jawab:

Kelas	Frek. Kumulatif	Kelas	Frek. Kumulatif
0.5-0.9	0.050	2.5-2.9	0.545
1.0-1.4	0.145	3.0-3.4	0.715
1.5-1.9	0.265	3.5-3.9	0.915
2.0-2.4	0.400	4.0-4.4	1.000

16. Berdasarkan soal (15), kira-kira berapa nilai tengah dari set data berat bayi di atas.

    Jawab:

    Nilai tengah dari berat bayi di atas sekitar 2.8 pounds.

17. Berdasarkan soal (15), jika bayi dengan berat di bawah 3 pounds membutuhkan inkubator untuk beberapa hari perawatan, berapa persentase bayi yang butuh inkubator.

    Jawab:

    Berdasarkan tabel frekuensi kumulatif di atas, sekitar 55% bayi membutuhkan inkubator.

18. Berikut ini adalah data dari produksi batubara (dalam ton) sebanyak 35 shift. Berdasarkan data tersebut, buatlah distribusi frekuensi relatif dan kumulatif dengan membuat 6 kelas yang seimbang.

356	331	299	391	364	317	386
360	281	360	402	411	390	362
311	357	300	375	427	370	383
322	380	353	371	400	379	380
369	393	377	389	430	340	368

Jawab:

Kelas	Frekuensi	Frek. Relatif	Frek. Kumulatif
281-305	3	0.09	0.09
306-330	3	0.09	0.18
331-355	3	0.09	0.26
356-380	15	0.43	0.69
381-405	8	0.23	0.92
406-430	3	0.09	1.00
	35	1.00	

19. Berdasarkan soal (18), jika diperkirakan produksi optimal antara 330-380 ton per shift, berapa persentase produksi di bawah dan di atas produksi optimal.

    Jawab:

    Berdasarkan tabel di atas dapat dilihat jumlah produksi yang di bawah produksi optimal sekitar 26%. Sedangkan produksi yang di atas produksi optimal sekitar 31%.

20. Berdasarkan soal (18), informasi apa yang perlu digali lagi berkaitan dengan produksi batu bara.

    Jawab:

    Data di atas hanya menunjukkan produksi optimal (antara 330-380 ton per shift) tanpa memberikan penjelasan bagaimana produksi optimal tersebut dicapai. Sehingga tidak bisa dilakukan evaluasi bagi shift yang belum mencapai produksi optimal.

MODUL 2
DISTRIBUSI FREKUENSI, UKURAN SENTRAL & UKURAN PENYEBARAN

Pokok Bahasan:

- Rerata Hitung Data Terkelompok dan Tidak Terkelompok
- Median Data Terkelompok dan Tidak Terkelompok
- Perhitungan Kuartil dan Persentil
- Modus (*Mode*) Terkelompok dan Tidak Terkelompok
- Pengertian tentang Persebaran

- Deviasi Standar Data Terkelompok dan Tidak Terkelompok
- Koefisien Variasi

2.1 Rerata Hitung dan Rerata Ukur Data

Mean atau nilai rerata, dinotasikan dengan $\overline{x}$, diperoleh dengan menjumlahkan semua data dan membaginya dengan jumlah pengamatan. Sederhananya dapat diilustrasikan sebagai berikut:

$$\overline{x} = \frac{\sum \overline{x}_i}{n}$$

dimana: *i =1, 2, ..., n*

$$\overline{x} = \frac{x_1 + x_2 + ... + x_n}{n}$$

Sebagai contoh, dicari rerata waktu yang dihabiskan responden untuk menonton TV selama seminggu. Di sini ada 5 responden dimana masing-masing orang menghabiskan waktu untuk menonton TV sebanyak (dalam jam) 5, 7, 3, 38, dan 7. Rerata dapat dicari sebagai berikut:

$$\overline{x} = \frac{5+7+3+38+7}{5} = \frac{60}{5} = 12$$

Jadi, rerata waktu dari 5 responden yang dihabiskan untuk menonton TV selama seminggu adalah 12 jam.

Rerata dapat diartikan sebagai nilai khas yang mewakili suatu himpunan data. Nilai khas cenderung terletak secara terpusat dan merupakan nilai yang mewakili seluruh nilai observasi. Ada dua jenis rerata, yaitu rerata hitung dan rerata ukur. Masing-masing rerata digunakan untuk data yang tidak dikelompokkan (data mentah) dan data yang dikelompokkan dalam distribusi frekuensi.

2.1.1 Rerata hitung

Data Tidak Dikelompokkan.

Rata-rata hitung dari nilai observasi x_1, x_2, ..., x_n adalah hasil penjumlahan semua nilai observasi dibagi dengan jumlah observasi. Secara matematis, dapat dinyatakan dengan:

$$\pi = \frac{\sum x_i}{N}$$

Data Dikelompokkan.

Pengelompokan observasi dalam distribusi frekuensi perlu dilakukan karena jumlah observasi cukup besar. Tujuannya adalah untuk memudahkan perhitungan. Secara matematis, dapat dinyatakan dengan:

(i)Metode Langsung, dirumuskan dengan $\pi = \dfrac{\sum fm}{\sum f}$

(ii) Metode *Short-Cut*, dirumuskan dengan $\pi = \pi_a + \dfrac{\sum fd}{N} i$

Keterangan:

π_a = rata-rata hitung yang diasumsikan

f = frekuensi kelas

f_m = frekuensi kelas median

d = penyimpangan nomor interval kelas

N = jumlah frekuensi

i = interval kelas

2.1.2 Rerata ukur

Rerata ukur (*geometric mean*) dari nilai sejumlah *n* observasi adalah akar pangkat *n* dari perkalian seluruh nilai data. Salah satu kegunaan rerata ukur adalah untuk melihat tingkat pertumbuhan *(rate of growth)*. Tujuannya adalah untuk mengurangi bias yang disebabkan nilai x_i yang ekstrim.

Data Tidak Dikelompokkan.

Rumus rerata ukur dinyatakan dengan:

$$G = \sqrt{n_1 \times n_2 \times n_3 ...}$$

dimana n_1, n_2 menunjukkan observasi ke-1 dan selanjutnya.

Data Dikelompokkan.

Rumus rerata ukur dinyatakan dengan:

$$\log G = \sum \frac{(\log m) f}{N}$$

dimana G merupakan *anti log* dari rumus di atas.

2.2 Median Data

Tahap awal yang harus dilakukan untuk memperoleh nilai median adalah mengurutkan data dari data yang terkecil sampai dengan yang terbesar. Secara definisi, median diartikan sebagai nilai yang berada di tengah ketika data diurutkan menurut besarnya. Jika jumlah titik data adalah genap, maka nilai tengah merupakan rata-rata dari dua nilai yang berada di tengah.

Sebagai contoh, dicari median dari data berat badan empat orang anak TK; masing-masing sebesar (dalam kg) 20, 25, 27, dan 30. Median dari berat badan anak TK tersebut adalah 26.

Seperti halnya pada rerata, median juga digunakan untuk data yang dikelompokkan dan data yang tidak dikelompokkan.

Data Tidak Dikelompokkan.

Dari sejumlah *N* data yang telah diurutkan, jika *N* adalah ganjil, maka median adalah data yang terletak di tengah. Tetapi jika *N* adalah genap, maka median adalah rerata hitung dari dua data yang terletak di tengah. Dirumuskan dengan :

$$Median = Md = \frac{(N+1)}{2}$$

Data Dikelompokkan.

Untuk data yang dikelompokkan, secara matematis, dapat dinyatakan dengan:

$$Median = Md = Lm + \frac{\left(N/2 - \sum f\right)}{fm} i$$

Keterangan:

Lm = batas bawah kelas median

N = jumlah frekuensi

$\sum f$ = frekuensi kumulatif dari atas pada kelas sebelum kelas median

fm = frekuensi kelas median

i = interval kelas median

2.3 Perhitungan Kuartil dan Persentil

Kuartil membagi serangkaian data yang sudah diurutkan menurut besarnya menjadi empat bagian. Secara notasi dapat dinyatakan dengan Q_1 (kuartil pertama), Q_2 (kuartil kedua), dan Q_3 (kuartil ketiga). Di sini, kuartil kedua tidak lain adalah median. Rumus kuartil ke-1 dan ke-3 dapat dinyatakan dengan

$$Q_1 = L_m + \frac{\left(\frac{N}{4} - \Sigma f\right)}{f_m} i$$

$$Q_3 = L_m + \frac{\left(\frac{3N}{4} - \Sigma f\right)}{f_m} i$$

Persentil adalah nilai yang membagi serangkaian data menjadi seratus bagian yang sama dan dinyatakan dengan P_1, P_2, P_3, ...,Pn. Rumus persentil dinyatakan dengan

$$P_k = L_m + \frac{\left({kN}/{100} - \Sigma f \right)}{f_m} i$$

2.4 Modus (*Mode*) Terkelompok dan Tidak Terkelompok

Modus adalah nilai yang paling sering muncul dari serangkaian data. Serangkaian data bisa memiliki dua modus (bimodal) atau lebih dari dua modus (multimodal). Munculnya bimodal kadang-kadang disebabkan oleh penggabungan dua distribusi yang berbeda.

Untuk data yang dikelompokkan, rumus modus adalah:

$$Mo = Lm + \left(\frac{d_1}{d_1 + d_2} \right) i$$

Keterangan:

Lm = batas bawah kelas modus

d_1 = selisih antara frekuensi kelas modus dengan frekuensi kelas sebelum modus

d_2 = selisih antara frekuensi kelas modus dengan frekuensi kelas setelah modus

i = interval kelas modus

2.5 Pengertian tentang Persebaran

Ukuran-ukuran persebaran merupakan salah topik penting selain ukuran pemusatan. Singkatnya, persebaran menunjukkan distribusi data pada suatu rentang tertentu. Persebaran data juga dikenal dengan dispersi atau variabilitas data.

Yang menjadi pertanyaan lebih lanjut adalah, mengapa variabilitas data merupakan salah satu karakteristik yang penting untuk dipelajari dan diukur? Pertama, memberikan tambahan informasi untuk menilai reliabilitas ukuran terpusat. Jika data tersebar dalam rentang yang sangat panjang, ukuran terpusat menjadi kurang representatif. Kedua, karena data tersebar dalam rentang yang panjang, kita harus mampu memahami karakteristik data tersebut sebelum menganalisis masalah. Ketiga, bisa dilakukan komparasi dispersi data dari berbagai sampel.

Dalam konteks dispersi, analis keuangan sangat memperhatikan pendapatan perusahaan. Apabila dispersi pendapatan perusahaan sangat luas (mulai dari pendapatan dengan jumlah yang sangat tinggi sampai dengan yang sangat rendah atau bahkan negatif), hal ini mengindikasikan resiko yang ditanggung pemegang saham sangat besar.

2.6 Deviasi Standar Data Terkelompok dan Tidak Terkelompok

Ukuran yang Data Tidak Dikelompokkan. Rumus deviasi standar dinyatakan dengan:

$$\sigma = \sqrt{\frac{\Sigma (x - \pi)^2}{N}}$$

Data Dikelompokkan. Rumus deviasi standar dinyatakan dengan:

$$\sigma = \sqrt{\frac{\Sigma\, f(m - \pi)^2}{N}} \rightarrow \text{ untuk populasi}$$

$$s = \sqrt{\frac{\Sigma\, f(m - \pi)^2}{N - 1}} \rightarrow \text{ untuk sampel}$$

Keterangan:

f = frekuensi kelas

m = titik tengah kelas

N = jumlah frekuensi

π = rata-rata hitung dari data yang telah dikelompokkan

2.7 Koefisien Variasi

Koefisien Variasi

Koefisien variasi merupakan ukuran penyebaran relatif yang menunjukkan persentase standar deviasi suatu distribusi terhadap rata-ratanya.

Rumus koefisien variasi adalah:

$$KV = \frac{\sigma}{\pi} x100\%$$

Koefisien variasi bebas dari satuan yang digunakan sehingga berguna untuk membandingkan distribusi dengan satuan yang berbeda.Dari contoh di atas, dapat dibuat ukuran pemusatan dan ukuran penyebaran baik untuk data yang tidak dikelompokkan maupun data yang dikelompokkan.

2.8 Latihan: Soal & Jawab

1. Berdasarkan group data besarnya belanja ibu rumah tangga (dalam US$), carilah rata-rata sampel group data berikut.

Kelas (US$)	Frekuensi
0-49.99	78
50.00-99.99	123
100.00-149.99	187
150.00-199.99	82
200.00-249.99	51
250.00-299.99	47
300.00-349.99	13
350.00-399.99	9
400.00-449.99	6
450.00-499.99	4

Jawab:

Kelas (US$)	Nilai Tengah (x)	Frekuensi (f)	*xf*
0-49.99	25	78	1,950

50.00-99.99	75	123	9,225
100.00-149.99	125	187	23,375
150.00-199.99	175	82	14,350
200.00-249.99	225	51	11,475
250.00-299.99	275	47	12,925
300.00-349.99	325	13	4,225
350.00-399.99	375	9	3,375
400.00-449.99	425	6	2,550
450.00-499.99	475	4	1,900
		600	85,350

$$\sum f = \sum n = 600$$

$$\sum(xf) = 85350$$

$$\bar{x} = \frac{\sum(xf)}{\sum f} = \frac{85350}{600} = 142.25$$

Jadi rata-rata sampel dari belanja ibu RT dari berbagai kategori di atas adalah US$ 142.25

2. Apa yang dimaksung dengan *coding?* Buatlah *coding* berdasarkan data di bawah.

Kelas	1-5	6-10	11-15	16-20	21-25	26-30	31-35	36-40	41-45

Jawab:

Coding secara sederhana dapat dinyatakan sebagai proses penyederhanaan data dengan membuat *kode* untuk setiap nilai tengah dari kelas data. Pada umumnya, diletakkan nilai nol pada distribusi frekuensi yang terletak di tengah.

Kelas	1-5	6-10	11-15	16-20	21-25	26-30	31-35	36-40	41-45
Kode (u)	-4	-3	-2	-1	0	1	2	3	4

3. Berdasarkan data di bawah, buatlah *coding*-nya. dan carilah nilai rata-ratanya.

Kelas	0-7	8-15	16-23	24-31	32-39	40-47
Nilai Tengah	3.5	11.5	19.5	27.5	35.5	43.5
Frekuensi	2	6	3	5	2	2

Jawab:

Kelas	0-7	8-15	16-23	24-31	32-39	40-47
Nilai Tengah	3.5	11.5	19.5	27.5	35.5	43.5
Koding	-2	-1	0	1	2	3

4. Berdasarkan soal (3), carilah rata-rata dari kelas data di atas.

   Jawab:

Kelas	Nilai Tengah	Koding (u)	Frekuensi	uf
0-7	3.5	-2	2	-4
8-15	11.5	-1	6	-6
16-23	19.5	0	3	0
24-31	27.5	1	5	5
32-39	35.5	2	2	4
40-47	43.5	3	2	6
			20	5

$$\sum f = \sum n = 20$$

$$\sum (xf) = 5$$

$$\overline{x} = x_0 + w\frac{\sum (uf)}{\sum f} = 19.5 + 8\frac{5}{20} = 19.5 + 2 = 21.5$$

Dimana:

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

x_0 = nilai tengah yang memiliki kode 0

w = lebar kelas interval

u = koding untuk setiap kelas

f = frekuensi untuk setip kelas observasi

n = jumlah total observasi dalam sampel

5. Berikut ini adalah daftar berat paket (dalam pounds) yang ada di perusahaan pengiriman "Sangat Cepat".

Kelas	Frekuensi	Kelas	Frekuensi
10.5-10.9	1	15.0-15.9	11
11.0-11.9	4	16.0-16.9	8
12.0-12.9	6	17.0-17.9	7
13.0-13.9	8	18.0-18.9	6
14.0-14.9	12	19.0-19.9	2

Berdasarkan data tersebut, tentukan nilai tengah kelas, *coding*, dan data yang dibutuhkan untuk menentukan berbagai ukuran statistik.

Jawab:

Kelas	Frekuensi (f)	(Jawaban No. 6) Nilai Tengah (x)	Fx	(No. 7) Koding (u)	uf	(No.8) Koding (u)	uf
10.5-10.9	1	10.5	10.5	-3	-3	-5	-5
11.0-11.9	4	11.5	46.0	-2	-8	-4	-16
12.0-12.9	6	12.5	75.0	-1	-6	-3	-18
13.0-13.9	8	13.5	108.0	0	0	-2	-16
14.0-14.9	12	14.5	174.0	1	12	-1	-12
15.0-15.9	11	15.5	170.5	2	22	0	0
16.0-16.9	8	16.5	132.0	3	24	1	8
17.0-17.9	7	17.5	122.5	4	28	2	14
18.0-18.9	6	18.5	111.0	5	30	3	18
19.0-19.9	2	19.5	39.0	6	12	4	8
	∑=65		∑=988.5		∑=111		∑=-19

6. Berdasarkan soal (5), hitunglah rata-rata sampel dengan menggunakan rumus berikut [$\bar{x} = \frac{\sum(fx)}{n}$].

   Jawab:

   $$\bar{x} = \frac{\sum(fx)}{n} = \frac{988.5}{65} = 15.2077\text{pounds}$$

   Jadi berat rata-rata dari paket yang akan dikirim adalah 15.2077 pounds.

7. Berdasarkan soal (5), hitunglah rata-rata sampel dengan menggunakan *coding*; dimana angka 0 diletakkan di kelas ke-4.

   Jawab:

   $$\bar{x} = x_0 + w\frac{\sum(uf)}{\sum f} = 13.5 + 1.0\frac{(111)}{65} = 15.2077$$

8. Berdasarkan soal (5), hitunglah rata-rata sampel dengan menggunakan *coding*; dimana angka 0 diletakkan di kelas ke-6.

   Jawab:

   $$\bar{x} = x_0 + w\frac{\sum(uf)}{\sum f} = 15.5 + 1.0\frac{(-19)}{65} = 15.2077$$

9. Jelaskan mengapa jawaban no (7) dan (8) adalah sama.

   Jawab:

   Karena penentuan *coding* tidak mempengaruhi hasil akhir. Jadi tidak masalah, apakah nilai 0 diletakkan di kelas ke-4, ke-6 ataupun kelas manapun, hasilnya akan sama.

10. Berikut ini adalah data yang menunjukkan persetujuan kredit (dalam US$) dari pihak manajemen bagi konsumen yang ingin membeli barang elektronik secara kredit. Perusahaan akan mengurangi tingkat bunga jika rata-rata kredit bulanan lebih dari US$ 65000. Apa yang diputuskan perusahaan berdasarkan data berikut.

Bulan	Kredit	Bulan	Kredit	Bulan	Kredit	Bulan	Kredit
Jan.	121300	April	72800	Juli	58700	Okt.	52800
Feb.	112300	Mei	72800	Agust.	61100	Nov.	49200
Maret	72800	Juni	57300	Sept.	50400	Des.	46100

   Jawab:

$$\bar{x} = \frac{\sum x}{n} = \frac{827600}{12} = US\$68967$$

   Jadi dapat disimpulkan bahwa pihak manajemen akan mengurangi tingkat bunga karena kredit bulanan lebih dari US$ 65000

11. Perusahaan roti “Sangat Enak” harus menempatkan berbagai level pekerja pada 2 jenis roti yang diproduksi, yaitu rati “Siap-Santap” dan roti “Siap-Siap”. Pekerja dibedakan menjadi pekerja tidak berpengalaman, cukup berpengalaman, dan berpengalaman. Berikut ini adalah jumlah pekerja yang dibutuhkan untuk tiap jenis roti (dalam jam/unit output). Tentukan biaya roti “Siap-Santap” per unit pekerja.

Tingkat Pekerja	(US$)	Jumlah jam kerja/unit output	
	Upah/Jam	Roti “Siap-Santap”	Roti “Siap-Siap”
Tidak Berpengalaman	5	1	4
Cukup Berpengalaman	7	2	3
Berpengalaman	9	5	3

   Jawab:

$$\left(\frac{4}{10} x\$5\right) + \left(\frac{3}{10} x\$7\right) + \left(\frac{3}{10} x\$9\right) = \$6.8 / jam$$

Jadi, biaya yang dibutuhkan untuk setiap unit roti “Siap-Santap” per pekerja adalah US$ 6.8/jam

12. Berdasarkan soal (11), tentukan biaya roti “Siap-Siap” per unit pekerja dengan menggunakan rumus berikut $\left[\bar{x}_w = \dfrac{\sum(wx)}{\sum w}\right]$.

Jawab:

$$\left[\bar{x}_w = \frac{\sum(wx)}{\sum w}\right]$$

$$= \frac{\left(\frac{1}{8} x\$5\right) + \left(\frac{2}{8} x\$7\right) + \left(\frac{5}{8} x\$9\right)}{\frac{1}{8} + \frac{2}{8} + \frac{5}{8}}$$

$$= \frac{\$8}{1} = \$8 / \; jam$$

Jadi, biaya yang dibutuhkan untuk setiap unit roti “Siap-Siap” per pekerja adalah US$ 8/jam.

Dimana:

$\bar{x}_w =$ rata-rata tertimbang

$w =$ bobot untuk setiap observasi

$\sum(wx) =$ jumlah bobot setiap elemen dikaliakn dengan elemen tersebut

$\sum w =$ jumlah seluruh bobot (atau timbangan)

13. Apa yang dapat anda simpulkan dari jawaban no. 11 dan 12.

Jawab:

Biaya pekerja per jam yang dibutuhkan untuk membuat roti “Siap-Siap” lebih mahal dibandingkan dengan roti “Siap-Santap”

14. Perusahaan Retail “Sangat Besar” mengumumkan bahwa untuk beberapa jenis produk (produk A-F), pihak manajemen memastikan bahwa harga produk mereka adalah yang paling murah. Jika ada kompetitornya, Retail “Lain”, yang memberikan harga lebih murah, maka konsumen akan mendapat barang tersebut secara cuma-cuma. Bapak A yang sangat cermat dan teliti menemukan beberapa produk yang dibeli di retail “Sangat Besar” ternyata lebih mahal dibandingkan

dengan toko lain. Dari struk pembelian dapat diperoleh data sebagai berikut (dalam US$).

Retail	A	B	C	D	E	F
Retail “Sangat Besar”	1.35	2.89	3.19	4.98	7.59	11.5
Jumlah unit yang dibeli	7	9	12	8	6	3

Sedangkan harga produk yang sama di Retail “Lain” adalah:

Retail “Lain”	1.29	2.97	3.49	5	7.5	10.95

Berdasarkan data di atas, hitung rata-rata tak tertimbang (*unweighted average)* dari belanja Bapak A.

Jawab:

$$\bar{x}_{SB} = \frac{\sum x}{n} = \frac{31.5}{6} = \$5.25$$

$$\bar{x}_{L} = \frac{\sum x}{n} = \frac{31.2}{6} = \$5.20$$

Note:

$\bar{x}_{SB}$ = rata-rata belanja Bapak A di Retail "Sangat Besar"

$\bar{x}_{L}$ = rata-rata belanja Bapak A di Retail "Lain"

15. Berdasarkan soal (14), hitung rata-rata tertimbang (*weighted average)* dari belanja Bapak A.

Jawab:

$$\bar{x}_{SB} = \frac{\sum(wx)}{\sum w}$$

$$= \frac{7(1.35)+9(2.89)+12(3.19)+8(4.98)+6(7.59)+3(11.5)}{7+9+12+8+6+3}$$

$$= \frac{193.62}{45} = \$4.303$$

$$\bar{x}_L = \frac{\sum(wx)}{\sum w}$$

$$= \frac{7(1.29)+9(2.97)+12(3.49)+8(5.00)+6(7.50)+3(10.95)}{7+9+12+8+6+3}$$

$$= \frac{195.49}{45} = \$4.344$$

16. Informasi apa yang dapat anda sampaikan kepada konsumen berdasarkan jawaban no. 14 dan no. 15.

    Jawab:

    Apabila menggunakan *ukuran* rata-rata tidak tertimbang, rata-rata harga di Retail "Sangat Besar" lebih mahal dibandingkan dengan Retail "Lain". Akan tetapi sebaliknya, apabila menggunakan *ukuran* rata-rata tertimbang, rata-rata harga di Retail "Sangat Besar" lebih murah dibandingkan dengan Retail "Lain".

17. Pihak Manajemen Retail "Sangat Besar" mendasarkan pernyataannya berdasarkan harga rata-rata tertimbang. Apakah ada kemungkinan terjadinya masalah berkaitan dengan pernyataan pihak manajemen tersebut.

    Jawab:

    Potensi munculnya masalah dari pernyataan tersebut cukup besar. Mengapa? Karena konsumen tidak terlalu peduli dengan *istilah* "rata-rata tidak tertimbang" ataupun "rata-rata tertimbang". Yang menjadi fokus perhatian konsumen adalah pihak Retail "Sangat Besar" bersedia memberikan gratis jika ada pihak lain yang menjual produk yang sama dengan harga yang lebih rendah.

18. Dari soal no. 17, anda sebagai pihak yang netral, apakah pihak manajemen Retail "Sangat Besar" bisa dikatakan sudah melakukan kecurangan?

    Jawab:

    Tidak ada alasan yang cukup kuat untuk menyatakan pihak manajemen Retail "Sangat Besar" melakukan kecurangan. Karena pernyataan yang dibuat hanya merupakan salah satu strategi pemasaran untuk mendongkrak penjualan.

19. Perusahaan perkakas rumah “Bersahaja” memiliki pangsa pasar di beberapa daerah. Bagian pemasaran menyiapkan berbagai promosi untuk meningkatkan angka penjualan tahun depan. Berikut ini adalah data penjualan tahun sebelumnya dan tingkat pertumbuhan penjualan yang diharapkan di beberapa region.

Region	Penjualan (US$)	Ekspektasi Pertumbuhan (%)
A	193.8	7.25
B	79.3	8.20
C	57.5	7.15

Berdasarkan data di atas, hitung rata-rata tingkat pertumbuhan penjualan tahun depan.

Jawab:

$$\bar{x} = \frac{\sum(wx)}{\sum w}$$

$$= \frac{193.8(7.25) + 79.3(8.20) + 57.5(7.15)}{193.8 + 79.3 + 57.5}$$

$$= \frac{2466.435}{330.6} = 7.46\%$$

20. Berdasarkan data di atas, apakah dapat diperoleh informasi bahwa penjualan di Region A sangat tinggi karena jumlah penduduk yang besar dan daya beli yang lebih tinggi dibandingkan dengan Region B dan C.

Jawab:

Tidak bisa. Karena data di atas tidak menyertakan data jumlah penduduk dan daya beli di masing-masing region. Jadi faktor jumlah penduduk dan daya beli belum bisa dijadikan alasan tingginya jumlah penjualan di Region A.

MODUL 3

EKSPLORASI DATA

Pokok Bahasan:

- Pendahuluan: Pengantar ke Eksplorasi Data
- Histogram (*A Set Univariate Data*)
- Diagram Dahan Daun
- Diagram Kotak Garis (*Box Plot*) dan Pemeriksaan *Outliers* (pencilan)

3.1 Pendahuluan: Pengantar ke Eksplorasi Data

Exploratory data analysis (EDA) adalah salah satu cakupan statistik yang menfokuskan pada review, komunikasi, dan penggunaan data. Sederhananya, EDA adalah salah satu pendekatan yang digunakan untuk analisis data. Teknik EDA berkaitan dengan pemanfaatan data mentah dan pola pikir statistik. Atau bisa dinyatakan, EDA lebih mengacu pada filosofi bagaimana membedah data, apa yang dapat digali dari data tersebut, bagaimana melihat data, dan bagaimana menginterpretasikan data. EDA juga bisa didefinisikan sebagai penggunaan teknik “statistik grafik”. Mengapa? Karena salah satu fokus EDA adalah menggali atau mengeksplorasi data; dan salah satu caranya adalah melalui sajian grafis. Akan tetapi EDA tidak sepenuhnya identik dengan grafik. Beberapa penyajian grafis yang digunakan EDA cukup sederhana, yaitu:

1. Plot data mentah (seperti: dan *histogram)*

2. Plot statistik sederhana (seperti: plot rata-rata, plot standar deviasi, *box plot,* dan sebagainya)

Penyajian grafis lain yang biasa digunakan adalah *scatter plot* dan *stem-and-leaf plot.* Plot data dapat mencerminkan kondisi data secara kasar.

Tujuan dari EDA adalah:

1. Mendukung hipotesis tentang penyebab dari fenomena yang diobservasi
2. Menilai asumsi yang menjadi patokan statistik inferensi
3. Mendukung pemilihan instrumen dan metode statistik yang tepat
4. Menyediakan pedoman pengumpulan data yang lebih lanjut melalui survei atau eksperimen

Yang menjadi pertanyaan lebih lanjut adalah apa yang membedakan EDA dengan analisis data klasik lainnya? Ada tiga pendekatan yang pada umumnya digunakan untuk analisis data, yaitu:

1. Pendekatan Klasik (*Classical Approach)*
2. EDA *Approach*
3. Pendekatan Bayesian (*Bayessian Approach)*

Ditinjau dari paradigma dalam melakukan teknik analisis, tiga pendekatan di atas serupa dalam konteks bahwa tiga pendekatan tersebut berawal dari permasalahan dan solusi umum di bidang ilmu pengetahuan. Perbedaannya terletak pada alur dan fokus pada prosesnya. Berikut ini adalah alur dari masing-masing pendekatan.

1. Alur Pendekatan Klasik (*Classical Approach)*

   Problem => Data => Model => Analysis => Conclusions

2. Alur EDA *Approach*

   Problem => Data => Analysis => Model => Conclusions

3. Alur Pendekatan Bayesian (*Bayessian Approach)*

   Problem => Data => Model => Prior Distribution => Analysis => Conclusions

Selanjutnya, untuk pendekatan analisis klasik, kumpulan data diikuti dengan persyaratan yang harus dipenuhi model (seperti: normalitas, linearitas dan sebagainya) dan analsisis, estimasi, pengujian yang secara keseluruhan fokus pada parameter model. Untuk pendekatan EDA, kumpulan data tidak diikuti dengan berbagai macam persyaratan; akan tetapi, diikuti dengan analisis model yang tepat untuk menghasilkan inferensi yang diinginkan. Sedangkan untuk pendekatan Bayesian, dibuat analisis dengan mengakomodir informasi/pengetahuan awal ke dalam analisis dengan menerapkan syarat pada data, sebagai contoh, distribusi independen terhadap parameter yang dipilih dalam model. Selanjutnya, analisis merupakan kombinasi dari distribusi awal (*prior distribution)* dan kumpulan data yang secara bersama-sama digunakan untuk membuat inferensi dan atau menguji parameter model. Dalam aplikasinya, analisis data bisa merupakan *mix* dari ketiga pendekatan tersebut. Perbedaan pendekatan di atas hanya untuk menekankan fokus studi dari masing-masing pendekatan.

Yang perlu dipahami, statistik dan prosedur analisis data pada dasarnya dibedakan menjadi dua, yaitu kuantitatif dan grafis. Teknik kuantitatif adalah prosedur statistik dimana outputnya berupa angka (numerik) atau tabulasi. Sebagai contoh:

1. pengujian hipotesis (*hypothesis testing)*
2. analisis varians (*analysis of variance)*
3. titik estimasi dan interval kepercayaan (*point estimates and confidence intervals)*
4. *least squares regression*

Dalam konteks analisis klasik, semua teknik penting. Di sisi lain, ada beberapa instrumen statistik yang pada umumnya mengacu pada teknik grafis. Beberapa teknis grafis tersebut mencakup:

1. *scatter plots*
2. *histograms*
3. *probability plots*
4. *residual plots*
5. *box plots*
6. *block plots*
7. *steam-and-leaf plots*

Pada modul ini, hanya tiga teknik grafis yang dibahas, yaitu: *histograms, steam-and-leaf plots,* dan *box plots.*

Pendekatan EDA berkaitan erat dengan teknik grafis. Prosedur grafis bukan hanya melekat pada EDA tetapi juga suatu instrumen yang penting untuk dipilih atau digunakan. Berbagai instumen grafis merupakan cara yang paling mudah dan singkat untuk mendapatkan *insight* dari data berkaitan dengan:

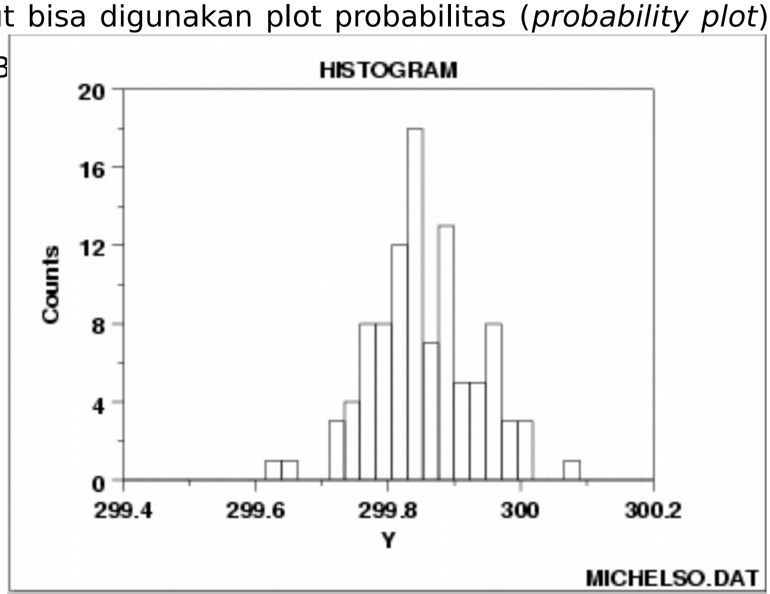
1. pengujian asumsi (*testing assumptions*)
2. pemilihan model (*model selection*)
3. validasi model (*model validation*)
4. pemilihan estimator (*estimator selection*)
5. identifikasi hubungan (*relationship identification*)
6. *factor effect determination*
7. deteksi outlier (*outlier detection*)

3.2 Histogram *(A Univariate Data Set)*

Tujuan histogram adalah membuat ringkasan secara grafis dari distribusi set data tunggal (*univariat data set).* Histogram secara grafis menunjukkan beberapa point berikut:

1. pemusatan data (*center of the data*)
2. persebaran data (*spread of the data*)
3. kecondongan data (*skewness of the data*)
4. adanya outlier (*presence of outliers*)
5. adanya modus ganda dalam data (*presence of multiple modes*)

Melalui histogram, dapat dilihat indikasi yang kuat bagaimana bentuk distribusi yang tepat pada data. Untuk melakukan verifikasi bentik distribusi dari data tersebut bisa digunakan plot probabilitas (*probability plot*) atau uji *goodness-of-fit.* B



Pada umumnya histogram diperoleh dengan membagi data menjadi beberapa kelas. Kelas tersebut bisa ditentukan berdasarkan tujuan ataupun berdasarkan aturan tertentu. Secara teoritis, aturan tersebut diajukan oleh Scott (Scott, 1992). Selanjutnya untuk setiap kelas, dihitung berapa frekuensi untuk setiap kelas. Dari gambar histogram, sumbu vertikal menunjukkan frekuensi dan sumbu horisontal menunjukkan variabel.

Histogram kumulatif adalah variasi histogram dimana sumbu vertikal tidak hanya terdiri dari satu kelas tetapi beberapa kelas. Semakin ke atas, frekuensi kelas makin kecil. Baik histogram atupun histogram kumulatif memiliki variasi yang disebut dengan histogram relatif dan histogram kumulatif relatif.

Ada dua cara untuk membuat normalisasi angka, yaitu:

1. Normalisasi angka adalah angka atau frekuensi dalam suatu kelas dibagi dengan total observasi. Angka atau frekuensi relatif tersebut jika dijumlahkan, hasilnya sama dengan satu (atau 100 jika digunakan skala persentase). Jadi ketinggian histogram menunjukkan proporsi data setiap kelas.
2. Normalisasi angka adalah angka atau frekuensi di kelas dibagi dengan total observasi dikalikan dengan lebar kelas (*the class width*). Untuk kategori normalisasi ini, area atau integral dibawah histogram sama dengan satu. Dilihat dari sudut pandang probabilitas, hasil normalisasi histogram relatif dikaitkan dengan *probability density function;*

sedangkan histogram kumulatif relatif dikaitkan dengan *cumulative distribution function.*

Penjelasan tentang histogram di atas dapat digunakan untuk menjawab beberapa pertanyaan berkut, yaitu:

1. Apa jenis distribusi populasi dari data tersebut?
2. Dimanakah data terpusat?
3. Bagaimanakah persebaran data tersebut?
4. Apakah data tersebut simetris atau memiliki kecondongan?
5. Apakah data tersebut mengandung *outliers*?

Berikut ini adalah beberapa contoh bentuk atau pola data, yaitu

1. *Normal*
2. *Symmetric, Non-Normal, Short-Tailed*
3. *Symmetric, Non-Normal, Long-Tailed*
4. *Symmetric and Bimodal*
5. *Bimodal Mixture of 2 Normals*
6. *Skewed (Non-Symmetric) Right*
7. *Skewed (Non-Symmetric) Left*
8. *Symmetric with Outlier*

3.3 Diagram Dahan Daun (*Stem-and-Leaf Plot*)

Diagram dahan daun (*stem-and-leaf plot*) adalah sajian grafis dari data kuantitatif yang serupa dengan histogram dan sangat berguna dalam menggambarkan bentuk distribusi. Diagram dahan daun ini disebut juga dengan *stemplots.* Diagram dahan daun ini memiliki informasi lebih dibandingkan dengan histogram karena nilai data individu disajikan dalam format tabel untuk data yang memiliki pengaruh lebih besar. Berbeda dengan histogram dimana data disajikan dalam bentuk batang (*bars).* Sederhananya, diagram dahan daun terdiri dari dua kolom yang dipisahkan dengan garis vertikal. Kolom sebelah kiri adalah dahan dan sebelah kanan adalah daun.

Untuk membuat diagram dahan daun, observasi harus disusun dari nilai terkecil ke nilai terbesar. Sebagai contoh adalah set data berikut.

54 56 57 59 63 64 66 68 68 72 72 75 76 81 84 88 106

Selanjutnya, ditentukan data yang mewakili dahan dan data yang mewakili daun. Daun terdiri dari digit terakhir suatu angka dan dahan seluruh digit lainnya. Dalam kasus data sangat besar atau sangat kecil, nilai data dapat disederhanakan dalam satuan tertentu (seperti ratusan) yang digunakan untuk daun. Sedangkan sisanya digunakan sebagai dahan. Dalam contoh berikut, daun mewakili satuan tertentu dan dahan mewakili sisanya (puluhan dan digit yang lebih tinggi). Berikut ini adalah contoh diagram dahan daun.

Contoh 3.1

```
5 | 4 6 7 9
6 | 3 4 6 8 8
7 | 2 2 5 6
8 | 1 4 8
9 |
10 | 6
```

Double stemplots merupakan variasi dari diagram dahan daun. *Double stemplots* dapat dibedakan menjadi *splitting stems* dan *the back-to-back stemplot.*

a. *Splitting Stems*

Untuk suatu set data, pemecahan atau pembagian setiap dahan menjadi bagian-bagian yang lebih kecil, dua atau lima dahan, akan memberikan ilustrasi yang lebih baik dari bentuk distribusi data. Ketika melakukan pemecahan dahan, yang penting dicatat adalah seluruh dahan harus dipecah dan pemecahan tersebut harus seimbang. Sebagai contoh, dahan dipecah menjadi dua dan lima. Ketika dahan dipecah menjadi dua bagian, satu dahan terdiri dari 0-4 daun dan dahan berikutnya terdiri dari 5-9 daun. Sedangkan ketika dahan dipecah menjadi lima bagian, satu dahan terdiri dari 0-1 daun, dahan berikutnya terdiri dari 2-3 daun, dan untuh dahan-dahan selanjutnya terdiri dari 4-5 daun, 6-7 daun, dan 8-9 daun. Berikut ini adalah contoh dari *split stemplot* dimana satu dahan dipecah menjadi dua dahan (untuk contoh ini, data masih mengacu pada data di atas).

Contoh 3.2

```
        5 | 4
        5 | 6 7 9
        6 | 3 4
        6 | 6 8 8
        7 | 2 2
        7 | 5 6
        8 | 1 4
        8 | 8
        9 |
        9 |
       10 |
       10 | 6
```

b. *Back-to-back stemplot*

Back-to-back stemplots digunakan untuk membandingkan dua distribusi *side-by-side*. Jenis *double stemplot* ini terdiri dari tiga kolom, masing-masing dipisahkan oleh garis vertikal. Kolom dibagian tengah merupakan dahan sedangkan kolom pertama dan ketiga terdiri dari daun yang memiliki distribusi yang berbeda. Berikut ini adaah contoh *back-to-back stemplot* yang membandingkan distribusi Set B dan Set A.

<u>Contoh 3.3</u>

```
  Set A          Set B
 -------        -------
   6 5 3 |  4 |
 8 7 6 5 |  5 | 4 6 7 9
   7 3 2 |  6 | 3 4 6 8 8
     4 2 |  7 | 2 2 5 6
       6 |  8 | 1 4 8
         |  9 |
         | 10 | 6
```

3.4 Diagram Kotak Garis (*Box Plot*) dan Pemeriksaan *Outliers* (pencilan)

Salah satu tujuan pembuatan diagram kotak garis (*box plot*) adalah untuk mengecek pergeseran lokasi dan variasi. *Box plots* (Chambers, 1983) adalah alat yang sangat bagus untuk menggambarkan informasi lokasi dan variasi suatu data, khususnya untuk mendeteksi dan mengilustrasikan perubahan lokasi dan variasi antara kelompok data yang berbeda. Contoh berikut menyatakan bahwa mesin memiliki efek signifikan terhadap energi berkaitan dengan lokasi dan variasi yang memungkinkan.

Gambar 4.2 *Box Plot*

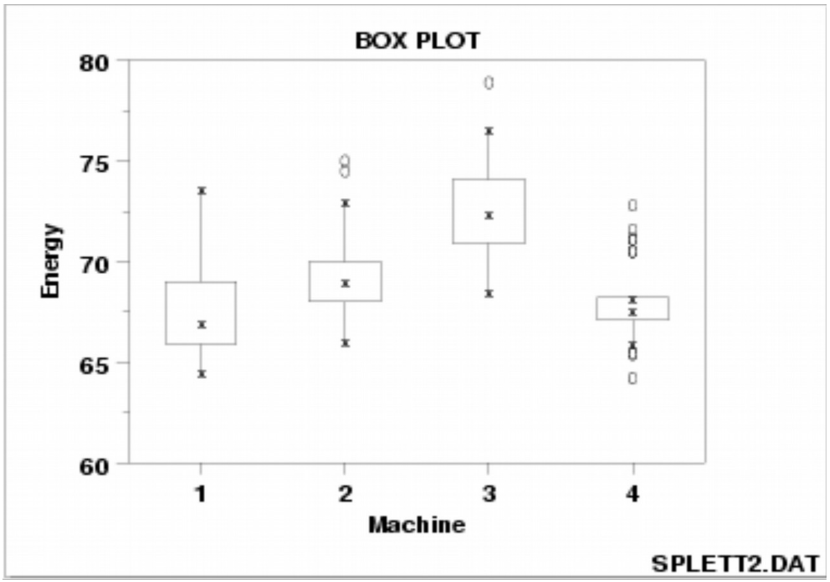


Diagram kotak garis di atas menunjukkan perbandingan empat mesin yang bisa digunakan untuk menghasilkan output. Mesin memiliki efek yang signifikan terhadap energi berkaitan dengan lokasi dan variasi. Mesin tiga memiliki respon energi terbesar (sekitar 72.5); mesin empat memiliki respon energi terkecil.

Diagram kotak garis dapat menjawab beberapa pertanyaan berikut:

1. Apakah faktor tersebut signifikan?
2. Apakah lokasinya berbeda antara dua sub kelompok?
3. Apakah variasinya berbeda antara dua sub kelompok?
4. Apakah ada outlier?

Diagram kotak garis dibentuk oleh sumbu vertikal (variabel respon, sebagai contoh adalah energi) dan horisontal (faktor yang mempengaruhi, sebagai contoh adalah mesin). Lebih spesifik, diagram kotak garis dapat dibuat berdasarkan tahapan berikut:

1. Menghitung median dan kuartil (kuartil terendah adalah 25% persentil dan kuartil tertinggi adalah 75% persentil).
2. Membuat plot simbol pada median (atau menggambar garis) dan menggambar sebuah boks (selanjutnya dikenal dengan istilah *box-plot*) antara kuartil terendah dan tertinggi; boks ini mewakili 50% data pertengahan (*the* "*body*" *of the data).*

3. Membuat garis dari kuartil terendah ke titik minimum dan membuat garis dari kuartil tertinggi ke titik maksimum. Secara spesifik, simbol digambarkan pada titik minimum dan maksimum, meskipun hal ini tidak harus dilakukan.
4. Dari gambar yang dibuat, *box-plot* mampu mengidentifikasi 50% data pertengahan, median dan titik ekstrim.

Diagram kotak-garis tunggal (*a single box plot*) dapat dibuat untuk sejumlah data yang tidak memiliki perbedaan. Sebagai alternatif, diagram kotak-garis ganda (*multiple box plots*) dapat dibuat secara bersama-sama untuk membandingkan beberapa set data atau berbagai kelompok data dalam set data tunggal. Untuk diagram kotak-garis tunggal, lebar boks bisa berubah-ubah. Sedangkan untuk diagram kotak-garis ganda, lebar boks dapat ditetapkan secara proporsional dengan jumlah sampel. Pada umumnya, lebar diagram kotak-garis ditetapkan sama untuk seluruh boks.

Salah satu manfaat dari diagram kotak-garis adalah mengidentifikasi *outliers*. Tahapan yang dilakukan untuk mengidentifikasi *outliers* adalah:

1. Menghitung median dan kuartil terendah dan tertinggi.
2. Membuat plot simbol median dan membuat boks antara kuartil terendah dan tertinggi.
3. Menghitung interval interkuartil (*interquartile range*) yang merupakan perbedaan antara kuartil terendah dan tertinggi, dinotasikan dengan *IQ*.
4. Menghitung beberapa point berikut:

   L_1 = kuartil terendah - *1.5*IQ*

   L_2 = kuartil terendah - *3.0*IQ*

   U_1 = kuartil tertinggi + *1.5*IQ*

   U_2 = kuartil tertinggi + *3.0*IQ*
5. Pada awalnya, garis dibuat dari kuartil terendah ketitik minimum, sekarang garis dibuat dari kuartil terendah ke titik yang paling kecil tetapi lebih besar dibandingkan dengan L_1. Hal yang sama juga berlaku bagi kuartil tertinggi. Pada awalnya garis dibuat dari kuartil tertinggi ke titik maksimum, sekarang garis dibuat dari kuartil tertinggi ke titik yang paling besar tetapi lebih kecil dibandingkan dengan U_1.

6. Titik antara L_1 dan L_2 atau antara U_1 dan U_2 digambarkan dalam lingkaran (*circle)* yang lebih kecil. Sedangkan titik yang lebih kecil dari L_2 atau lebih besar dari U_2 digambarkan dalam lingkaran yang lebih besar.

Diagram kotak-garis merupakan instrumen EDA yang penting untuk menentukan apakah suatu faktor memiliki efek yang signifikan dengan lokasi dan variasi yang berbeda. Selain itu diagram kotak-garis juga efektif untuk meringkas informasi yang sangat besar.

MODUL 4
PROBABILITAS

Pokok Bahasan:

- Pengantar tentang Probabilitas
- Konsep Dasar
- Aturan-Aturan Dasar Probabilitas *(Basic Rules for Probability)*
- Aturan-aturan Perhitungan: Permutasi dan Kombinasi
- Operasi-operasi Probabilitas
- Konsep Lanjutan: *The Law of Total Probability* dan Teori Bayes

4.1 Pengantar tentang Probabilitas

Probabilitas adalah sesuatu yang lazim diperbincangkan dalam kehidupan sehari-hari. Sebagai contoh, ketika dua orang pecandu sepakbola memperdebatkan **peluang** kesebelasan favoritnya untuk menjadi pemenang, **substansi** dari pembicaraan tersebut adalah probabilitas. Mengapa? Karena setiap orang selalu menyatakan berapa peluang kesebelasan-nya untuk menang atau kalah. Tentunya, kedua orang tersebut tidak ada yang mengatakan bahwa kesebelasannya “pasti menang” atau “pasti kalah”. Semua dibicarakan dalam konteks peluang untuk “menang” atau “kalah”.

Singkatnya, probabilitas dinyatakan sebagai “ukuran” atas ketidakpastian (*uncertainty*). Probabilitas bisa digunakan sebagai ukuran “besarnya keyakinan (*strength of belief*)” atas terjadinya suatu peristiwa. Atau juga, probabilitas digunakan sebagai ukuran “tingkat peluang (*degree of chance*) terjadinya sesuatu yang tidak pasti. Probabilitas dinyatakan dalam angka antara “0 dan 1” atau juga “0% dan 100%”.

Apabila membahas aliran probabilitas, ada dua pembedaan atau klasifikasi, yaitu probabilitas objektif (dikenal juga dengan istilah probabilitas klasik) dan probabilitas subjektif. Karakteristik dari kedua klasifikasi probabilitas tersebut dapat dijelaskan sebagai berikut:

1. Probabilitas Objektif atau Klasik *(Objective or Classical Probability)*
   - memiliki peluang yang sama (*based on equally-likely events*)
   - berdasarkan frekuensi yang relatif panjang
   - tidak berdasarkan penilaian seseorang (*personal beliefs*)
   - probabilitas berlaku sama bagi seluruh observasi atau subjek

     Contoh: pelemparan koin, pelemparan dadu, permainan kartu
2. Probabilitas Subjektif *Subjective Probability*
   - berdasarkan penilaian seseorang (*personal beliefs*) atau pengalaman
   - probabilitas berbeda untuk seluruh observasi atau subjek

     Contoh: pemilihan pemimpin, perkiraan cuaca, pel

4.2 Konsep Dasar

Ada beberapa konsep dasar probabilitas yang harus dipahami, yaitu:

- ***Set*** – adalah kumpulan elemen atau objek studi

  Set juga bisa dibedakan menjadi:
  - Set kosong (*empty set*) dinotasikan dengan ∅

    Yaitu set yang tidak memiliki elemen
  - Set universal (*universal set*) dinotasikan dengan S

    Yaitu set yang terdiri dari seluruh elemen yang memungkinkan
  - Komplemen (*Complement)*

    Komplemen A adalah set dari elemen S yang bukan A
- **Perpotongan** (*Intersection*) dinyatakan dengan “dan”

Untuk kasus dua set, A dan B, *intersection* adalah set yang merupakan irisan elemen A dan B

- ***Union*** dinyatakan dengan "atau"

  Yaitu set yang terdiri dari seluruh elemen A atau B atau keduanya

- ***Mutually exclusive or disjoint sets***

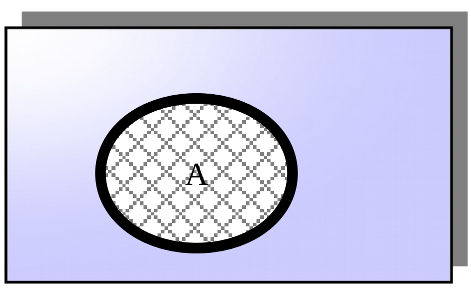
  Yaitu beberapa set yang tidak memiliki irisan. Irisan dari set tersebut berupa himpunan kosong (*empty set*)

- ***Partition***

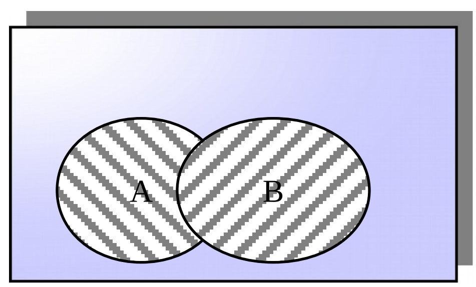
  Yaitu kumpulan set *mutually exclusive* yang elemennya secara bersama-sama membuat set universal

Set, *Intersection* dan *Union* dapat dilihat pada gambar 4.1.a-c di bawah.

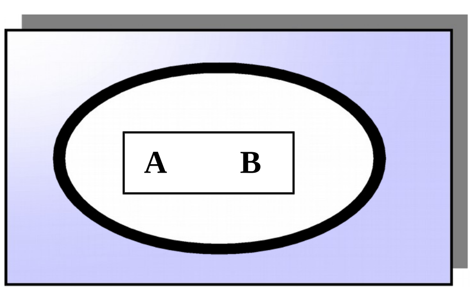
Gambar 4.1.a Set



Gambar 4.1.b *Intersection*



Gambar 4.1.c *Union*



- **Eksperimen (*Experiment*)**

Yaitu suatu proses yang menghasilkan output yang memungkinkan. Sebagai contoh: **output** dari pelemparan koin adalah “muka” dan “ekor”. Setiap eksperimen hanya menghasilkan satu output. Sebagai contoh, **satu kali** pelemparan koin hanya akan menghasilkan **satu output**, yaitu “muka atau ekor”. Output dari eksperimen yang random tidak diketahui sebelum dilakukan percobaan.

- **Kejadian (*Events*)**

  Sering disebut dengan “Ruang sampel” (*Sample Space*) atau “set kejadian” (*Event Set*). Kejadian merupakan set seluruh output yang memungkinkan pada percabaan tertentu.

  Contoh:

  1. Pelemparan dadu

     Kejadian yang memungkinkan adalah S = (1,2,3,4,5,6)

  2. Kumpulan dari angka genap

     Kejadian yang memungkinkan adalah A = (2,4,6)

Probabilitas dari suatu kejadian adalah jumlah dari probabilitas kejadian yang ada. Contoh: probabilitas A merupakan penjumlahan dari beberapa kejadian. Dinotasikan dengan,   P(A) = P(2) + P(4) + P(6).

4.3 Aturan-Aturan Dasar Probabilitas *(Basic Rules for Probability)*

1. Nilai probabilitas ada pada interval

2. Komplemen (*Complements*)

   Prbabilitas yang ***bukan*** A dinyatakan dengan

3. Perpotongan (*Intersection)*

   Untuk set A dan B, probabilitasnya adalah A **dan** B. Dinyatakan dengan

4. *Mutually exclusive events* (A and C) :

Untuk set A dan C, probabilitasnya dari keduanya adalah **nol.** Dinyatakan dengan

Karena **irisan** dari probabilitas A dan C adalah nol, maka *union* dari set A dan C adalah penjumlahan dari probabilitas A dan C.

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

5. *Union*

Untuk set A dan B, probabilitas dari *union* adalah probabilitas A **atau** B atau **keduanya**.

Union dinyatakan dengan

6. Probabilitas Bersyarat (*Conditional Probability*)

Untuk set A dan B, probabilitas bersyarat merupakan probabilitas A **apabila** B **terjadi**. Atau diterjemahkan juga dengan probabilitas A dengan syarat terjadi B. Dinyatakan dengan

Untuk kejadian yang independen, berlaku aturan

<u>Contoh 4.1:</u>

Probabilitas kartu As (*Ace)* apabila terambil kartu Hati (*Heart);* dan sebaliknya, probabilitas kartu Hati (*Heart)* apabila terambil kartu As (*Ace).*

Irisan dari kartu As (*Ace)* dan kartu Hati (*Heart)* dapat dinyatakan dengan

4.4 Aturan-aturan Perhitungan: Permutasi dan Kombinasi

1. Konsep Kombinasi (*Combinatorial Concepts*)

Untuk memudahkan penjelasan konsep kombinasi digunakan ilustrasi sebuah dadu. Dari dadu tersebut, untuk setiap pelemparan, ada enam kemungkinan output, yaitu (1,2,3,4,5,6). Jadi untuk dua dadu yang dilemparkan secara bersama-sama dalam setiap pelemparan, ada 36 kemungkinan output.

Secara umum, jika ada ***n*** **kejadian** dengan kemungkinan **output** sebanyak ***i***, maka kemungkinan output total sebanyak ***ni.*** Sebagai contoh adalah permainan kartu baik **dengan** pengembalian atau **tanpa** pengembalian.

a. **dengan** pengembalian

   Jika diambil 5 buah kartu dari dalam bungkus **dengan** pengembalian, maka total probabilitas output yang memungkinkan adalah 52*52*52*52*52= 380.204.032

b. **tanpa** pengembalian

   Jika diambil 5 buah kartu dari dalam bungkus **tanpa** pengembalian, maka total probabilitas output yang memungkinkan adalah 52*51*50*49*48 = 311.875.200

Jadi dapat dijelaskan bahwa total output yang memungkinkan dari pengambilan lima kartu dari dalam bungkus **dengan** dan **tanpa** pengembalian adalah berbeda.

Contoh lain adalah, kombinasi dari tiga huruf yang diinginkan (misal: B, C, D) dari total enam huruf (A, B, C, D, E, dan F) adalah BCD, BDC, CBD, CDB, DBC, dan DCB. Atau dengan kata lain, membuat susunan dari tiga huruf yang sama.

Singkatnya, kombinasi adalah pilihan yang memungkinkan dari ***r*** objek atau unit **dari total *n* objek** atau group *n* **sesuai** urutan yang diinginkan. Jumlah kombinasi dinotasikan dengan *n* kombinasi *r*, atau ***nCr.***

Rumus kombinasi *r* dari total *n* dinyatakan dengan

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

Sebagai contoh:

$$\binom{n}{r} = {}_6C_3 = \frac{6!}{3!(6-3)!} = \frac{6!}{3!3!}$$

$$= \frac{6*5*4*3*2*1}{(3*2*1)(3*2*1)} = \frac{6*5*4}{3*2*1} = \frac{120}{6} = 20$$

2. Konsep Faktorial (*Factorial*)

Pembahasan konsep faktorial dapat dimulai dengan pertanyaan berikut. Berapa banyak kita dapat menyusun tiga huruf, A, B, dan C?

Susunan dari ketiga huruf di atas dapat dinyatakan dengan: 3 pilihan untuk huruf pertama (misal: A), 2 pilihan untuk huruf kedua (misal: B), dan 1 pilihan untuk huruf ketiga (misal: C). Jadi, susunan yang mungkin dibuat dari huruf-huruf di atas berjumlah 3*2*1 = 6.

Demikian juga susunan huruf berjumlah enam, yaitu: A, B, C, D, E dan F. Total susunan yang bisa kita buat dari enam huruf tersebut adalah (6*5*4*3*2*1 = 720).

Secara definisi, n faktorial untuk bilangan positif integer dinyatakan dengan

$$n(n-1)(n-2)\ldots(1)$$

Notasi untuk *n* faktorial adalah ***n!***. Angka *n!* adalah angka dimana *n* objek dapat ditentukan (disusun). Sebagai contoh, berdasarkan definisi tersebut, *1! = 1.*

3. Konsep Permutasi (*Permutations*)

Untuk menjelaskan konsep permutasi, dapat digunakan ilustrasi berikut. Jika kita memiliki enam huruf (A, B, C, D, E, dan F), dan ***hanya*** tiga huruf yang diinginkan (tidak harus spesifik B, C, D sebagaimana contoh kombinasi), cara yang bisa dilakukan untuk mendapat hasil tersebut adalah ada 6 cara untuk memilih huruf pertama, ada 5 cara untuk memilih huruf kedua, dan ada 4 cara untuk memilih huruf ketiga. Total susunan yang bisa dibuat sebanyak 6*5*4=120. Ini yang dinamakan permutasi.

Singkatnya, definisi permutasi adalah pemilihan susunan (*ordered)* yang memungkinkan atas *r* objek dari total *n* objek. Berdasarkan contoh di atas, *n* objek dinyatakan sebagai banyaknya huruf, yaitu 6; dan *r* objek dinyatakan sebagai banyaknya objek yang diinginkan, yaitu 3. Jumlah permutasi atas *r* objek dari total *n* objek dinotasikan dengan ***nPr.***

Rumus dari permutasi dinyatakan dengan

$$_6P_3 = \frac{6!}{(6-3)!}$$

Sebagai contoh:

$$_6P_3 = \frac{6!}{(6-3)!} = \frac{6!}{3!} = \frac{6*5*4*3*2*1}{3*2*1} = 6*5*4 = 120$$

4.5 Operasi-operasi Probabilitas

Sebelum masuk ke dalam operasi probabilitas, ada kata-kata kunci yang biasa digunakan dalam probabilitas, yaitu: (Contoh kasus set E dan F)

- E ***dan*** F : Peristiwa E maupun F sama-sama terjadi
- E ***atau*** F : Salah satu dari peristiwa E *atau* F terjadi; *atau* bisa juga keduanya terjadi
- ***Tidak*** E : Peristiwa E tidak terjadi

Rumus operasi probabilitas dapat dirangkum sebagai berikut:

1. Aturan Penjumlahan

P(E ATAU F) = P(E) + P(F) - P(E DAN F)

Apabila set E dan F *saling meniadakan atau independen*, maka aturan penjumlahan adalah

P(E ATAU F) = P(E) + P(F)

2. Aturan Pengurangan

P(E) = 1- P(TIDAK E)

3. Aturan Perkalian

P(E DAN F) = P(E/F) P(F)

Apabila set E dan F *saling meniadakan atau independen*, maka aturan perkalian adalah

P(E DAN F) = P(E) P(F)

4.6 Konsep Lanjutan: *The Law of Total Probability* dan Teori Bayes

Konsep lanjutan dari probabilitas adalah Hukum Probabilitas Total *(The Law of Total Probability)* dan Teori Bayes (*Bayes' Theorem*).

4.6.1 Hukum Probabilitas Total *(The Law of Total Probability)*

Hukum probabilitas total dirumuskan sebagai berikut:

Dalam konteks probabilitas bersyarat, berlaku:

Lebih umumnya, jika B_i merupakan suatu proporsi, berlaku:

Untuk menjelaskan rumus di atas, digunakan contoh **pasar saham dan kinerja ekonomi tahun depan.** Ada dua kondisi yang dapat dinyatakan dari keadaan pasar saham dan kinerja ekonomi. Pasar saham bisa tumbuh positif (kejadian *U*) atau negatif (kejadian bukan *U*); dan kinerja ekonomi bisa mengalami peningkatan (kejadian *W)* atau penurunan (kejadian bukan *W)* tahun depan. Kejadian *bukan* U dinotasikan dengan $\bar{U}$ dan Kejadian *bukan* W dinotasikan dengan $\bar{W}$. Atau lebih singkatnya, pernyataan di atas dapat dinyatakan dengan kejadian *U* dan *W* berikut.

Kejadian *U*: pasar saham akan tumbuh positif tahun depan
Kejadian *W*: kinerja ekonomi akan mengalami peningkatan tahun depan

<u>Diketahui:</u>

Probabilitas pasar saham akan tumbuh positif jika kinerja ekonomi mengalami peningkatan tahun depan adalah 0.75. Sedangkan probabilitas kinerja ekonomi akan mengalami peningkatan tahun depan adalah 0.8. Jadi, besarnya probabilitas pasar saham akan tumbuh positif tahun depan adalah:

<u>Jawab:</u>

4.6.2 Teori Bayes (*Bayes' Theorem)*

Dengan mengetahui probabilitas *A given B,* melalui teori Bayes, dapat diperoleh probabilitas *B given A.* Berdasarkan definisi probabilitas bersyarat dan Hukum Probabilitas Total dapat dirumuskan:

Sebagai contoh adalah uji medis HIV, dinotasikan dengan *I,* yang mempengaruhi 0.1% populasi adalah tidak sempurna. Uji medis tersebut dilakukan kepada dua kategori, yaitu seseorang yang dinyatakan sakit (Z) dan dinyatakan tidak sakit (bukan Z).

Dinotasikan, probabilitas dari uji tersebut adalah $P(I) = 0.001$.

- Ketika diujikan pada seseorang yang **dinyatakan sakit**, uji medis tersebut menunjukkan probabilitas sebesar 0.92. Dinotasikan dengan

$$P(Z|I) = 0.92 \Rightarrow P(\bar{Z}|I) = 0.08$$

Kejadian $P(Z|I)$ adalah negatif tidak benar

- Ketika diujikan pada seseorang yang **dinyatakan tidak sakit**, uji medis tersebut menunjukkan probabilitas sebesar 0.04. Dinotasikan dengan

$$P(Z|\bar{I}) = 0.04 \Rightarrow P(\bar{Z}|\bar{I}) = 0.96$$

Kejadian $P(\bar{Z}|\bar{I})$ adalah positif tidak benar

Contoh di atas dapat diselesaikan sebagai berikut:

Jadi probabilitas hasil uji medis yang mempengaruhi 0.1% populasi dengan kategori orang yang dinyatakan tidak sakit adalah 0.0225.

4.7 Latihan: Soal & Jawab

1. Ada sembilan anak SD yang terdiri dari 5 pria dan 4 wanita yang akan diurutkan dalam satu baris. Berapa banyak cara yang bisa dilakukan jika Guru mereka meminta wanita menempati urutan bernomor genap.

   Jawab:

   $${}_5P_5 \cdot {}_4P_4 = 5!4! = (120)(24) = 2880$$

   Jadi untuk mengurutkan sembilan anak tersebut, ada ${}_5P_5$ cara untuk mengatur anak pria dan ${}_4P_4$ cara untuk mengatur anak wanita.

2. Ada berapa cara 5 buah bendera yang berbeda warna dapat dideretkan dalam satu baris.

   Jawab:

   $$\begin{aligned} {}_nP_n &= n! \\ &= n(n-1)(n-2)\ldots1 \\ &= 5\cdot4\cdot3\cdot2\cdot1 = 120 \end{aligned}$$

   Untuk mengurutkan bendera tersebut, ada 5 cara untuk mengisi urutan pertama, ada 4 cara untuk mengisi urutan kedua, ada 3 cara untuk mengisi urutan ketiga, ada 2 cara untuk mengisi urutan kedua dan ada 1 cara untuk mengisi urutan terakhir.

3. Di sebuah meja makan, hanya tersedia empat kursi. Ada berapa cara yang bisa dilakukan jika ada 10 orang yang hendak menempati meja tersebut.

   Jawab:

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

   Jika $n = 10$ dan $r = 4$ maka

   $${}_{10}P_{4} = \frac{10!}{(10-4)!}$$

   $$= 10 \cdot 9 \cdot 8 \cdot 7 = 5040$$

   Untuk menempati meja makan tersebut, ada 10 cara untuk mengisi tempat duduk pertama, ada 9 cara untuk mengisi tempat duduk kedua, ada 8 cara untuk mengisi tempat duduk ketiga, dan ada 7 cara untuk mengisi tempat duduk keempat.

4. Ada berapa cara untuk menyusun angka yang terdiri dari 4 digit dari 10 angka berikut [0, 1, 2, 3, ..., 9]; jika pengulangan diperkenankan.

   Jawab:

   Digit pertama dapat diisi dengan salah satu dari 9 angka yang tersedia. Mengapa hanya 9 angka untuk digit pertama? Karena angka 0 tidak boleh menempati digit pertama. Selanjutnya, digit kedua sampai keempat bisa diisi dengan 10 angka di atas. Jadi jumlah angka 4 digit yang dapat dibentuk adalah

   $$= 9 \cdot 10 \cdot 10 \cdot 10 = 9000$$

5. Berdasarkan soal (4), jika pengulangan tidak diperkenankan.

   Jawab:

   Karena pengulangan tidak diperkenankan, maka penyusunan angka 4 digit memiliki cara sebagai berikut. Di sini, seperti penjelasan di atas, angka nol tidak bisa digunakan pada digit pertama. Sedangkan pada digit kedua, angka nol sudah bisa digunakan. Pada digit ke-4, tinggal ada 7 cara, karena sudah terambil tiga angka untuk menempati digit pertama-ketiga.

   $$= 9 \cdot 9 \cdot 8 \cdot 7 = 4536$$

6. Berdasarkan soal (4), jika pengulangan tidak diperkenankan dan digit terakhir harus angka nol.

   Jawab:

   Digit pertama dapat dipilih dengan 9 cara, digit kedua dapat dipilih dengan 8 cara dan digit ketiga dapat dipilih dengan 7 cara. Atau

$$= 9 \cdot 8 \cdot 7 = 504$$

7. Dari berbagai penerbit, dikumpulkan empat buku matematika yang berbeda, 6 buku fisika yang berbeda, dan 2 buku kimia yang berbeda. Ada berapa cara penyusunan yang mungkin jika masing-masing kelompok (buku) harus disusun bersama.

   Jawab:

   Buku Matematika: ${}_4P_4 = 4!$

   Buku Fisika: ${}_6P_6 = 6!$

   Buku Kimia: ${}_2P_2 = 2!$

   Dari 3 kelompok ilmu: ${}_3P_3 = 3!$

   Jadi jumlah penyusunan yang mungkin adalah: $4!6!2!3! = 207360$ cara

8. Berdasarkan soal (7), jika hanya buku matematika yang disatukan.

   Jawab:

   Diasumsikan, buku matematika digabung jadi satu. Jadi empat buku matematika yang digabung, dihitung sebagai satu buku matematika.

   Sehingga cara penyusunan buku ${}_9P_9 = 9!$

   Untuk buku Matematika sendiri: ${}_4P_4 = 4!$

   Buku Fisika: ${}_6P_6 = 6!$

   Jadi jumlah penyusunan yang mungkin adalah: $9!4! = 8709120$ cara

9. Amir memiliki 6 buku matematika dan 4 buku fisika. Hitunglah probabilitas bahwa 3 buku matematika tertentu akan tersusun bersama-sama.

   Jawab:

Seluruh buku dapat disusun dengan ${}_{10}P_{10} = 10!$

Diasumsikan ke-3 buku Matematika diperlakukan sebagai 1 buku.

Sehingga diperoleh 8 buku yang dapat disusun dengan ${}_{8}P_{8} = 8!$ cara

Untuk buku Matematika sendiri: ${}_{3}P_{3} = 3!$ cara

Jadi probabilitas bahwa 3 buku matematika tertentu akan tersusun secara bersama-sama adalah $\frac{8!3!}{10!} = \frac{6}{90} = \frac{1}{15}$

10. Ada lima kotak berwarna merah, 2 kotak berwarna putih, dan 3 kotak berwarna biru disusun dalam satu baris. Jika seluruh kotak yang memiliki warna sama tidak dibedakan satu sama lain, ada berapa cara penyusunan yang memungkinkan.

Jawab:

$${}_{n}P_{(n_1 \times n_2 \times n_3)} = \frac{n!}{n_1!n_2!n_3!}$$

Dimana $n = n_1 + n_2 + n_3$

$n =$ jumlah seluruh kotak

$n_1 =$ kotak merah

$n_2 =$ kotak merah

$n_3 =$ kotak merah

maka ${}_{10}P_{(5\times 2\times 3)} = \frac{10!}{5!2!3!} = 2520$

11. Seorang ketua panitia lomba memasak harus memilih 5 dari 9 siswa yang tersedia. Berapa banyak cara yang tersedia.

Jawab:

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

$${}_{9}C_{5} = \frac{9!}{5!(9-5)!} = \frac{9!}{5!4!}$$

$$= \frac{9\cdot 8\cdot 7\cdot 6\cdot 5}{5\cdot 4\cdot 3\cdot 2\cdot 1} = \frac{15(120)}{120} = 126$$

12. Anak TK diberikan berbagai warna untuk membuat lukisan. Ada 6 warna dasar yang dapat digunakan, yaitu warna emas, biru, merah, putih, kuning dan hijau. Setiap komposisi dapat menggunakan 4 warna. Apabila pengulangan warna tidak diperkenankan (seperti: biru, biru, merah, biru), berapa kemungkinan komposisi yang dapat dipakai.

    <u>Jawab:</u>

    $${}_{n}P_{r} = \frac{n!}{(n-r)!} = \frac{6!}{(6-4)!} = 360$$

    $\text{dimana } n = 6 \text{ dan } r = 4$

    Singkatnya, urutan pertama dari komposisi warna dapat digunakan salah satu dari 6 warna yang tersedia (jadi ada 6 cara). Urutan kedua dapat digunakan 5 warna yang tersedia. Demikian juga untuk urutan ketiga dan keempat, masing-masing ada 4 dan 3 cara.

13. Berdasarkan soal (12), apabila pengulangan warna diperkenankan, berapa kemungkinan komposisi yang dapat dipakai.

    <u>Jawab:</u>

    $$P = n^{r} = 6^{4} = 1296$$

    $\text{dimana } n = 6 \text{ dan } r = 4$

    Jika pengulangan warna diperbolehkan, ada 6 cara untuk membuat komposisi, mulai dari urutan pertama sampai dengan keempat.

14. Ada berapa banyak cara 10 mainan dapat dikelompokkan dalam 2 group jika setiap group terdiri dari 4 dan 6 mainan.

    <u>Jawab:</u>

    $${}_{10}C_{4} = \frac{10!}{4!(10-4)!} = \frac{10 \cdot 9 \cdot 8 \cdot 7}{4 \cdot 3 \cdot 2 \cdot 1} = \frac{5040}{24} = 210$$

15. Seorang pemerhati hidangan ingin membuat kombinasi dari berbagai jenis lalapan yang tersedia, yaitu: slada, kubis, ketimun, tomat dan kacang panjang. Berapa banyak kombinasi lalapan yang memungkinkan.

    <u>Jawab:</u>

    $${}_{5}C_{1} + {}_{5}C_{2} + {}_{5}C_{3} + {}_{5}C_{4} + {}_{5}C_{5}$$
    $$= 5 + 10 + 10 + 5 + 1 = 31$$

16. Sebuah perusahaan konsultan memiliki 6 ahli matematika dan 7 ahli fisika. Dari komposisi ahli yang dimiliki, perusahaan tersebut ingin membuat tim yang terdiri dari 2 ahli matematika dan 3 ahli fisika. Ada berapa cara yang bisa dilakukan jika tim tersebut **harus** mencakup 1 ahli matematika dan fisika.

    Jawab:

    $${}_5C_2 \cdot {}_7C_3 = \left[\frac{5!}{2!(5-2)!}\right]\left[\frac{7!}{3!(7-3)!}\right]$$

    $$= \left(\frac{5\cdot 4}{2\cdot 1}\right)\left(\frac{7\cdot 6\cdot 5}{3\cdot 2\cdot 1}\right) = (10)(35) = 350$$

17. Berdasarkan soal (16), jika tim tersebut **harus** mencakup 1 ahli fisika.

    Jawab:

    $${}_5C_2 \cdot {}_6C_2 = \left[\frac{5!}{2!(5-2)!}\right]\left[\frac{6!}{2!(6-2)!}\right]$$

    $$= \left(\frac{5\cdot 4}{2\cdot 1}\right)\left(\frac{6\cdot 5}{2\cdot 1}\right) = (10)(15) = 150$$

18. Berdasarkan soal (16), jika 2 ahli matematika tidak bisa masuk dalam tim tersebut.

    Jawab:

    $${}_3C_2 \cdot {}_7C_3 = \left[\frac{3!}{2!(3-2)!}\right]\left[\frac{7!}{3!(7-3)!}\right]$$

    $$= \left(\frac{3\cdot 2}{2\cdot 1}\right)\left(\frac{7\cdot 6\cdot 5}{3\cdot 2\cdot 1}\right) = (3)(35) = 105$$

19. Seorang sekretaris ingin mengatur tempat duduk dari 7 peserta rapat. Ada berapa banyak cara yang dapat digunakan jika peserta rapat bisa duduk dimana saja.

    Jawab:

    Singkatnya, ada satu orang yang bisa duduk dimana saja. Masih ada 6 kursi yang tersisa untuk peserta rapat lainnya. Jadi ada 6! =720 cara untuk menempatkan 7 peserta di sekeliling meja.

20. Berdasarkan soal (19), jika ada dua peserta tertentu tidak diperkenankan duduk bersebelahan.

    Jawab:

Singkatnya, ada dua orang yang diperlakukan sebagai satu orang. Masih ada 6 orang yang tersisa yang dapat diatur dengan 5! cara. Selanjutnya, diantara 2 orang yang digabung di atas, ada 2! cara. Jadi total cara untuk mengatur peserta rapat adalah 5! 2! =240 cara di sekeliling meja.

Jadi untuk mengatur duduk peserta rapat agar ada dua orang tertentu yang tidak bisa duduk bersebelahan adalah 720-240=480

MODUL 5
PROBABILITAS (LANJUTAN)

Pokok Bahasan:

- Pengertian tentang *joint, marginal* dan*, conditional probability*
- *Expected Value, Tree Diagram,* dan *Conditional Pay-off table*

5.1 Pengertian tentang *joint, marginal* dan *conditional probability*

5.1.1.Joint dan *Marginal probability*

Mereview beberapa konsep dasar probabilitas pada pembahasan sebelumnya, dalam konteks probabilitas *bivariate,* probabilitas irisan (*intersection probabilities)* dikenal dengan *joint probabilities.* Untuk set A dan B, ***joint probabilities*** (irisan) dinotasikan dengan $P(A \cap B)$. Sedangkan probabilitas untuk kejadian masing-masing (individu), yaitu probabilitas kejadian A dan B, dinotasikan dengan, *P(A)* dan *P(B),* disebut dengan ***marginal probabilities*.**

Marginal probabilities diperoleh dengan menjumlahkan *joint probabilities* yang memiliki karakteristik *mutually exclusive. Marginal probabilities* dirumuskan dengan

$$P(Ai) = P(A_i \cap B_1) + P(A_i \cap B_2) + ... + P(A_i \cap B_k)$$

Sebagai contoh, ada tiga kategori restoran (premium, moderat dan murah) dan tiga kategori pendapatan (tinggi, sedang dan rendah). Masing-masing kategori memiliki probabilitas sebagaimana tabel 5.1.

Tabel 5.1 Probabilitas Restoran dan Tingkat Pendapatan

Frekuensi	Pendapatan Tinggi (B_1)	Pendapatan Sedang (B_2)	Pendapatan Rendah (B_3)	Total
Premium (A_1)	0.04	0.13	0.04	0.21
Moderat (A_2)	0.10	0.11	0.06	0.27
Murah (A_3)	0.13	0.17	0.22	0.52
Total	0.27	0.41	0.32	1.00

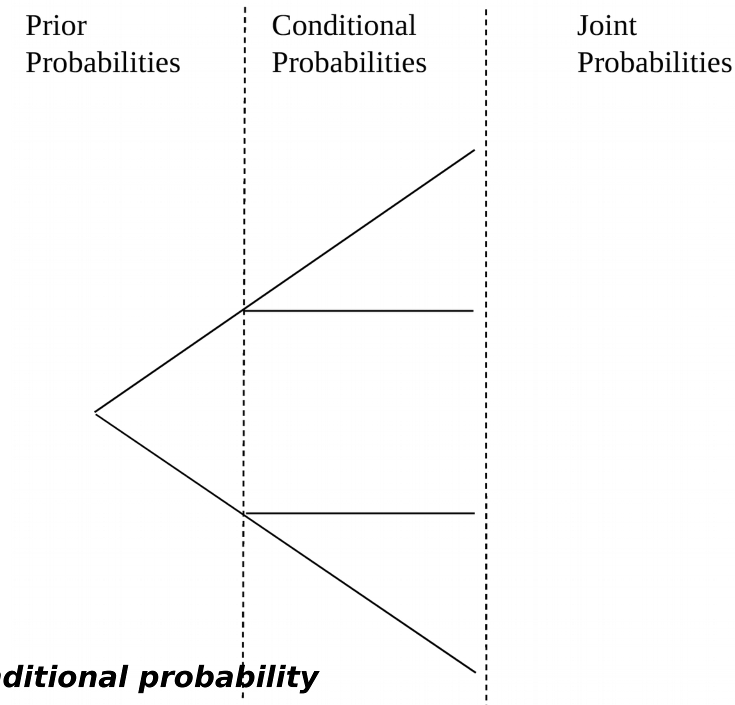
Berdasarkan tabel 5.1 di atas, dapat dicari probabilitas dari rumah tangga yang gemar makan di restoran dengan tingkat harga moderat, *$P(A_2)$.*

$$\begin{aligned} & P(A_2) \\ & = P(A_2 \cap B_1) + P(A_2 \cap B_2) + P(A_2 \cap B_3) \\ & = 0.10 + 0.11 + 0.06 \\ & = 0.27 \end{aligned}$$

Probabilitas marginal dapat diperoleh dengan menjumlahkan **nilai baris** dari setiap kategori. Jadi probabilitas dari rumah tangga yang gemar makan di restoran dengan tingkat harga tinggi [$P(A_1)$] dan murah [$P(A_3)$] adalah

$P(A_1)$= 0.21 dan $P(A_3)$= 0.52

Selain contoh di atas, juga bisa diambil contoh *conditional* dan *joint probabilities* dengan menggunakan diagram pohon. Di sini, untuk menentukan probabilitas, pada umumnya digunakan informasi awal dari probabilitas yang sudah ada sebelumnya (*prior probabilities).* Sekali lagi, digunakan contoh uji medis HIV (dinotasikan dengan *I)* seperti pada bab sebelumnya. Uji medis HIV dinyatakan mempengaruhi 0.1% populasi adalah tidak sempurna. Uji medis tersebut dilakukan kepada dua kategori, yaitu seseorang yang dinyatakan sakit (Z) dan dinyatakan tidak sakit (bukan Z).



5.1.2 *Conditional probability*

Probabilitas bersyarat (*Conditional Probability*) dari *A given B* dinyatakan dengan

Untuk **kejadian yang independen** (*independent events)* berlaku

Aturan pada probabilitas bersyarat adalah:

Rumus di atas dapat dinyatakan juga dengan:

Untuk kejadian A dan D dengan karakteristik ***statistically independent*** dapat dijelaskan sebagai berikut. Probabilitas dari kejadian A jika terjadi kejadian D adalah probabilitas kejadian A itu sendiri. Demikian juga probabilitas dari kejadian D jika terjadi kejadian A adalah probabilitas kejadian D itu sendiri. Atau dapat dinotasikan dengan

$\rightarrow$

Jadi persyaratan *statistically independent* untuk kejadian A dan B adalah

$$P(A|B) = P(A)$$
$$P(B|A) = P(B)$$

Sehingga

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

Contoh kejadian ***statistically independent*** adalah pengambilan kartu As dan kartu hati berikut. Berapa probabilitas kartu As (*Ace*) jika terambil kartu hati (*Heart*); dan sebaliknya, probabilitas kartu Hati (*Heart*) jika terambil kartu As (*Ace*) adalah sebagai berikut:

Jadi, probabilitas dari perpotongan kartu As (*Ace*) dan kartu hati (*Heart*) adalah

5.2 *Expected Value, Tree Diagram,* dan *Conditional Pay-off table*

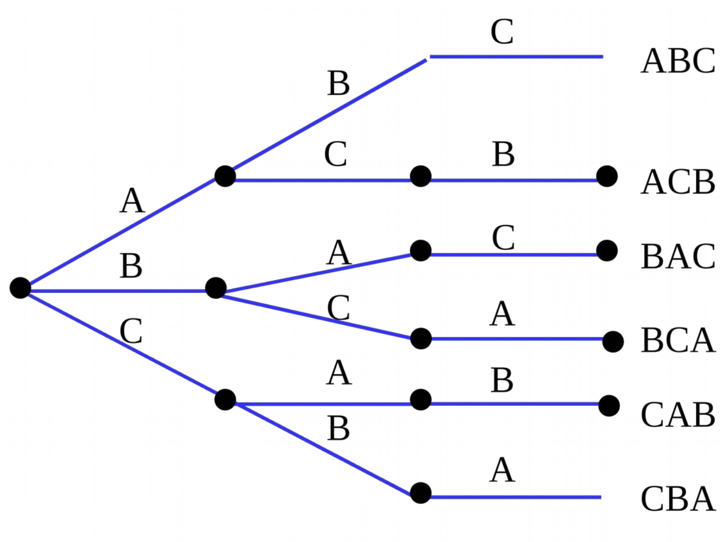
5.2.1 ***Expected Value***

Seperti yang kita pahami, informasi awal (sampel) sangat penting dalam proses pengambilan keputusan. Informasi tersebut dapat digunakan sebagai dasar dalam pengambilan keputusan. Seberapa penting suatu informasi tergantung dari relevansinya menerjemahkan permasalahan yang ada. Untuk mendapatkan informasi awal, perusahaan mengeluarkan dana yang cukup besar sehingga perusahaan tersebut harus mempertimbangkan apakah manfaat yang diperoleh lebih besar dibandingkan dengan biaya yang harus dikeluarkan.

5.2.2 Diagram Pohon (*Tree Diagram)*

Diagram Pohon (*Tree Diagram*) merupakan pembahasan lanjutan dari konsep kombinasi. Untuk diagram pohon, dapat dijelaskan melalui contoh sebagai berikut.

Berapa banyak susunan yang bisa dibuat dari tiga huruf (A, B dan C) berikut.



Jadi, hasil susunan dari tiga huruf di atas adalah ABC, ACB, BAC, BCA, CAB, dan CBA.

5.2.3 *Conditional Pay-off table*

Untuk menjelaskan *Conditional Pay-Off Table,* dapat digunakan contoh sebagai berikut.

Ada dua perusahaan besar yang bergerak di bidang telekomunikasi dan komputer, yaitu AT&T dan IBM. Tabel pertama menunjukkan jumlah pangsa pasar dari masing-masing perusahaan. Sedangkan tabel kedua menunjukkan proporsinya. Berapakah probabilitas dari proyek yang diambil IBM *given* proyek telekomunikasi?

Jumlah Pangsa Pasar Perusahaan

Jumlah	AT&T	IBM	Total
Telekom.	40	10	50
Komputer	20	30	50
Total	60	40	100

Proporsi Pangsa Pasar Perusahaan

Prob.	AT&T	IBM	Total
Telekom.	0.4	0.1	0.5
Komputer	0.2	0.3	0.5
Total	0.6	0.4	1.0

Probabilitas proyek yang diambil oleh IBM *given* proyek telekomunikasi adalah:

$$P(IBM \mid T) = \frac{P(IBM \cap T)}{P(T)}$$

$$= \frac{0.1}{0.5} = 0.2$$

Teori Bayes: Lanjutan

Dengan adanya *partisi* kejadian, misal: B_1, B_2, ..., Bn, dapat dirumuskan probabilitas masing-masing partisi (misal: B1) jika kejadian A terjadi.

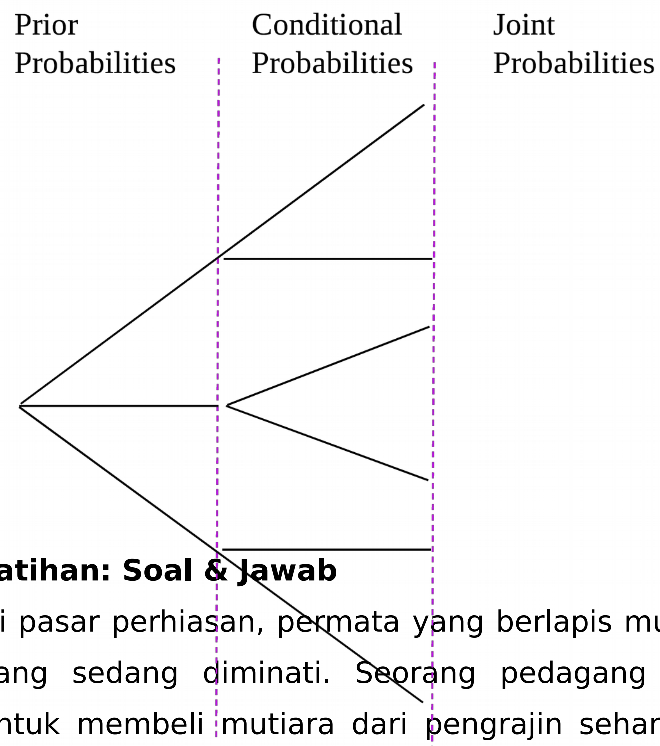
Soal 1:

- Seorang ekonom memprediksikan ketika pertumbuhan ekonomi tinggi (*high growth (P(H))*, US$ apresiasi dengan probabilitas 0.7; ketika pertumbuhan ekonomi moderat *(moderate growth (P(M))*), US$ apresiasi dengan probabilitas 0.4; dan ketika pertumbuhan ekonomi rendah *(low growth P(L)*), US$ apresiasi dengan probabilitas 0.2.
- Pada kurun waktu tersebut, probabilitas pertumbuhan ekonomi tinggi sebesar 0.3; moderat sebesar 0.5; dan rendah 0.2.
- Dengan asumsi bahwa dolar akan apresiasi, berapakah probabilitas terjadinya pertumbuhan ekonomi tinggi?

Jawab:

Partisi	*Event* A - *Appreciation*
H - *High growth* P(H) = 0.30	P (A/H) = 0.70
M - *Moderate growth* P(M) = 0.50	P (A/M) = 0.40
L - *Low growth* P(L) = 0.20	P (A/L) = 0.20

Apabila digunakan diagram pohon, soal di atas dapat diselesaikan sebagai berikut:



5.3 Latihan: Soal & Jawab

1. Di pasar perhiasan, permata yang berlapis mutiara merupakan produk yang sedang diminati. Seorang pedagang perhiasan memutuskan untuk membeli mutiara dari pengrajin seharga US$ 20 ribu. Penjual dapat menjual permata tersebut seharga US$ 24 ribu, US$ 22 ribu, US$ 20 ribu dan US$ 18 ribu dengan probabilitas masing-masing sebesar 0.22, 0.47, 0.26, 0.05. Berapa besarnya laba bruto yang diharapkan.

   Jawab:

$$E(X) = [24000(0.22) + 22000(0.47) + 20000(0.26)$$
$$+18000(0.05)] - 20000$$
$$= (5280 + 10340 + 5200 + 900) - 20000$$
$$= 21720 - 20000 = 1720$$

2. Pemilik perkebunan memiliki 3 area perkebunan untuk ditanami bibit unggul strawbery yang dibelinya. Jika dipilih area 1, peluang bibit unggul tumbuh subur adalah 0.6. Dari kebun tersebut, jika bibit tumbuh subur, potensi profitnya sebesar Rp 300 ribu dan jika gagal, kerugian yang ditanggung Rp 50 ribu. Jika dipilih area 2, peluang bibit unggul tumbuh subur adalah 0.5. Dari kebun tersebut, jika bibit tumbuh subur, potensi profitnya sebesar Rp 400 ribu dan jika gagal,

kerugian yang ditanggung Rp 100 ribu. Jika dipilih area 3, peluang bibit unggul tumbuh subur adalah 0.55. Dari kebun tersebut, jika bibit tumbuh subur, potensi profitnya sebesar Rp 350 ribu dan jika gagal, kerugian yang ditanggung Rp 60 ribu. Berdasarkan harapan profit tertinggi, kebun mana yang akan dipilih.

Jawab:

$$E(I) = 300000(0.6) + 50000(0.4)$$
$$= 180000 - 20000 = 160000$$
$$E(II) = 400000(0.5) + 100000(0.5)$$
$$E(III) = 350000(0.55) = 60000(0.45)$$
$$= 192500 - 27000 = 165500$$

Jadi pemilik kebun bisa memilih area perkebunan III karena menghasilkan potensi keuntungan terbesar dibandIngkan dengan dua kebun lainnya.

3. Berdasarkan soal (2), jika peluang area 3 berubah menjadi 0.5, apakah petani akan merubah keputusan.

   Jawab:

   $$E(III) = 350000(0.5) + 60000(0.5)$$
   $$= 175000 - 30000 = 145000$$

   Berdasarkan hasil perhitungan di atas, tentunya, petani akan merubah keputusannya, yaitu dari kebun III menjadi kebun I; karena potensi keuntungan kebun I menjadi lebih tinggi.

4. Perusahaan cola “Sejuk” memproduksi botol untuk kemasan produk mereka. Satu mesin menghasilkan 12000 botol/hari dan banyaknya botol yang rusak rata-rata sebesar 3%. Dari 600 botol yang dipilih secara random, berapa probabilitas 12 diantaranya adalah rusak.

   Jawab:

   Dari 12000 botol, 3% rusak atau 360 botol rusak. Selebihnya, 11640 botol tidak rusak. Jadi probabilitas dari pengambilan 600 botol secara random untuk mendapat 12 botol rusak adalah $\frac{({}_{360}C_{12})({}_{11640}C_{588})}{{}_{12000}C_{600}}$

5. Seorang penjudi sedang menghitung untung rugi dari uang yang harus dikeluarkan dalam satu kali permainan. Dalam permainan tersebut kemungkinan menang Rp 500 rb dan Rp 200 rb dengan probabilitas

masing-masing sebesar 0.3 dan 0.4. Berapa uang yang pantas dikeluarkan untuk satu kali main.

Jawab:

$$E(X) = \sum_{i=1}^{n} X_i \cdot P(X_i)$$

$$X_1 = 500000 \qquad P(X_1) = 0.3$$

$$X_2 = 200000 \qquad P(X_2) = 0.4$$

$$E(X) = 500000(0.3) + 200.000(0.4)$$

$$= 150000 + 80000$$

$$= 230000$$

6. Ada 3 tempat magang yang disediakan bagi 10 anak yang sudah hampir menyelesaikan kuliahnya. Masing-masing dapat menampung 5, 2, dan 3 anak. Berapa macam susunan penempatan yang dapat dibentuk.

   Jawab:

$${}_{n}C_{(r_1 \times r_2 \times r_3)} = \frac{n!}{r_1!\,r_2!\,r_3!}$$

$${}_{10}C_{(5\times 2\times 3)} = \frac{10!}{5!2!3!} = 2520$$

7. Seorang guru statistik sedang mengajarkan topik probabilitas melalui permainan kartu. Dalam permainan kartu tersebut, diambil 5 kartu dari setumpuk kartu (berjumah 52). Berapakah probabilitas bahwa kartu yang terambil adalah 4 kartu As.

   Jawab:

$$\text{P(4 kartu As)} = \frac{({}_{4}C_{4})({}_{48}C_{1})}{{}_{52}C_{5}}$$

$$= \frac{\left(\dfrac{4!}{4!(4-4)!}\right)\left(\dfrac{48!}{1!(48-1)!}\right)}{\dfrac{52!}{5!(52-5)!}}$$

$$= \frac{48}{2598960} = \frac{1}{54145}$$

8. Berdasarkan soal (7), berapakah probabilitas bahwa kartu yang terambil adalah 4 kartu As dan 1 kartu *King*.

   Jawab:

$$\text{P(4 kartu As dan 1 kartu King)} = \frac{({}_4C_4)({}_4C_1)}{{}_{52}C_5}$$

$$= \frac{\left(\frac{4!}{4!(4-4)!}\right)\left(\frac{4!}{1!(4-1)!}\right)}{\frac{52!}{5!(52-5)!}}$$

$$= \frac{4}{2598960} = \frac{1}{649740}$$

9. Berdasarkan soal (7), berapa probabilitas bahwa kartu yang terambil adalah 3 kartu Ten dan 2 kartu Jack.

   Jawab:

$$\text{P(3 kartu Ten dan 2 kartu Jack)} = \frac{({}_4C_3)({}_4C_2)}{{}_{52}C_5}$$

$$= \frac{\left(\frac{4!}{3!(4-3)!}\right)\left(\frac{4!}{2!(4-2)!}\right)}{\frac{52!}{5!(52-5)!}}$$

$$= \frac{24}{2598960} = \frac{1}{108290}$$

10. Berdasarkan soal (7), berapa probabilitas bahwa kartu yang terambil adalah 1 kartu *nine, ten, jack, queen, king* dengan urutan yang mana saja.

    Jawab:

$$\text{P(1 kartu Nine, Ten, Jack, Queen, King dengan urutan yang mana saja)}$$

$$= \frac{({}_4C_1)({}_4C_1)({}_4C_1)({}_4C_1)({}_4C_1)}{{}_{52}C_5}$$

$$= \frac{1024}{2598960} = \frac{64}{162435}$$

11. Berdasarkan soal (7), Berapa probabilitas bahwa kartu yang terambil adalah 3 kartu dengan pasangan yang mana saja dan 2 kartu lainnya.

    Jawab:

P(3 kartu dengan pasangan yang mana saja dan 2 kartu lainnya)

$$= \frac{(4 \cdot {}_{13}C_3)(3 \cdot {}_{13}C_2)}{{}_{52}C_5}$$

$$= \frac{(1144)(234)}{2598960} = \frac{267696}{2598960} = \frac{429}{4165}$$

12. Berdasarkan soal (7), Berapa probabilitas bahwa kartu yang terambil paling sedikit adalah 1 kartu As.

<u>Jawab:</u>

P(paling sedikit 1 Kartu As)=1-P(bukan kartu As)

$$= 1 - \frac{({}_{48}C_5)}{{}_{52}C_5} = 1 - \frac{\frac{48!}{5!(48! - 5)!}}{\frac{52!}{5!(52! - 5)!}}$$

$$= 1 - \frac{1712304}{2598960} = 1 - \frac{35673}{54145} = \frac{18472}{54145}$$

13. Sebuah kotak mainan terdiri dari 8 bola merah, 3 bola putih, dan 9 bola biru. Apabila secara random diambil 3 bola tanpa pengembalian, berapakah probabilitas terambilnya 3 bola merah.

<u>Jawab:</u>

P(3 bola merah)=jumlah kombinasi 3 bola merah yang diambil dari 8 bola merah dibagi dengan jumlah kombinasi 3 bola yang diambil dari 20 bola.

$$P(\text{semua (3) bola merah}) = \frac{{}_{8}C_3}{{}_{20}C_3} = \frac{\frac{8!}{3!(8-3)!}}{\frac{20!}{3!(20-3)!}}$$

$$= \frac{56}{1140} = \frac{14}{285}$$

14. Berdasarkan soal (13), berapa probabilitas terambil 3 bola putih.

<u>Jawab:</u>

$$P(\text{semua (3) bola putih}) = \frac{{}_{3}C_3}{{}_{20}C_3} = \frac{\frac{3!}{3!(3-3)!}}{\frac{20!}{3!(20-3)!}}$$

$$= \frac{1}{1140}$$

15. Berdasarkan soal (13), berapa probabilitas terambilnya 2 bola merah dan 1 bola putih.

Jawab:

$$=\frac{({}_8C_2)({}_3C_1)}{{}_{20}C_3}=\frac{\left(\frac{8!}{2!(8-2)!}\right)\left(\frac{8!}{2!(8-2)!}\right)}{\frac{20!}{3!(20-3)!}}$$

$$=\frac{28(3)}{1140}=\frac{84}{1140}=\frac{7}{95}$$

16. Berdasarkan soal (13), berapakah probabilitas terambilnya paling tidak 1 bola putih.

Jawab:

P(paling sedikit 1 bola putih)=1-P(tidak ada bola putih)

$$=1-\frac{{}_{17}C_3}{{}_{20}C_3}=1-\frac{\frac{17!}{3!(17-3)!}}{\frac{20!}{3!(20-3)!}}$$

$$=1-\frac{680}{1140}=1-\frac{34}{57}=\frac{23}{57}$$

17. Berdasarkan soal (13), berapakah probabilitas terambilnya masing-*masing* warna 1 bola.

Jawab:

P(masing-masing warna 1 bola)

$$=\frac{\left(\frac{8!}{1!(8-1)!}\right)\left(\frac{3!}{1!(3-1)!}\right)\left(\frac{9!}{1!(9-1)!}\right)}{\frac{20!}{3!(20-3)!}}$$

$$=\frac{(8)(3)(9)}{1140}=\frac{216}{1140}=\frac{18}{95}$$

18. Berdasarkan soal (13), berapakah probabilitas terambilnya berturut-turut bola merah, bola putih, dan bola biru.

Jawab:

P(masing-masing 1 bola dengan urutan merah, putih dan biru)

$$= \frac{1!}{3!} \cdot P(\text{masing-masing warna 1 bola})$$

$$= \frac{1}{6}\left(\frac{18}{95}\right) = \frac{3}{95}$$

19. Bagian kriminal kepolisian di distrik A menghitung probabilitas terjadinya pencurian kendaraan bermotor. Probabilitas dalam 1 hari tidak ada pencurian, 1, 2, 3, 4, dan 5 pencurian memiliki probabilitas masing-masing berturut-turut sebesar 0.23, 0.34, 0.26, 0.12, 0.04 dan 0.01. Berapa banyak kendaraan bermotor diperkirakan tercuri per hari, bila dianggap pencurian lebih dari 5 kendaraan bermotor tidak mungkin terjadi.

    Jawab:

    $$E(X) = 0(0.23) + 1(0.34) + 2(0.26) + 3(0.12) + 4(0.04) + 5(0.01)$$
    $$= 0 + 0.34 + 0.52 + 0.36 + 0.16 + 0.05$$
    $$= 1.43$$

20. Seorang anak SMA yang sedang melamar magang di perusahaan “Sangat Kecil” sedang mempertimbangkan untuk menerima atau menolak gaji yang di tawarkan kepadanya. Jika diterima, gajinya sebesar Rp 97200/bulan. Jika ditolak, dia berharap mendapat pekerjaan di tempat lain dengan gaji Rp 135 rb/bulan. Berapa probabilitas dia mendapat pekerjaan dengan gaji RP 135 rb/bulan.

    Jawab:

    Dia akan menolak gaji sebesar Rp 97.2 rb/bulan jika ada harapan mendapat gaji Rp 135 rb/bulan.

    Maka probabilitas dia mendapat pekerjaan dengan gaji RP 135 rb/bulan:

    $$X \cdot E(X) > 97200$$
    $$135000P > 97200$$
    $$P > \frac{97200}{135000}$$
    $$p > 0.72$$

MODUL 6

DISTRIBUSI PROBABILITAS

Pokok Bahasan:

- Pengertian tentang Distribusi Probabilitas
- Distribusi Binomial
- Distribusi Poisson

6.1 Pengertian tentang Distribusi Probabilitas

Kumpulan data yang sangat banyak dan beragam dapat disederhanakan melalui penyajian yang praktis. Salah satunya adalah dengan membuat "kelas" untuk data yang sangat beragam. Selanjutnya dapat dihitung berapa frekuensi (jumlah data) di masing-masing kelas. Di sini, distribusi probabilitas berkaitan erat dengan distribusi frekuensi. Secara teoritis dapat dinyatakan bahwa distribusi frekuensi adalah distribusi probabilitas yang menggambarkan bagaimana output diharapkan beragam. Karena berkaitan erat dengan ekspektasi, distribusi ini sangat berguna untuk membuat inferensi dan keputusan dalam kondisi *uncertainty.*

Ada perbedaan yang dapat dilihat antara distribusi frekuensi dan distribusi probabilitas. Distribusi frekuensi adalah daftar frekuensi untuk seluruh output percobaan yang secara riil terjadi ketika percobaan dilakukan sedangkan distribusi probabilitas adalah daftar probabilitas untuk seluruh output yang mungkin dapat menjadi hasil jika eksperimen dilakukan. Contoh sederhana dari distribusi probabilitas adalah pemungutan suara. Dalam pemilihan bupati periode ke depan, jumlah suara dan probabilitas yang diperoleh kandidat A adalah sebagai berikut.

Jumlah Suara	1000	2000	3000	4000
Probabilitas yg akan terjadi	0.1	0.3	0.4	0.2

Total probabilitas dari seluruh jumlah suara yang memilih kandidat A sama dengan satu.

Distribusi probabilitas bisa berdasarkan pertimbangan teoritis (seperti pelemparan koin) ataupun penilaian subjektif berdasarkan kemungkinan tertentu (seperti pemungutan suara). Selain itu, distribusi probabilitas juga bisa dibuat berdasarkan pengalaman. Sebagai contoh adalah penetapan premi asuransi.

Distribusi probabilitas bisa dikategorikan menjadi dua, yaitu diskrit dan kontinus. Distribusi probabilitas diskrit hanya mencakup angka tertentu. Sebagai contoh adalah probabilitas seseorang lahir pada hari Senin. Sedangkan distribusi probabilitas kontinus mencakup angka dalam suatu interval tertentu; jadi kita “tidak dapat” membuat daftar seluruh angka yang memungkinkan.

6.2 Distribusi Binomial

Ketika menonton perlombaan masak, pertandingan sepak bola dan sebagainya, peristiwa yang mungkin terjadi hanya dua, yaitu sukses dan gagal. Itu merupakan beberapa aplikasi rill dari ditribusi probabilitas diskrit, dikenal dengan distribusi binomial. Berbagai percobaan yang dilakukan disebut dengan proses Bernaulli.

Berikut ini adalah beberapa karakteristik dari percobaan Bernaulli, yaitu:

- Setiap percobaan hanya menghasilkan *dua* peristiwa yang mungkin, seperti: gagal dan berhasil, ya atau tidak, baik atau cacat.
- Probabilitas munculnya suatu peristiwa adalah konstan sepanjang waktu. Sebagai contoh, probabilitas sukses untuk perlombaan masak adalah 0.5; terlepas kapanpun perlombaan masak dilakukan.
- Semua percobaan bersifat independen secara statistik (*statistically independent)*, artinya peristiwa dari suatu percobaan tidak mempengaruhi atau dipengaruhi peristiwa dalam percobaan lain.

Secara matematis, rumus distribusi binomial dapat dinyatakan sebagai berikut.

$$p(x!n, p) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

Keterangan:

$p =$ probabilitas sukses

$q =$ probababilitas gagal $= 1 - p$

$n =$ jumlah percobaan yang dilakukan

$x =$ jumlah sukses yang diinginkan

Rumus rata-rata dan standar deviasi dari distribusi binomial dapat dinyatakan sebagai berikut.

Rata-rata: $\mu = n.p$

Deviasi Standar: $\sigma = \sqrt{npq}$

Contoh 6.1

Di sebuah perusahaan konsultan "TERKENAL", ada sembilan indikator kinerja yang bisa menaikkan status karyawan, salah satunya adalah karya ilmiah. Manager ingin melihat kinerja si A. Diambil secara random 4 buah indikator yang menunjukkan kinerja si A tersebut. Berapa probabilitas bahwa 3 diantara indikator tersebut adalah berasal dari karya ilmiah?

Jawab:

Probabilitas dari indikator 'karya ilmiah' adalah:

$p = 1/9$

Probabilitas dari indikator yang bukan 'karya ilmiah' adalah:

$q = 1 - p = 1 - 1/9 = 8/9$

Berdasarkan soal tersebut dapat ditentukan:

$n = 4$ dan $x = 3$

Jadi probabilitas tiga diantara indikator tersebut adalah karya ilmiah adalah:

$$p(x = 3) = \frac{4!}{3!(4-3)!}(1/9)^3(8/9)$$

$$= \frac{4x1}{729} x \frac{8}{9}$$

$$= \frac{32}{656}$$

Dengan berbagai skenario probabilitas, bentuk distribusi binomial dapat dijelaskan sebagai berikut, yaitu:

- Jika p makin mendekati 0, distribusi binomial makin condong ke kanan.
- Jika $p = 0{,}5$, distribusi binomial akan simetris.
- Jika p makin mendekati 1, distribusi binomial makin condong ke kiri.
- Jika p konstan dan n bertambah, distribusi binomial akan mendekati simetris (normal).

Dalam distribusi binomial, x (banyaknya sukses) adalah variabel random, sedangkan p dan n adalah parameter.

6.3 Distribusi Poisson

Distribusi Poisson digunakan untuk menggambarkan sebuah proses seperti distribusi panggilan telepon, rata-rata kedatangan pasien di UGD, jumlah kecelakaan di tol, dan lain sebagainya. Pendekatan Poisson akan memberikan hasil yang semakin baik jika *n* ditambah dan *p* mendekati 0 atau 1. Hasil perhitungan dari distribusi Poisson tidak akan berbeda jauh dengan distribusi Binomial.

Suatu proses dikategorikan sebagai distribusi Poisson apabila memiliki beberapa karakteristik berikut (sebagai contoh adalah kedatangan truk di jalan tol):

1. Tingkat kedatangan rata-rata untuk setiap satuan waktu dapat diestimasi dengan menggunakan data di masa lalu.
2. Jika periode waktu dibagi ke dalam interval waktu tertentu, seperti detik, menit atau jam, maka dapat dibuktikan bahwa pernyataan berikut adalah benar:
   - Rata-rata tingkat kedatangan ***per satuan waktu*** tertentu adalah konstan. Sebagai contoh, dalam satu jam, ada 120 truk yang masuk jalan tol. Maka dalam waktu 30 menit ada 60 truk yang masuk jalan tol.
   - Banyaknya kedatangan dalam suatu interval waktu ***bersifat independen*** dengan apa yang terjadi (atau banyaknya kedatangan) pada ***interval waktu sebelumnya.***
   - Banyaknya kedatangan dalam suatu interval waktu bersifat independen terhadap waktu pada ***interval waktu tersebut.***
   - Probabilitas suatu kedatangan pada interval waktu yang sangat pendek adalah sangat kecil. Dengan demikian, probabilitas lebih dari satu kedatangan dalam interval waktu yang sangat pendek tersebut akan mendekati nol.

Rumus distribusi Poisson adalah:

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

Keterangan:

$P(x) =$ probabilitas terjadinya x

$e =$ bilangan Napierian = 2.71828

$x =$ variabel random yang memiliki nilai integer 0, 1, 2, 3, 4, dan seterusnya

$\mu =$ rata-rata jumlah kejadian dalam interval waktu tertentu

Yang perlu dicatat, distribusi Poisson dapat digunakan untuk mendekati distribusi Binomial ketika $n \geq 20$ dan probabilitas sukses (p) adalah $p \leq 0{,}05$. Rumusnya sama dengan rumus distribusi Poisson di atas dengan $\mu = np$.

<u>Contoh 6.2</u>

Dika bekerja sebagai analis di perusahaan jasa tol Cikampek. Berdasarkan data yang diterima, dari 1000 mobil yang melalui tol Cikampek, rata-rata ada 1 mobil yang mengalami kerusakan ban. Jika ada 10.000 mobil yang masuk tol Cikampek, berapa probabilitas bahwa ada 8 mobil yang mengalami kerusakan ban?

Jawab:

$n = 10.000$ dan $\mu = 10$

$$p(x = 8) = \frac{10^8 e^{-10}}{8!} = 0{,}1126$$

6.4 Latihan: Soal & Jawab

1. Dari bagian satuan lalu lintas distrik B diperoleh informasi bahwa rata-rata kematian karena kecelakaan lalu lintas per tahun adalah 4 kematian dari 100 rb penduduk. Dari penduduk yang berjumlah 200 rb, hitunglah probabilitas bahwa sepuluh kematian terjadi karena kecelakaan lalu lintas.

   Jawab:

   $$p(x=10) = \frac{8^{10} \cdot e^{-8}}{10!} = 0.0993$$

2. Berdasarkan soal (1), hitunglah probabilitas terjadi 4 sampai 6 kematian karena kecelakaan lalu lintas.

   Jawab:

   $$p(4 \le x \le 6) = p(x=4) + p(x=5) + p(x=6)$$

   $$p(x=4) = \frac{8^4 \cdot e^{-8}}{4!} = 0.0573$$

   $$p(x=5) = \frac{8^5 \cdot e^{-8}}{5!} = 0.0916$$

   $$p(x=6) = \frac{8^6 \cdot e^{-8}}{6!} = 0.1221$$

   Jadi $p(4 \le x \le 6) = 0.0573 + 0.0916 + 0.1221 = 0.2710$

3. Berdasarkan soal (1), hitunglah probabilitas terjadi kurang dari 5 kematian karena kecelakaan lalu lintas.

Jawab:

$$p(x<5) = p(x=0) + p(x=1) + p(x=2) + p(x=3) + p(x=4)$$

Jadi $p(x<5) = 0.0003 + 0.0027 + 0.0107 + 0.0286 + 0.0573 = 0.0996$

4. Berdasarkan soal (1), hitunglah probabilitas terjadi lebih dari 2 kematian karena kecelakaan lalu lintas.

Jawab:

$$p(x>2) = 1 - [p(x=0) + p(x=1) + p(x=2)]$$

Jadi $p(x>2) = 1 - (0.0003 + 0.0027 + 0.0107)$

$= 1 - 0.0137$

$= 0.9863$

5. Bagian pemasaran apartemen “Depok Tenang” mampu menjual rata-rata 2 rumah per minggu. Hitunglah probabilitas bahwa dalam satu minggu tertentu dia hanya dapat menjual satu rumah.

Jawab:

$\mu = 2$

$$p(x=1) = \frac{2^1 \cdot e^{-2}}{1!} = 0.2707 (atau\ \ 2x0.13534)$$

6. Bagian kecelakaan Tol Cikampek melaporkan bahwa rata-rata 1 mobil dari 1000 mobil yang melalui tol mengalami kerusakan ban. Apabila pada hari tertentu ada 10.000 mobil yang melalui Tol Cikampek, berapa probabilitas bahwa 8 mobil mengalami kerusakan ban.

Jawab:

$n = 10000$

$\mu = 10$

$$p(x=8) = \frac{10^8 \cdot e^{-10}}{8!} = 0.1126$$

7. Dari sekian serum yang ditujukan untuk kekebalan tubuh, probabilitas seseorang akan menderita reaksi buruk dari serum tersebut adalah 0.001. Dari 2000 orang yang mendapat injeksi dari serum tersebut, hitunglah probabilitas 3 orang menderita reaksi buruk.

Jawab:

$$p(x|\mu) = \frac{\mu^x e^{-u}}{x!}$$

$$\mu = np$$

$$n = 2000 \qquad p = 0.001$$

$$\mu = 2000(0.001) = 2$$

$$p(x=3) = \frac{2^3 e^{-2}}{3!} = \frac{8(0.13534)}{6} = 0.1804$$

8. Berdasarkan soal (7), hitunglah probabilitas lebih dari 2 orang menderita reaksi buruk.

   Jawab:

$$p(x>2) = 1 - [p(x=0) + p(x=1) + p(x=2)]$$

$$p(x>2) = 1 - \left(\frac{2^0 e^{-2}}{0!} + \frac{2^1 e^{-2}}{1!} + \frac{2^2 e^{-2}}{2!} \right)$$

$$p(x>2) = 1 - (0.1353 + 0.2707 + 0.2707)$$

$$p(x>2) = 1 - 0.6767 = 0.3233$$

9. YLKI akan melakukan dengar pendapat dengan 10 orang konsumen yang terdiri dari 6 pria dan 4 wanita. Diambil sampel random sebanyak 4 orang yang akan menjadi peserta diskusi diantara 10 orang tersebut. Hitunglah probabilitas bahwa seluruh peserta diskusi adalah wanita.

   Jawab:

$$n = 4 \qquad p = 0.4 \qquad q = 0.6$$

$$p(x=4) = \frac{4!}{4!(4-4)!}(0.4)^4(0.6)^0 = 0.0256$$

10. Berdasarkan soal (9), hitunglah probabilitas bahwa 2 orang peserta diskusi adalah wanita.

    Jawab:

$$n = 4 \qquad p = 0.4 \qquad q = 0.6$$

$$p(x=2) = \frac{4!}{2!(4-2)!}(0.4)^2(0.6)^2 = 0.3456$$

11. Berdasarkan soal (9), hitunglah probabilitas bahwa 3 orang peserta diskusi adalah pria.

    Jawab:

$n = 4 \quad\quad p = 0.4 \quad\quad q = 0.6$

$$p(x=1) = \frac{4!}{1!(4-1)!}(0.4)^1(0.6)^3 = 0.3456$$

Note: Tiga orang peserta pria sama artinya dengan satu orang peserta wanita.

12. Berdasarkan soal (9), hitunglah probabilitas bahwa peserta diskusi paling banyak adalah 2 pria.

Jawab:

$p(x \geq 2) = p(x=2) + p(x=3) + p(x=4)$

$= 0.3456 + 0.1536 + 0.0256$

$= 0.5248$

Note: Paling banyak terdiri dari 2 peserta pria sama artinya dengan dua atau lebih peserta wanita.

13. Berdasarkan soal (9), hitunglah probabilitas bahwa peserta diskusi paling sedikit adalah 3 wanita.

Jawab:

$p(x \geq 3) = p(x=3) + p(x=4)$

$$p(x=3) = \frac{4!}{3!(4-3)!}(0.4)^3(0.6)^1 = 0.1536$$

$$p(x=4) = \frac{4!}{4!(4-4)!}(0.4)^4(0.6)^0 = 0.0256$$

Jadi $p(x \geq 3) = 0.1536 + 0.0256 = 0.1792$

14. Dari total produksi alat-alat pertanian yang dihasilkan pabrik ABC, diketahui 10%-nya tidak memenuhi standar. Dari sampel random sebanyak 10 peralatan, hitunglah probabilitas bahwa 2 peralatan tidak memenuhi standar dengan menggunakan distribusi binomial.

Jawab:

Probabilitas diperoleh peralatan yang tidak memenuhi standar (rusak) adalah p=0.1. Apabila x menunjukkan jumlah peralatan yang tidak memenuhi standar dari 10 kali pengambilan, maka dengan menggunakan rumus binomial:

$$p(x=2) = \frac{10!}{2!(10-2)!}(0.1)^2(0.9)^8 = 0.1937$$

15. Berdasarkan soal (14), hitunglah probabilitas bahwa 2 peralatan tidak memenuhi standar dengan menggunakan distribusi Poisson untuk distribusi binomial.

    Jawab:

    Diperoleh U=np =(10)(0.1)=1

    $$p(x=2)=\frac{(1)^2 e^{-1}}{2!}=0.1839$$

    Note:

    Pada umumnya pendekatan dianggap baik jika $p \le 0.1$ dan $U = np \le 5$

16. Perusahaan farmasi "Terkenal" mengetahui bahwa rata-rata 5% dari tablet yang dijual tidak memenuhi layak uji sehingga harus disortir. Berapa probabilitas kurang dari 8 tablet harus disortir dari 200 sampel tablet.

    Jawab:

    Digunakan pendekatan kurva normal:

    $n = 200 \qquad p = 5\%$

    $\mu = np = 5\%(200) = 10$

    $\sigma = \sqrt{npq} = \sqrt{200(5\%)(95\%)} = 3.08$

    $$Z = \frac{7.5-10}{3.08} = -0.81$$

    Luas kurva adalah 0.2910

    Luas kurva normal 7.5 atau kurang =0.5-0.2910=0.2090

17. Dalam proses pengobatan, 22% pasien darah tinggi mengalami efek samping dari obat yang diberikan. Dari 120 pasien, berapa probabilitas lebih dari 30 orang terkena efek samping dari obat tersebut.

    Jawab:

Digunakan pendekatan kurva normal (untuk soal binomial):

$n = 120 \qquad p = 22\%$

$\mu = np = (120)g22\% = 26.4$

$\sigma = \sqrt{npq} = \sqrt{120(22\%)(78\%)} = 4.54$

$Z = \dfrac{30.5 - 26.4}{4.54} = 0.90$

Luas kurva adalah 0.3159

Luas kurva normal lebih dari 30.5 adalah

=0.5-0.3159=0.1841

18. Untuk membantu siswa yang kesulitan memahami topik probabilitas, seorang guru menggunakan contoh pelemparan dadu. Dadu dilempar sebanyak 120 kali. Hitunglah probabilitas bahwa yang muncul sisi 4 adalah 18 kali atau kurang.

Jawab:

Digunakan pendekatan kurva normal (untuk soal binomial):

$\mu = np = (120)(1/6) = 20$

$\sigma = \sqrt{npq} = \sqrt{120(1/6)(5/6)} = 4.08$

$Z = \dfrac{18.5 - 20}{4.08} = -0.39$

Luas kurva adalah 0.1443

Luas kurva normal 18.5 atau kurang adalah

=0.5-0.1443=0.3557

19. Berdasarkan soal (18), hitunglah probabilitas bahwa yang muncul sisi 4 adalah 14 kali atau kurang.

Jawab:

Digunakan pendekatan kurva normal (untuk soal binomial):

$Z = \dfrac{14.5 - 20}{4.08} = -1.35$

Luas kurva adalah 0.4115

Luas kurva normal 14.5 atau kurang adalah

=0.5- 0.4115=0.0885

20. Sekali lagi, agar topik probabilitas lebih jelas, seorang guru menggunakan contoh pelemparan koin. Dari 500 kali pelemparan koin,

hitunglah probabilitas bahwa yang muncul adalah sisi “kepala” yang muncul dari 250 kali pelemparan adalah tidak lebih dari 30.

Jawab:

Digunakan pendekatan kurva normal (untuk soal binomial):

$\mu = np = (500)(1/2) = 250$

$\sigma = \sqrt{npq} = \sqrt{500(1/2)(1/2)} = 11.18$

$Z_1 = \dfrac{219.5 - 250}{11.18} = -2.73$

Luas kurva adalah 0.4968

$Z_2 = \dfrac{280.5 - 250}{11.18} = 2.73$

Luas kurva normal 2.73 adalah 0.4968

Jadi luas kurva normal 219.5 sampai dengan 280.5 adalah

=0.4968+0.4968=0.9936

MODUL 7
DISTRIBUSI PROBABILITAS (LANJUTAN)

Pokok Bahasan:

- Distribusi Hipergeometrik
- Distribusi Normal
- Penggunaan Distribusi Normal sebagai pendekatan Distribusi Binomial dan Poisson
- Variabel Normal Standar
- Menggunakan Tabel Distribusi Normal Standar
- Distribusi Normal sebagai pendekatan Distribusi Binomial

7.1 Distribusi Hipergeometrik

Distribusi binomial sangat sering digunakan dalam konteks pengambilan sampel. Yang perlu diingat adalah, distribusi binomial mensyaratkan independensi dan probabilitas konstan. Apabila syarat tersebut tidak terpenuhi, dapat digunakan probabilitas hipergeometris. Dalam hal ini, apabila persyaratan distribusi binomial tidak terpenuhi, rumus binomial masih bisa digunakan jika ukuran sampel n tidak lebih dari 5% dari elemen populasi N atau $n \leq 0{,}05N$.

Rumus Probabilitas Hipergeometris:

$$P(x) = \frac{C_x^X C_{n-x}^{N-R}}{C_n^N} = \frac{\dfrac{S!}{x!(S-x)!} \, x \, \dfrac{(N-S)!}{(n-x)!(N-S-n+x)!}}{N!\big/n!(N-n)!}$$

Keterangan:

N = jumlah populasi

n = jumlah sampel

X = jumlah sukses dalam populasi

x = jumlah sukses dalam sampel

<u>Contoh 1</u>

Pak Andi ingin memastikan bahwa produksi lampu taman di pabriknya memiliki mutu yang baik, artinya jumlah produk cacat relatif kecil. Sebelum dikirim ke toko-toko retail, Pak Andi memeriksa salah satu kotak siap kirim. Satu kotak berisi 100 lampu, 90 diantaranya baik dan 10 cacat. Kemudian dilakukan sampling dengan mengambil 6 lampu untuk setiap kotaknya. Berapa probabilitas jumlah barang baik sebanyak empat?

Jawab:

$n = 6$

$p = 90/100$

$$P(4) = \frac{C_4^{90} C_{6-4}^{100-90}}{C_6^{100}}$$

$$P(4) = \frac{\dfrac{90!}{4!86!} \; \dfrac{10!}{2!8!}}{\dfrac{100!}{6!94!}}$$

$$P(4) = 0{,}0965$$

7.2 Distribusi Normal

Distribusi normal merupakan distribusi yang sering digunakan dan memiliki peranan penting dalam statistik. Alasan dari penggunaan distribusi normal adalah:

- Distribusi normal memiliki karakteristik yang dapat diterapkan pada berbagai situasi. Terlebih lagi dalam konteks sampling untuk tujuan inferensia.
- Dalam banyak kasus, distribusi ini sama dengan distribusi frekuensi aktual.

Suatu distribusi dikatakan distribusi normal apabila memiliki karakteristik sebagai berikut:

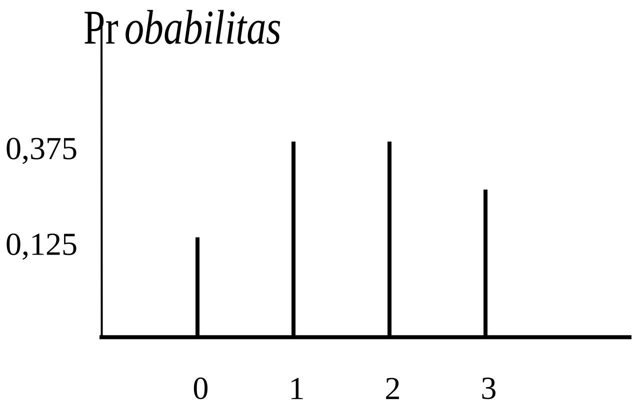
- Kurvanya memiliki puncak tunggal dan modus tungal dengan bentuk seperti lonceng.
- Rata-rata terletak di tengah distribusi dan distribusi bersifat simetris di sekitar garis tegak lurus yang ditarik melalui rata-rata.
- Dengan distribusi yang simetris tersebut, distribusi normal memiliki nilai median, mean dan modus yang sama.
- Kedua ekor (ujung) kurva memanjang tak terbatas dan tidak pernah menyentuh sumbu horisontal.
- Distribusi normal juga dikenal dengan distribusi Gauss karena penemunya adalah Karl Gauss.

7.3 Area dan Parameter pada Distribusi Probabilitas Normal

Area dibawah kurva normal menunjukkan probabilitas dimana total area untuk kurva normal adalah satu untuk berapapun nilai μ dan σ. Pada distribusi normal, probabilitas untuk suatu interval digambarkan dalam suatu luas wilayah.

Distribusi kontinu berbeda dengan distribusi diskrit dimana distribusi diskrit memiliki probabilitas nilai setiap variabel yang ditunjukkan oleh panjang garis tegak lurus di atasnya. Sehingga jumlah panjang garis tegak lurus di atasnya harus sama dengan satu.

Gambar 1
Distribusi Probabilitas Diskrit



Distribusi normal dari variabel random $x(-\infty < x < \infty)$ seperti pada gambar 1 memiliki fungsi kepadatan yang dapat dirumuskan sebagai berikut:

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

dengan rata-rata μ dan varian σ^2. Fungsi kepadatan distribusi normal harus memenuhi syarat bahwa

$$P(-\infty < x < \infty) = \int_{-\infty}^{\infty} f(x)dx = 1$$

Disitribusi normal variabel random x biasanya dinotasikan dengan $X \sim N(\mu, \sigma^2)$.

7.4 Variabel Normal Standar

Distribusi normal memiliki jumlah yang tak terhingga. Untuk mencari probabilitas suatu interval dari variabel random kontinu, dapat digunakan distribusi normal standar dengan rata-rata $\mu = 0$ dan deviasi standar $\sigma = 1$. Selanjutnya, variabel random dalam distribusi normal standar dinotasikan dengan Z. Rumus variabel normal standar Z adalah:

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

Keterangan:

Z = jumlah standar deviasi (variabel x) terhadap rata-rata
(Note: Nilai Z dapat diartikan sebagai berapa kali deviasi standar suatu nilai variabel random menyimpang dari rata-ratanya.)

X = nilai variabel random

$\mu =$ rata-rata distribusi variabel random

$\sigma =$ standar deviasi

Contoh 2

TK "Harapan Kita Semua" melakukan penimbangan berat badan anak-anaknya. Hal ini dilakukan untuk meyakinkan bahwa semua anak tercukupi kebutuhan gizinya. Berat badan terdistribusi normal dengan rata-rata dan deviasi standar adalah 25 dan 5. Berapa nilai variabel normal standar beberapa anak yang memiliki berat badan 22, 27 dan 30 (satuan kg).

Jawab:

Untuk $x = 22$ → $Z = \frac{22 - 25}{5} = -0,6$

Untuk $x = 27$ → $Z = \frac{27 - 25}{5} = 0,4$

Untuk $x = 30$ → $Z = \frac{30 - 25}{5} = 1$

7.5 Menggunakan Tabel Distribusi Normal Standar

Tabel distribusi normal standar dapat dilihat di berbagai buku statistik. Luas seluruh wilayah di bawah kurva harus sama dengan satu. Karena kurva simetris, luas wilayah di sebelah kanan dan kiri garis tegak lurus adalah sama yaitu 0,5. Untuk lebih jelasnya, dapat dilihat contoh 6.

Contoh 3

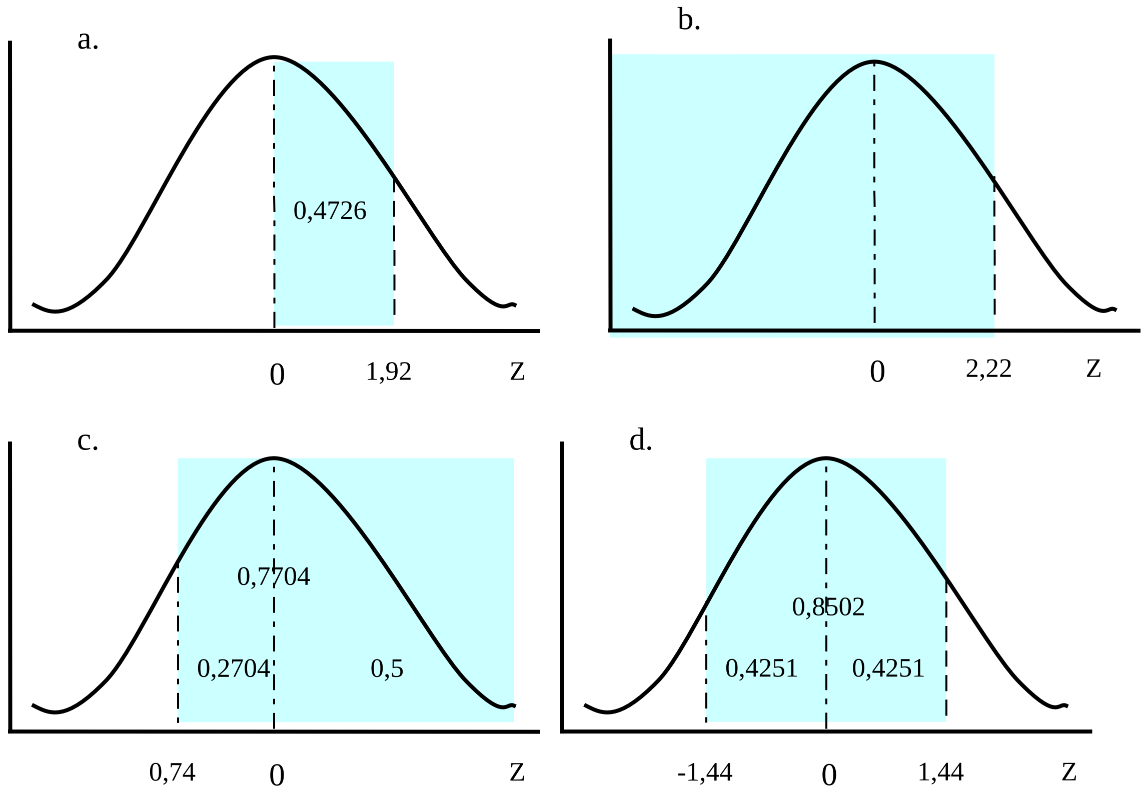
Hitung nilai Z apabila:

a. luas kurva normal antara 0 dan Z adalah 0,4726
b. luas kurva normal ke kiri dari Z adalah 0,9868
c. luas kurva normal ke kanan dari Z adalah 0,7704
d. luas kurva normal antara -Z dan Z adalah 0,8502

Jawab: (Cek di tabel)

a. luas kurva normal antara 0 dan Z sebesar 0,4726 berasal dari nilai Z =1,92
b. luas kurva normal ke kiri dari Z adalah 0,9868 dapat diartikan sebagai luas wilayah antara 0 dan Z yaitu 0,9868 - 0,5 = 0,4868. Luas daerah 0,4868 berasal dari nilai Z = 2,22.

Jawaban dari soal pada contoh 3 (a,b,c, dan d) dapat ditunjukkan pada gambar berikut:



c. Luas kurva normal ke kanan dari Z adalah 0,7704 dapat diartikan sebagai luas wilayah antara 0 dan Z, yaitu 0,7704 - 0,5 = 0,2704 (yang berada di sebelah kiri dari 0). Luas daerah 0, 2704 berasal dari nilai Z = -0,74.

d. Luas kurva normal antara -Z dan Z adalah 0,8502 dapat diartikan sebagai luas wilayah antara –Z, dan 0 dan Z, yaitu 0,5 x 0,8502 = 0,4251. Luas daerah 0, 4251 berasal dari nilai Z = 1,44 dan Z= -1,44.

7.6 Distribusi Normal sebagai pendekatan Distribusi Binomial

Distribusi normal juga bisa digunakan sebagai pendekatan dalam distribusi diskrit, seperti distribusi binomial. Pendekatan ini akan memberikan hasil yang semakin baik jika parameter n memiliki jumlah yang semakin besar dan p mendekati 0,5 baik dari atas maupun dari bawah. Atau dengan kata lain, $np \geq 5$. Semakin besar nilai n, distribusi probabilitas cenderung mendekati distribusi normal. Untuk mencari nilai rata-rata dan standar deviasi, dapat digunakan rumus sebagai berikut:

$$\mu = np \text{ dan } \sigma = \sqrt{npq}$$

7.7 Latihan: Soal & Jawab

1. Suatu variabel random mempunyai distribusi normal dengan rata-rata 80 dan standar deviasi 4.8. Berapa probabilitas variabel random akan mempunyai nilai kurang dari 87.2.

   Jawab:

   $\mu = 80 \qquad \sigma = 4.8$

   Probabilitas kurang dari 87.2 adalah

   $$Z = \frac{x-\mu}{\sigma} = \frac{87.2-80}{4.8} = 1.5$$

   Luas kurva 1.5 adalah 0.4332

   Luas kurva 87.2 ke kiri $= 0.5 + 0.4332 = 0.9332$

2. Suatu distribusi normal memiliki rata-rata 62.4. Hitunglah standar deviasi jika luas kurva normal di sebelah kanan 79.2 adalah 20%.

   Jawab:

   Luas daerah antara 62.4 dan 79.2 adalah 50%-20%=30%

   Luas yang mendekati adalah 0.2995, ditunjukkan oleh nilai Z=0.84

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

   $$0.84 = \frac{79.2-62.4}{\sigma}$$

   $$0.84\sigma = 16.8$$

   $\sigma = 20$ (nilai deviasi standar)

3. Seorang petugas pengalengan sedang menyeleksi ikan yang masuk ke pabrik. Pabrik menerima ikan yang memiliki panjang rata-rata 4.54 inci dengan standar deviasi 0.25 inci. Apabila distribusi panjang ikan tersebut mendekati distribusi normal, berapa persentase dari ikan-ikan tersebut memiliki panjang lebih dari 5 inci.

   Jawab:

$\mu = 4.54 \qquad \sigma = 0.25$

Persentase lebih dari 5 inci:

$$Z = \frac{x-\mu}{\sigma} = \frac{5-4.54}{0.25} = 1.84$$

Luasnya adalah 46.71%

Luas lebih dari 5 inci adalah

50%-46.71%=3.29%

4. Berdasarkan soal (3), berapa persentase dari ikan-ikan tersebut memiliki panjang kurang dari 4 inci.

   Jawab:

   $\mu = 4.54 \qquad \sigma = 0.25$

   Persentase kurang dari 4 inci:

   $$Z = \frac{x-\mu}{\sigma} = \frac{4-4.54}{0.25} = -2.16$$

   Luasnya adalah 48.46%

   Luas kurang dari 4 inci adalah

   50%-48.46%=1.54%

5. Berdasarkan soal (3), berapa persentase dari ikan-ikan tersebut memiliki panjang 4.4 sampai 4.6 inci.

   Jawab:

   $\mu = 4.54 \qquad \sigma = 0.25$

   Persentase antara 4.4 sampai 4.6 inci:

   $$Z_1 = \frac{x-\mu}{\sigma} = \frac{4.4-4.54}{0.25} = -0.56$$

   Luasnya adalah 21.23%

   $$Z_2 = \frac{x-\mu}{\sigma} = \frac{4.6-4.54}{0.25} = 0.24$$

   Luasnya adalah 9.48%

   Luas 4.4 inci sampai 4.6 inci = 21.23%+9.48%=30.71%

6. Pabrik minuman botol teh "Kami' mengeset mesin pengisi botol sehingga rata-rata isi botol adalah 200 ml. Jika volume minuman tersebut terdistribusi secara normal dengan standar deviasi 15 ml, berapa bagian yang berisi lebih dari 224 ml.

Jawab:

$\mu = 200 \qquad \sigma = 15$

Bagian yang berisi lebih dari 224 ml:

Luas kurva normal 200-224

$$Z = \frac{x-\mu}{\sigma} = \frac{224-200}{15} = 1.6$$

Luasnya adalah 0.4452

Bagian yang melebihi 224 adalah

= 50%-44.52%=5.48%

7. Berdasarkan soal (6), berapa probabilitas seluruh botol akan terisi 191 ml sampai 209 ml.

   Jawab:

   $\mu = 200 \qquad \sigma = 15$

   Bagian yang berisi 191 ml sampai 209 ml:

   $$Z_1 = \frac{x-\mu}{\sigma} = \frac{91-200}{15} = -0.60$$

   Luasnya adalah 0.2257

   $$Z_2 = \frac{x-\mu}{\sigma} = \frac{209-200}{15} = 0.60$$

   Luasnya adalah 0.2257

   Luas seluruhnya (191-209) adalah

   = 0.2257+0.2257=0.4514

8. Berdasarkan soal (6), berapa probabilitas botol yang terisi lebih dari 230 ml bila diketahui produksi pabrik tersebut 1000 botol.

   Jawab:

   $\mu = 200 \qquad \sigma = 15$

   Dari 1000 botol, yang berisi melebihi 230 ml:

   $$Z = \frac{x-\mu}{\sigma} = \frac{230-200}{15} = 2.00$$

   Luasnya adalah 0.4772

   Luas lebih dari 230 adalah

   = 0.5+0.4772=0.0228

   Banyak minuman (botol) adalah

   =0.0228 x 1000 botol

   =23 botol

9. Berdasarkan soal (6), di bawah nilai berapakah untuk diperoleh 25% isi terendah.

<u>Jawab:</u>

$\mu = 200 \qquad \sigma = 15$

Nilai X untuk diperoleh 25% isi terendah adalah sbb.

Luas kurva normal yang mendekati 25% adalah 0.2486,

ditunjukkan oleh nilai Z= -0.67

$$-0.67 = \frac{X - 200}{15}$$

X=189.95 (dibulatkan 190)

10. Di sebuah SMA "Unggulan", diketahui nilai rata-rata hasil ujian matematika adalah 82 dengan standar deviasi adalah 5. Semua siswa dengan nilai dari 88 sampai 94 mendapat nilai B. Bila nilai matematika tersebut terdistribusi normal dan ada 8 siswa yang mendapat nilai B, berapa banyak siswa yang mengikuti ujian tersebut.

<u>Jawab:</u>

$\mu = 82 \qquad \sigma = 5$

Nilai 88-94 dapat nilai B (adalah 8 orang)

$$Z_1 = \frac{94 - 82}{5} = 2.4$$

Luasnya adalah 0.4918

$$Z_2 = \frac{88 - 82}{5} = 1.20$$

Luasnya adalah 0.3849

Luas 88-94 =0.4918-0.3849=0.1069

Apabila jumlah siswa yang menempuh ujian matematika sebanyak X, maka

10.69%gX=8

X=74.8 (ada 75 siswa)

11. Bagian HRD Perusahaan "Semangat" menyeleksi para pelamar di perusahaannya. Hasil ujian seleksi menunjukkan skor rata-rata adalah 500 dengan standar deviasi adalah 50. Distribusi skor mengikuti distribusi normal. Manajemen perusahaan mempertimbangkan bahwa 6% pelamar yang memiliki skor tertinggi akan menduduki posisi

penting. Berapa skor terendah yang harus dicapai pelamar untuk mendapatkan posisi tersebut.

Jawab:

$\mu = 500 \qquad \sigma = 50$

Luas kurva normal yang mendekati 44% adalah 0.4394,

ditunjukkan oleh nilai Z=1.55

$$1.55 = \frac{X - 500}{50}$$

$X = 577.5$ (skor terendah)

12. Berdasarkan soal (11), diputuskan bahwa pelamar yang memiliki skor 400 atau kurang akan langsung ditolak. Berapa persen kira-kira pelamar yang ditolak.

Jawab:

$\mu = 500 \qquad \sigma = 50$

Persentase yang ditolak:

$$Z = \frac{470 - 500}{50} = -0.6$$

Luasnya adalah 0.2257

Luas dari 470 ke bawah=$0.5 - 0.2257$=0.2743

atau 27.43%

13. Berdasarkan soal (11), diputuskan bahwa pelamar yang memiliki skor antara 400 sampai 485 akan dipertimbangkan sebagai cadangan. Jika ada 1000 pelamar, berapa pelamar yang masuk sebagai cadangan.

Jawab:

$\mu = 500 \qquad \sigma = 50$

Jumlah pelamar yang masuk cadangan:

$$Z_1 = \frac{485 - 500}{50} = -0.30$$

Luasnya adalah 0.1179

$$Z_2 = \frac{525 - 500}{50} = 0.50$$

Luasnya adalah 0.1915

Luas 484-525=0.1179+0.1915=0.3094

Jumlah cadangan=0.3094(1000)=309 pelamar (dibulatkan)

14. Berdasarkan pengalaman yang dimiliki, seorang sopir taksi menyatakan bahwa jumlah penumpang yang diantar sore hari rata-rata sebanyak 23.7 penumpang dengan standar deviasi 4.2. Dengan asumsi penumpang taksi terdistribusi secara normal, berapa probabilitas sopir taksi tersebut mengantar 20 penumpang pada sore hari.

    Jawab:

    $\mu = 23.7 \qquad \sigma = 4.2$

    Digunakan pendekatan kurva normal

    Probabilitas mengantarkan 20 penumpang:

    $$Z_1 = \frac{19.5 - 23.7}{4.2} = -1.00$$

    Luasnya adalah 0.3413

    $$Z_2 = \frac{20.5 - 23.7}{4.2} = -0.76$$

    Luasnya adalah 0.2764

    Luas 19.5-20.5=0.3413-0.2764=0.0649

15. Berdasarkan soal (14), berapa probabilitas sopir taksi tersebut mengantar paling sedikit 18 penumpang pada sore hari.

    Jawab:

    $\mu = 23.7 \qquad \sigma = 4.2$

    Digunakan pendekatan kurva normal

    Probabilitas paling sedikit 18 penumpang:

    $$Z = \frac{17.5 - 23.7}{4.2} = -1.48$$

    Luasnya adalah 0.4306

    Luas 17.5 ke kanan= 0.5 + 0.4306= 0.9306

16. Berdasarkan soal (14), berapa probabilitas sopir taksi tersebut mengantar paling banyak 25 penumpang pada sore hari.

    Jawab:

$\mu = 23.7 \qquad \sigma = 4.2$

Digunakan pendekatan kurva normal

Probabilitas paling banyak 25 penumpang:

$$Z = \frac{25.5 - 23.7}{4.2} = -0.43$$

Luasnya adalah 0.1664

Luas 25.5 ke kiri= 0.5 + 0.1664= 0.6664

17. Berdasarkan soal (14), berapa probabilitas sopir taksi tersebut mengantar 15 sampai 25 penumpang pada sore hari.

Jawab:

$\mu = 23.7 \qquad \sigma = 4.2$

Digunakan pendekatan kurva normal

Probabilitas 15 sampai 24 penumpang:

$$Z_1 = \frac{14.5 - 23.7}{4.2} = -2.19$$

Luasnya adalah 0.4857

$$Z_2 = \frac{24.5 - 23.7}{4.2} = -0.19$$

Luasnya adalah 0.0753

Luas 14.5 sampai dengan 24.5= 0.4857 + 0.0753= 0.5610

18. Saat ini, seluruh wilayah Indonesia sangat rawan dengan gempa. Jika frekuensi gempa besar dalam setahun di seluruh Indonesia mengikuti distribusi normal dengan rata-rata 20.8 dan standar deviasi 4.5, hitunglah probabilitas terjadinya 18 kali gempa besar dalam satu tahun tertentu.

Jawab:

$\mu = 20.8 \qquad \sigma = 4.5$

Digunakan pendekatan kurva normal

Probabilitas terjadi 25 kali gempa bumi:

$$Z_1 = \frac{25.5 - 20.8}{4.5} = 1.04$$

Luasnya adalah 0.3508

$$Z_2 = \frac{24.5 - 20.8}{4.5} = 0.82$$

Luasnya adalah 0.2939

Luas 24.5 sampai dengan 25.5= 0.3508 -0.2939= 0.0569

19. Berdasarkan soal (18), hitunglah probabilitas terjadi paling sedikit 22 kali gempa besar.

<u>Jawab:</u>

$\mu = 20.8 \qquad \sigma = 4.5$

Digunakan pendekatan kurva normal

Probabilitas paling sedikit terjadi 22 kali gempa bumi:

$$Z = \frac{21.5 - 20.8}{4.5} = 0.16$$

Luasnya adalah 0.0636

Luas 21.5 ke kanan adalah

= 0.5-0.0636= 0.4364

20. Berdasarkan soal (18), hitunglah probabilitas terjadi paling sedikit 20 sampai 25 kali gempa besar.

<u>Jawab:</u>

$\mu = 20.8 \qquad \sigma = 4.5$

Digunakan pendekatan kurva normal

Probabilitas dari 20 sampai 25 kali gempa bumi adalah:

$$Z_1 = \frac{19.5 - 20.8}{4.5} = -0.29$$

Luasnya adalah 0.1141

$$Z_2 = \frac{25.5 - 20.8}{4.5} = 1.04$$

Luasnya adalah 0.3508

Luas 19.5 sampai dengan 25.5 adalah

= 0.1141+0.3508= 0.4649

MODUL 8
DISTRIBUSI SAMPLING

Pokok Bahasan:

- Pengertian umum tentang Distribusi Sampling
- Distribusi Sampling tentang Rerata
- Distribusi Sampling tentang Proporsi

8.1 Pengertian umum tentang Distribusi Sampling

Distribusi sampling adalah topik yang sangat penting dalam statistik inferensi. Karena pengambilan sampel merupakan salah satu hal yang memberikan kontribusi besar dalam pengambilan keputusan yang dibuat. Singkatnya**, sampling dilakukan karena ada keterbatasan waktu dan dana untuk setiap studi**. Sebagai contoh, dampak kenaikan harga BBM terhadap masyarakat kecil. Di sini perlu dilakukan definisi secara jelas siapa yang menjadi target dari “masyarakat kecil” tersebut. Melalui indikator yang dibuat dapat dilakukan spesifikasi “masyarakat kecil” yang sesuai dengan kebutuhan studi, selanjutnya dapat dilakukan sampling.

Sebagai contoh, diambil 100 sampel dari 1000 populasi, yaitu wanita pekerja yang tidak memiliki pembantu RT di distrik A. Untuk populasi yang sangat besar perlu dilakukan sampling, dari sampling tersebut, dihitung berapakah rata-rata sampel dan standar deviasi sampel. Untuk sampel yang berbeda, diperoleh rata-rata dan standar deviasi yang berbeda. **Distribusi probabilitas untuk seluruh sampel yang memungkinkan disebut dengan distribusi rata-rata sampel atau dikenal juga dengan distribusi sampling rata-rata.**

8.2 Distribusi Sampling tentang Rerata

Suatu variabel random, dalam hal ini adalah rata-rata sampel $\bar{x}$, memiliki distribusi normal dengan rata-rata $\mu_{\bar{x}}$ dan standar deviasi $\sigma_{\bar{x}}$. Atau dapat ditulis $\bar{x} \sim N(\mu_{\bar{x}}, \sigma_{\bar{x}})$.

Variabel random standar Z dapat dirumuskan dengan

$$Z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} =$$

<u>Contoh 1</u>

Perusahaan kecap “Manis Asin” ingin menduga rata-rata penjualan per bulan berdasarkan rata-rata sampel yang dilakukan selama 10 tahun 1 bulan (121 bulan). Misalnya, rata-rata penjualan sebenarnya adalah 6200 dengan standar deviasi 770. Berapa banyak bulan dari sampel tersebut yang memiliki rata-rata penjualan antara 6150 dan 6240?

Jawab:

(i) Berdasarkan *Central Limit Theorem*, diketahui

$\mu = \mu_{\bar{x}} = 6200$

(ii) Untuk populasi yang infinit $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{770}{\sqrt{121}} = 70$

(note: 10 tahun 1 bulan =121 bulan)

Jika $\bar{x} = 6150,$ maka $Z = \frac{6150 - 6200}{70} = -0,71$

Jika $\bar{x} = 6240,$ maka $Z = \frac{6240 - 6200}{70} = 0,57$

Sehingga $P(6150 < x < 6240) = P(-0,71 < X < 0,57)$

$= 0,2611 + 0,2157$

$= 0,4768$

Jadi, terdapat 0,4768 x 121 = 57 bulan dengan rata-rata sampel antara 6150 sampai 6240.

Mengacu pada soal di atas, ada beberapa aturan yang perlu diketahui dalam *Central Limit Theorem*, yaitu:

- Rata-rata distribusi sampling akan sama dengan rata-rata populasi

  $\mu_{\bar{x}} = \mu$.

- Standar deviasi rata-rata $(\sigma_{\bar{x}})$ adalah:

(i) $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ → untuk populasi infinite

(ii) $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N - n}{N - 1}}$ → untuk populasi finite dan n < 0,05N

- Distribusi sampling rata-rata akan mendekati normal jika ukuran sampel cukup besar $(n \geq 30)$, apapun bentuk distribusi populasinya.
- Distribusi sampling rata-rata akan normal jika distribusi populasi normal, berapapun ukuran sampelnya.

8.3 Distribusi Sampling tentang Proporsi

Proporsi sampling akan memiliki distribusi mendekati normal jika nP ataupun $n(1 - P)$ lebih besar atau sama dengan 5. Distribusi sampling proporsi ini dapat ditulis $P \sim N(P, \sigma_p)$. Variabel random standar Z dapat dirumuskan dengan

$$Z = \frac{p - \mu_p}{\sigma_p} = \frac{p - P}{\sigma_p} \Rightarrow \mu_p = P$$

Proporsi merupakan variabel random diskrit, sehingga pendekatan distribusi normal membutuhkan faktor koreksi kontinuitas yang besarnya $\frac{1}{2n}$, sehingga nilai variabel random standar Z adalah

$$Z = \frac{(p - \frac{1}{2n}) - P}{\sigma_p}$$

Contoh 2

Perusahaan mie "Sangat Lezat" memiliki data bahwa dari mesin A, ada 2% produk mie yang tidak bagus. Untuk produksi sebanyak 400 unit, berapa probabilitas 3% atau lebih dari produk mie tersebut tidak bagus

Jawab:

$\mu_p = P = 0{,}02$

$$\sigma_p = \sqrt{\frac{P(1 - P)}{n}} = \sqrt{\frac{0{,}02x0{,}98}{400}} = 0{,}007$$

Untuk 3% atau lebih dari produk mie tersebut tidak bagus, artinya $P = 3\%$

$$Z = \frac{(p - \frac{1}{2n}) - P}{\sigma_p} = \frac{(0{,}03 - \frac{1}{800}) - 0{,}02}{0{,}07} = 1{,}25$$

Sehingga, untuk $P(p \geq 0{,}03) = P(Z \geq 1{,}25) = 0{,}5 - 0{,}3944 = 0{,}1056$

Soal tersebut juga bisa diselesaikan dengan menggunakan pendekatam distribusi binomial.

Diketahui: $P = 2\%$ dan $n = 400$

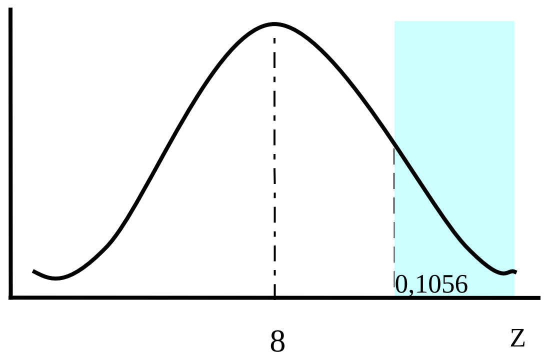
$\mu = np = 2\% x 400 = 8$

$\sigma = \sqrt{npq} = \sqrt{400x2\%x98\%} = \sqrt{7{,}84} = 2{,}8$

Untuk 3% atau lebih dari produk mie tersebut tidak bagus, artinya $3\% x 400 = 12$

$$Z = \frac{11{,}5 - 8}{2{,}8} = 1{,}25$$ → luasnya adalah 0,3944

Jadi luas daerah 11,5 ke kanan (untuk $p \geq 3\%$)adalah 0,5 - 0,3944 = 0,1056.



Note: Rumus Standar Deviasi adalah

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

Jika populasi infinite (*sampling with replacement)* atau populasi finite dengan $n < 5\%$.

(ii) $\sigma_p = \sqrt{\frac{P(1-P)}{n}} x \sqrt{\frac{N-n}{N-1}}$

Jika populasi finite (*sampling without replacement)*

8.4 Latihan: Soal & Jawab

1. Sebuah perlombaan pengusaha muda "Cerdas dan Smart" diambil 300 juri yang terdiri dari 150 orang akademisi, 100 orang praktisi bisnis dan 50 orang praktisi perbankan. Dari 300 orang tersebut diambil *stratified sampling* sebesar 20%. Bagaimana penentuan jumlah/besar sampel setiap strata jika 1/3 dari sampel dialokasikan untuk masing-masing strata.

   Jawab:

   $N = 300 \quad N_1 = 150 \quad N_2 = 100 \quad N_3 = 50$

   $n = 20\%(300) = 60$

   $n_1 = (1/3)(60) = 20$

   $n_2 = (1/3)(60) = 20$

   $n_3 = (1/3)(60) = 20$

   Jadi total n =60

2. Berdasarkan soal (1), bagaimana penentuan jumlah/besar sampel setiap strata jika alokasi untuk masing-masing strata dilakukan secara proporsional.

   Jawab:

   $N = 300 \quad N_1 = 150 \quad N_2 = 100 \quad N_3 = 50$

   Dengan alokasi yang proporsional

   $Rumus : n_i = \frac{N_i}{N} g n$

   $n_1 = (150/300)(60) = 30$

   $n_2 = (100/300)(60) = 20$

   $n_3 = (50/300)(60) = 10$

   Jadi total n =60

3. Di sebuah survei makanan, dilakukan *stratified sampling* dengan ukuran n=80 diambil dari total populasi yang memiliki ukuran N=2000. Populasi tersebut memiliki 4 strata yang terdiri dari N_1= 500, N_2=1200, N_3=200, dan N_4=100. Apabila digunakan alokasi yang proporsional, berapa sampel yang harus dialokasikan untuk setiap strata.

   Jawab:

   $N = 2000 \quad N_1 = 500 \quad N_2 = 1200 \quad N_3 = 200 \quad N_4 = 100 \quad n=80$

   Dengan alokasi yang proporsional

   $Rumus : n_i = \frac{N_i}{N} g n$

   $n_1 = (500/2000)(80) = 20$

   $n_2 = (1200/2000)(80) = 48$

   $n_3 = (200/2000)(80) = 8$

   $n_4 = (100/2000)(80) = 4$

   Jadi total n =80

4. Diketahui suatu populasi bisa dibedakan dalam 2 strata yang terdiri dari N1=10000 dan N2= 30000 dengan standar deviasi masing-masing adalah 45 dan 60. Bagaimana sampel dengan ukuran 100 dialokasikan pada strata tersebut jika digunakan alokasi optimum.

   Jawab:

$N_1 = 10000 \quad N_2 = 30000$

$\sigma_1 = 45 \qquad \sigma_2 = 60$

Dengan alokasi yang optimum

$$Rumus: n_i = \frac{n \cdot N_i \cdot \sigma_i}{N_1\sigma_1 + N_2\sigma_2}$$

$$n_1 = \frac{100(10000)(45)}{(45)(10000) + 60(30000)} = 20$$

$$n_2 = \frac{100(30000)(60)}{(45)(10000) + 60(30000)} = 80$$

Total jumlah n=100

5. Diketahui suatu populasi dibagi dalam 3 strata yaitu N1=5000, N2=2000 dan N3=3000 dengan standar deviasi masing-masing strata sebesar 15, 18 dan 5. Bagaimana sampel berjumlah 85 dialokasikan pada 3 strata bila digunakan alokasi optimum.

Jawab:

$N_1 = 5000 \quad N_2 = 2000 \quad N_3 = 3000$

$\sigma_1 = 15 \qquad \sigma_2 = 18 \qquad \sigma_3 = 5$

n=84

Dengan alokasi yang optimum

$$n_1 = \frac{84(5000)(15)}{(15)(5000) + 18(2000) + 5(3000)} = 50$$

$$n_2 = \frac{84(2000)(18)}{(15)(5000) + 18(2000) + 5(3000)} = 24$$

$$n_3 = \frac{84(3000)(5)}{(15)(5000) + 18(2000) + 5(3000)} = 10$$

Total jumlah n=84

6. Dalam periode tertentu, sebuah perusahaan asuransi "Semoga Selamat" menerima klaim sebanyak 3800 klaim. Dari jumlah tersebut, 2600 merupakan klaim ringan dan 1200 klaim berat. Klaim ringan

memiliki nominal di bawah Rp 2 juta dan klaim berat memiliki nominal di atas Rp 2 juta. Untuk menduga rata-rata besarnya klaim, perusahaan mengambil sampel sebnayak 1% dengan alokasi yang proporsional. Hasilnya adalah sebagai berikut (dalam puluhan ribu rupiah):

<u>Klaim ringan:</u>

42, 115, 63, 78, 45, 148, 195, 66, 18, 73, 55, 89, 170, 41, 92, 103, 22, 138, 49, 62, 88, 113, 29, 71, 58 dan 83

<u>Klaim berat:</u>

246, 355, 872, 649, 253, 338, 491, 869, 755, 502, 488 dan 311

Hitunglah rata-rata dari dua sampel tersebut. Selanjutnya hitunglah rata-rata tertimbang. Sebagai pembobot (timbangan), gunakan ukuran dari kedua strata tersebut.

<u>Jawab:</u>

Rata-rata dari sampel klaim ringan:

$$\bar{x}_1 = \frac{42+115+63+...+83}{26}$$

$$= \frac{2106}{26} = 81$$

Rata-rata dari sampel klaim berat:

$$\bar{x}_2 = \frac{246+355+872+...+311}{12}$$

$$= \frac{6120}{12} = 510$$

Rata-rata Tertimbang:

$$\bar{x}_w = \frac{n_1 x_1 + n_2 x_2}{n_1 + n_2}$$

$$\bar{x}_w = \frac{2600(81)+1200(510)}{2600+1200}$$

$$\bar{x}_w = \frac{822600}{3800} = 216.47$$

7. Berdasarkan soal (6), apakah hasil perhitungan pada no. 6 di atas sama dengan nilai rata-rata dari 38 klaim (untuk sampel secara keseluruhan).

   Jawab:

   Rata-rata dari sampel secara keseluruhan:

$$\bar{x} = \frac{42+115+63+...+488+311}{38}$$

$$= \frac{8226}{38} = 216.47$$

   Berdasarkan hasil di atas, dapat dinyatakan bahwa rata-rata dari sampel secara keseluruhan sama dengan rata-rata tertimbang.

8. Sebuah populasi terdiri dari 5 angka, yaitu 2, 3, 6, 8, dan 11. Dari populasi tersebut ditarik sampel yang beranggotakan 2 dengan pengembalian yang mungkin dapat diambil dari populasi tersebut. Hitunglah rata-rata populasi tersebut.

   Jawab:

$$\mu = \frac{\sum X}{N} = \frac{2+3+6+8+11}{5} = \frac{30}{5} = 6$$

9. Berdasarkan soal (8), hitunglah standar deviasi dari populasi tersebut.

   Jawab:

$$\sigma = \sqrt{\frac{\sum (X-U)^2}{N}} = \sqrt{\frac{(2-6)^2+(3-6)^2+(6-6)^2+(8-6)^2+(11-6)^2}{5}}$$

$$= \sqrt{\frac{16+9+0+4+25}{5}} = \sqrt{10.8} = 3.29$$

10. Berdasarkan soal (8), hitunglah rata-rata dan standar deviasi dari distribusi sampling harga rata-rata.

   Jawab:

   Ada 25 sampel (5^2), yakni:

(2,2)	(2,3)	(2,6)	(2,8)	(2,11)
(3,2)	(3,3)	(3,6)	(3,8)	(3,11)
(6,2)	(6,3)	(6,6)	(6,8)	(6,11)
(8,2)	(8,3)	(8,6)	(8,8)	(8,11)
(11,2)	(11,3)	(11,6)	(11,8)	(11,11)

   Rata-rata dari masing-masing sampel adalah:

$$\mu_{\bar{x}} = \frac{\sum X}{N} = \frac{2+2.5+4+...+11}{25} = \frac{150}{25} = 6$$

$$\mu = \mu_{\bar{x}} = 6$$

$$\sigma = \sqrt{\frac{\sum(\bar{X}_1-\mu_{\bar{x}})^2}{25}} = \sqrt{\frac{(2-6)^2+(2.5-6)^2+...+(11-6)^2}{25}}$$

$$= \sqrt{\frac{135}{25}} = 2.32$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3.29}{\sqrt{2}} = 2.32$$

11. Diasumsikan bahwa tinggi badan dari 3000 siswa di sebuah sekolah memiliki distribusi normal dengan rata-rata 68 cm dan standar deviasi 3 cm. Apabila diambil sampel 80 siswa dengan ukuran n=25 diambil dari populasi tersebut, berapa besarnya rata-rata dan standar deviasi yang diharapkan berasal dari distribusi sampling nilai rata-rata jika sampling ditarik dengan pengembalian.

Jawab:

$$\mu_{\bar{x}} = \mu = 68$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{35}} = 0.6$$

12. Berdasarkan soal (11), jika sampling ditarik tanpa pengembalian.

Jawab:

$$\mu_{\bar{x}} = \mu = 68$$

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

$$= \frac{3}{\sqrt{25}} \sqrt{\frac{3000-25}{3000-1}}$$

jauh lebih kecil dari 0.6

13. Berdasarkan soal (11), berapa banyak sampel bila diharapkan bahwa nilai rata-rata terletak antara 66.8 dan 68.3 cm.

Jawab:

$$z = \frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

$$z_1 = \frac{66.8 - 68}{0.6} = -2$$

Luasnya adalah 0.4772

$$z_2 = \frac{68.3 - 68}{0.6} = 0.5$$

Luasnya adalah 0.1915

Luas total =0.4772+0.1915=0.6687

Banyak sampel yang diharapkan =0.6687(80)=53

14. Berdasarkan soal (11), berapa banyak sampel bila diharapkan bahwa nilai rata-rata kurang dari 66.4 cm.

Jawab:

$$z = \frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

$$z = \frac{66.4 - 68}{0.6} = -2.67$$

Luasnya adalah 0.4962

Luas daerah 66.4 ke kiri adalah 0.5- 0.4962=0.0038

15. Dari total produk yang dihasilkan oleh mesin A, 2%-nya adalah rusak. Berapa probabilitas dari pengiriman 400 produk, 3% atau lebih ternyata rusak.

<u>Jawab:</u>

$\mu = np = 2\%(400) = 8$

$\sigma = \sqrt{npq} = \sqrt{2\%(98\%)(400)} = 2.8$

$z = \dfrac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$

$z = \dfrac{11.5 - 8}{2.8} = 1.25$

Luasnya adalah 0.3944

Luas daerah 11.5 ke kanan adalah 0.5- 0.3944=0.01056

16. Berdasarkan soal (15), berapa probabilitas 2% atau kurang dari total 400 produk adalah rusak.

<u>Jawab:</u>

$\mu = np = 2\%(400) = 8$

$\sigma = \sqrt{npq} = \sqrt{2\%(98\%)(400)} = 2.8$

$z = \dfrac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$

$z = \dfrac{8.5 - 8}{2.8} = 0.18$

Luasnya adalah 0.0714

Luas daerah 8.5 ke kiri adalah 0.5- 0.714=0.5714

17. Dalam pemilihan walikota, diketahui bapak A mendapat 46% suara. Berapa probabilitas jika dari 200 penduduk yang dipilih secara random dari total pemilih memberikan suara untuk bapak A.

Jawab:

$\mu = P = 0.46$

$$\sigma_p = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.46(0.54)}{200}} = 0.0352$$

$$z = \frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

$$z = \frac{0.5025 - 0.46}{0.0352} = 1.21$$

Luasnya adalah 0.3869

Luas daerah 0.5025 ke kanan adalah 0.5- 0.3869=0.1131

18. Berdasarkan soal (17), berapa probabilitas jika dari 1000 penduduk yang dipilih secara random dari total pemilih memberikan suara untuk bapak A.

Jawab:

$\mu = P = 0.46$

$n = 1000$

$$\sigma_p = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.46(0.54)}{1000}} = 0.0158$$

$$z = \frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

$$z = \frac{0.5025 - 0.46}{0.0158} = 2.69$$

Luasnya adalah 0.4964

Besarnya probabilitas adalah 0.5- 0.4964=0.0036

19. Bola lampu "Sangat Terang" mempunyai daya nyala rata-rata 1500 jam dengan standar deviasi 150 jam. Tiga bola lampu dihubungkan, apabila yang satu menyala maka yang lain akan menyala. Dengan asumsi bahwa daya nyala lampu terdistribusi secara normal, berapa probabilitas daya nyala nyampu tersebut paling sedikit 5000 jam.

Jawab:

$$\mu_{L1+L2+L3} = \mu_{L1} + \mu_{L2} + \mu_{L3}$$

$$\sigma_{L1+L2+L3} = \sqrt{\sigma^2{}_{L1} + \sigma^2{}_{L2} + \sigma^2{}_{L3}}$$

$$\sigma_p = \sqrt{3(150)^2} = 260 \text{ jam}$$

$$z = \frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

$$z = \frac{5000 - 4500}{260} = 1.92$$

Luasnya adalah 0.4726

Luas 5000 atau lebih adalah 0.5- 0.4726=0.274

20. Berdasarkan soal (19), berapa probabilitas daya nyala nyampu tersebut paling banyak 4200 jam.

Jawab:

$$\mu_{L1+L2+L3} = \mu_{L1} + \mu_{L2} + \mu_{L3}$$

$$\sigma_{L1+L2+L3} = \sqrt{\sigma^2{}_{L1} + \sigma^2{}_{L2} + \sigma^2{}_{L3}}$$

$$\sigma_p = \sqrt{3(150)^2} = 260 \text{ jam}$$

$$z = \frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

$$z = \frac{4200 - 4500}{260} = -1.15$$

Luasnya adalah 0.3749

Luas 4200 atau kurang adalah 0.5- 0.3749=0.1251

MODUL 9
STATISTIK INFERENSI: PENDUGAAN INTERVAL I-III

<u>Pokok Bahasan:</u>

- Pengertian Inferens: Pendugaan Interval dan Pengujian Hipotesis
- Ciri-ciri Penduga yang Baik
- Pendugaan Interval Rerata Populasi
- Pendugaan Interval <u>Selisih Dua Rerata</u> Populasi
- Pendugaan Interval Proporsi Populasi
- Pendugaan Interval Selisih Dua Proporsi Populasi
- Penentuan *Sample Size*

9.1 Pengertian Inferens: Pendugaan Interval dan Pengujian Hipotesis

Mengacu pada pengantar tentang statistika inferensia pada bab pertama, dalam pokok bahasan ini akan dibahas lebih spesifik beberapa topik penting berkaitan dengan statistika inferensia, seperti metode sampling dan distribusi sampling, teori pendugaan statistik dan pengujian hipotesis.

Melalui statistik inferensia, karakteristik populasi dapat "dipotret" melalui karakteristik sampel random. Di sini, sampel yang random adalah syarat penting yang harus dipenuhi agar kesimpulan yang diambil dari sampel bisa diterima sebagai kesimpulan populasi. Dan faktor kerandoman sampel sangat ditentukan oleh metode sampling dan distribusi sampling.

9.2 Pendugaan Interval Rerata Populasi

Jika suatu populasi normal memiliki distribusi sampling rata-rata mendekati normal dan standar deviasi populasi (σ) diketahui, maka standar deviasi rata-rata $(\sigma_{\bar{x}})$ juga dapat diketahui. Rumus pendugaan interval rata-rata populasi dibedakan untuk (σ) diketahui dan (σ) tidak diketahui, yaitu:

(i) Untuk (σ) diketahui

$$\Pr ob(\bar{x} - Z_{\alpha/2}\sigma_{\bar{x}} \leq \mu \leq \bar{x} + Z_{\alpha/2}\sigma_{\bar{x}} = C \quad \text{atau} \quad \Pr ob(\bar{x} \pm Z_{\alpha/2}\sigma_{\bar{x}}) = C$$

(ii) Untuk (σ) tidak diketahui

$$\Pr ob(\bar{x} - Z_{\alpha/2}\frac{S}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{\alpha/2}\frac{S}{\sqrt{n}} = C$$

Keterangan:

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

$\sigma =$ standar deviasi populasi

$\sigma_{\bar{X}} =$ standar deviasi rata-rata populasi

S =standar deviasi sampel

$S_{\bar{X}} =$ standar deviasi rata-rata sampel

$n =$ ukuran sampel

Untuk standar deviasi populasi diketahui, perhitungan standar deviasi adalah

(i) $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ untuk populasi infinite

(ii) $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ untuk populasi finite dan n < 0,05N

Untuk standar deviasi populasi tidak diketahui, perhitungan standar deviasi populasi diduga dengan standar deviasi sampel, yaitu

(i) $S = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$ untuk populasi infinite

(ii) $S_{\bar{x}} = \frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ untuk populasi finite dan n < 0,05N

Karena $\sigma_{\bar{x}}$ diduga dengan $S_{\bar{X}}$, maka variabel random $(\bar{x}-\mu)/S_{\bar{X}}$ tidak mengikuti distribusi normal. Variabel random $(\bar{x}-\mu)/S_{\bar{X}}$ mempunyai distribusi standar t. Distribusi t mirip dengan distribusi Z, hanya saja distribusi t tergantung pada ukuran sampel (n). Jika ukuran sampel semakin mendekati 30, bentuk distribusi t akan mendekati bentuk distribusi Z.

Jika standar deviasi populasi tidak diketahui, standar deviasi dihitung dari sampel dan $Z_{\alpha/2}$ diganti dengan $t_{\alpha/2}, v$. Dalam hal ini *v* menyatakan *degree* *of freedom* dimana nilai *v* diperoleh dengan rumus *v*=*n*-*k.* Jumlah observasi dinyatakan dengan *n* dan jumlah variabel bebas dinyatakan dengan *k*. Bentuk interval keyakinan dari distribusi t adalah

$$\Pr ob[\bar{x} - t_{\alpha/2}, v(\frac{S}{\sqrt{n}}) \leq \mu \leq \bar{x} + t_{\alpha/2}, v(\frac{S}{\sqrt{n}})] = C$$

9.3 Pendugaan Interval Selisih Dua Rerata Populasi

Interval Selisih Dua Rerata. Jika $\bar{x}_1$ dan $\bar{x}_2$ adalah rata-rata dua sampel dari dua populasi dengan distribusi sampling mendekati normal, pendugaan interval selisih rata-rata dua populasi dirumuskan dengan:

$$\Pr ob(\bar{x}_1 - \bar{x}_2) - Z_{\alpha/2}\sigma_{\bar{x}_1 - \bar{x}_2} \leq \mu_1 - \mu_2 \leq (\bar{x}_1 - \bar{x}_2) + Z_{\alpha/2}\sigma_{\bar{x}_1 - \bar{x}_2} = C$$

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

Keterangan:

$\bar{x}_1, \bar{x}_2$ = rata-rata sampel dari dua populasi
$\sigma_{\bar{x}_1 - \bar{x}_2}$ = standar deviasi selisih rata-rata
σ_1, σ_2 = standar deviasi dua populasi
S_1, S_2 =standar deviasi dua sampel
n_1, n_2 = ukuran sampel dari dua populasi

Jika standar deviasi populasi ($\sigma_{\bar{x}_1 - \bar{x}_2}$) tidak diketahui, standar deviasi dapat diduga dengan rumus:

(i) $S_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$ untuk proporsi infinite

(ii) $S_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} x \sqrt{\frac{N - n}{N - 1}}$ untuk proporsi finite dan $n \leq 0{,}05N$

9.4 Pendugaan Interval Proporsi Populasi

Interval Proporsi. Jika ukuran sampel cukup besar, yaitu ketika nP maupun $n(1 - P)$ lebih besar 5, untuk P adalah proporsi populasi, maka distribusi sampling proporsi akan mendekati distribusi normal dengan rata-rata P.

Untuk standar deviasi (proporsi) populasi diketahui, perhitungan standar deviasi adalah

(i) $\sigma_p = \sqrt{\dfrac{P(1-P)}{n}}$ untuk populasi infinite

(ii) $\sigma_p = \sqrt{\dfrac{P(1-P)}{n}}\sqrt{\dfrac{N-n}{N-1}}$ untuk populasi finite

Karena P pada umumnya tidak diketahui, standar deviasi (proporsi) populasi (σ_p) juga tidak diketahui. Perhitungan standar deviasi (proporsi) populasi diduga dengan standar deviasi sampel (S_P), yaitu: (dimana P diganti dengan p)

(i) $S_P = \sqrt{\dfrac{p(1-p)}{n-1}}$ → untuk populasi infinite

(ii) $S_p = \sqrt{\dfrac{p(1-p)}{n-1}}\sqrt{\dfrac{N-n}{N-1}}$ → untuk populasi finite

Note:

Karena rata-rata distribusi sampling populasi sama dengan proporsi populasi, maka proporsi sampel adalah penduga tak bias terhadap proporsi populasi. Ketika nilai P tidak diketahui, digunakan p untuk mengganti P. Ukuran sampel dikatakan cukup besar jika np dan $n(1-p)$ keduanya ≥ 15.

Bentuk pendugaan proporsi populasi untuk P yang tidak diketahui adalah:

$$\Pr ob(p - Z_{\alpha/2} S_p < P < p + Z_{\alpha/2} S_p) = C$$

<u>Contoh 1</u>

Suatu populasi memiliki distribusi normal dengan standar deviasi 5. Suatu sampel random sebanyak 36 memiliki rata-rata 30. Buatlah interval keyakinan rata-rata populasi dengan tingkat keyakinan 90%.

Jawab:

$\sigma = 5; n = 36; \bar{x} = 30; C = 90\% \Rightarrow \alpha = 10\%$

Yang harus dilakukan adalah mencari standar deviasi rata-rata $\sigma_{\bar{x}}$ dan menentukan $Z_{\alpha/2}$.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{36}} = 0,8333$$

$$Z_{\alpha/2} = (Z | P = 0,05) = 1,645$$

Jadi, interval keyakinan dengan tingkat keyakinan 90% adalah:

$$30 - 1,645(0,8333) \le \mu \le 30 + 1,645(0,8333)$$
$$28,63 \le \mu \le 31,37$$

<u>Contoh 2</u>

Diketahui proporsi sampel adalah 0,07. Berapakah interval keyakinan proporsi sampel tersebut untuk C=90%?

Jawab:

$$p = 0,07; C = 90\% \Rightarrow \alpha = 10\%$$

Yang harus dilakukan adalah mencari standar deviasi proporsi σ_p dan menentukan Z_p.

$$S_p = \sqrt{\frac{0,07(0,93)}{499}} = 0,0144$$
$$Z_p = 0,5 - 0,05 = 1,645$$

Jadi, interval keyakinan dengan tingkat keyakinan 90% adalah:

$$0,07 - 1,645(0,0114) < P < 0,07 + 1,645(0,0114)$$
$$0,051 < P < 0,089$$

9.5 Pendugaan Interval Selisih Dua Proporsi Populasi

Interval Selisih Dua Proporsi. Jika p_1 dan p_2 adalah proporsi dua sampel dari dua proporsi populasi dengan distribusi sampling mendekati normal, pendugaan interval selisih dua proporsi populasi dirumuskan dengan:

$$\Pr ob(p_1 - p_2) - Z_{\alpha/2} S_{p_1 - p_2} \le P_1 - P_2 \le (p_1 - p_2) + Z_{\alpha/2} S_{p_1 - p_2} = C$$

Keterangan:

p_1, p_2 = proporsi sampel dari dua populasi

S_{p1-p2} = penduga standar deviasi selisih proporsi (σ_{p1-p2})

n_1, n_2 = ukuran sampel dari dua populasi

Rumus standar deviasi yang digunakan adalah:

(i) $S_{p1-p2} = \sqrt{\dfrac{P_1(1-P_1)}{n_1-1} + \dfrac{P_2(1-P_2)}{n_2-1}}$ untuk proporsi yang infinite

(ii) $S_{p1-p2} = \sqrt{\dfrac{P_1(1-P_1)}{n_1-1} + \dfrac{P_2(1-P_2)}{n_2-1}}\, x \sqrt{\dfrac{N-n}{N-1}}$ untuk proporsi finite dan $n \leq 0{,}05N$

Untuk aplikasi dari pendugaan interval selisih rata-rata dan selisih proporsi populasi, digunakan contoh dua populasi. Untuk selanjutnya, prosedur penyelesaian sesuai dengan rumus di atas.

9.6 Penentuan *Sample Size*

Penentuan ukuran sampel penting dilakukan karena sampel yang tidak tepat bisa mengarah pada inferensi yang tidak tepat pula. “Sembilan dari 10 wanita Indonesia menggunakan sabun “Rupa-Rupa”, adalah pernyataan yang bisa saja digunakan oleh perusahaan iklan. Pernyataan tersebut merupakan kajian yang sangat menarik untuk mengantarkan kita dalam belajar ukuran sampel.

9.7 Latihan: Soal & Jawab

1. Tingkat IQ dari 200 siswa SMA yang diambil secara random dari populasi 800 siswa SMA memiliki rata-rata 107 dan standar deviasi 12.4. Hitunglah interval konfidensi 95% untuk rata-rata IQ dari 800 siswa SMA tersebut.

   Jawab:

   $N = 800$ $\quad \overline{X} = 107$

   $n = 200$ $\quad S = 12.4$

   Digunakan penyesuaian $= \sqrt{\dfrac{N-n}{N-1}}$

   Interval konfidensi 95 untuk rata-rata IQ dari 800 murid SLTA di kota A:

Rumus:

$$\bar{X} - 1.96\frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} < \mu < \bar{X} + 1.96\frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

$$107 - 1.96\frac{12.4}{\sqrt{200}}\sqrt{\frac{800-200}{800-1}} < \mu < 107 + 1.96\frac{12.4}{\sqrt{200}}\sqrt{\frac{800-200}{800-1}}$$

$$105.5 < \mu < 108.5$$

2. Dari *stratified random sampling* dengan ukuran n=33 dialokasikan secara proporsional ke dalam 4 strata dari populasi yang terbatas yaitu N1= 200, N2=600, N3= 1000, N=400 dengan standar deviasi masing-masing strata sebesar 2, 2.5, 1.5 dan 3. Jika rata-rata dari sampel digunakan untuk menduga rata-rata populasi, apa yang dapat diperkirakan tentang besarnya kesalahan maksimum dengan prob. 95%.

   Jawab:

   $$E < Z_{\frac{\alpha}{2}}.\sigma_{\bar{x}}$$

   $$E < 1.96(0.37)$$

   $$E < 0.73$$

3. Berdasarkan soal (2), jika rata-rata dari sampel tersebut adalah 117.39 cm, hitunglah interval konfidensi 99% untuk rata-rata populasi.

   Jawab:

   $$\bar{X} - Z_{\frac{\alpha}{2}}\sigma_{\bar{x}} < \mu < \bar{X} + Z_{\frac{\alpha}{2}}\sigma_{\bar{x}}$$

   $$117.39 - 2.58(0.37) < \mu < 117.39 + 2.58(0.37)$$

   $$116.44 \ \ cm < \mu < 118.34 \ \ cm$$

4. Dari sampel random sebanyak 250 murid SMA, 165 diantaranya ingin melanjutkan ke perguruan tinggi. Hitunglah interval konfidensi 99% untuk proporsi yang sesunguhnya.

   Jawab:

   $$\frac{165}{250} - 2.58\sqrt{\frac{\frac{165}{250}(1-\frac{165}{250})}{250}} < P < \frac{165}{250} + 2.58\sqrt{\frac{\frac{165}{250}(1-\frac{165}{250})}{250}}$$

   $$0.66 - 0.08 < P < 0.66 + 0.08$$

   $$0.58 < P < 0.74$$

$$\frac{X}{n} - Z_{\frac{\alpha}{2}} \sqrt{\frac{\frac{X}{n}(1-\frac{X}{n})}{n}} < P < \frac{X}{n} + Z_{\frac{\alpha}{2}} \sqrt{\frac{\frac{X}{n}(1-\frac{X}{n})}{n}}$$

5. Diambil sampel random 100 butir telur dari total 1000 telur yang akan dikirim dari kota A ke B. Dari sampel tersebut ada 18 butir telur yang rusak. Hitunglah interval konfidensi 95% untuk telur yang rusak dari 1000 telur tersebut.

   Jawab:

$$\frac{X}{n} - Z_{\frac{\alpha}{2}} \sqrt{\frac{\frac{X}{n}(1-\frac{X}{n})}{n}} \sqrt{\frac{N-n}{N-1}} < P < \frac{X}{n} + Z_{\frac{\alpha}{2}} \sqrt{\frac{\frac{X}{n}(1-\frac{X}{n})}{n}} \sqrt{\frac{N-n}{N-1}}$$

$$\frac{18}{100} - 1.96 \sqrt{\frac{\frac{18}{100}(1-\frac{18}{100})}{100}} \sqrt{\frac{1000-100}{1000-1}} < P < \frac{18}{100} - 1.96 \sqrt{\frac{\frac{18}{100}(1-\frac{18}{100})}{100}} \sqrt{\frac{1000-100}{1000-1}}$$

$$0.18 - 0.07 < P < 0.18 + 0.07$$

$$0.11 < P < 0.25$$

6. Dari 600 wanita yang lulus PT ”Terbaik”, 378 diantaranya ingin bekerja di kantor pemerintah. Jika 378/600 digunakan sebagai penduga proporsi yang sesungguhnya, apa yang dapat dikatakan dengan probabilitas 95% tentang besarnya maksimum *error*.

   Jawab:

$$E < 1.96 \sqrt{\frac{\frac{378}{600}(1-\frac{378}{600})}{600}}$$

$$E < 0.039$$

   Besarnya maksimum error kurang dari 0.039

7. Dari sebuah riset Lembaga Demografi “Tepat Sekali”, 6% dari sampel random sebanyak 1800 keluarga di pulau Jawa ingin bertransmigrasi ke Sumatera. Hitunglah interval konfidensi 90% untuk memperkirakan jumlah keluarga yang akan bertransmigrasi untuk populasi berjumlah 20 juta keluarga.

   Jawab:

$$0.06 - 1.64\sqrt{\frac{0.06(0.94)}{1800}} < \mu < 0.06 + 1.64\sqrt{\frac{0.06(0.94)}{1800}}$$

$$0.06 - 0.009 < P < 0.06 + 0.009$$

$$0.051 < P < 0.06$$

Dari populasi N= 20 jt keluarga, jumlah yang akan bertransmigrasi diperkirakan antara 0.051(20 jt) sampai 0.069(20 jt) atau antara 1.02 jt sampai 1.38 jt orang

8. Seorang pengusaha sudah memperkirakan bahwa para pekerja tidak akan setuju dengan rencana pengurangan upah sebesar 5%. Jika diambil sampel random 60 orang dari 300 orang pekerja dan hanya 18 orang yang bersedia menerima pengurangan upah, hitunglah interval konfidensi 95% untuk proporsi buruh yang bersedia menerima pengurangan upah tersebut.

   Jawab:

   $N = 300$ X=18

   $n = 60$

   Rumus:

$$\frac{X}{n} - 1.96\frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\sqrt{\frac{\text{N-n}}{\text{N-1}}} < \mu < \frac{X}{n} + 1.96\frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\sqrt{\frac{\text{N-n}}{\text{N-1}}}$$

$$\frac{18}{60} - 1.96\frac{12.4}{\sqrt{200}}\sqrt{\frac{\frac{18}{60}(1-\frac{18}{60})}{60}}\sqrt{\frac{300\text{-}60}{300\text{-}1}} < \mu < \frac{18}{60} + 1.96\frac{12.4}{\sqrt{200}}\sqrt{\frac{\frac{18}{60}(1-\frac{18}{60})}{60}}\sqrt{\frac{300\text{-}60}{300\text{-}1}}$$

$$0.30 - 0.10 < P < 0.30 + 0.10$$

$$0.2 < P < 0.4$$

9. Dari sampel random sebanyak 150 buah bola lampu merek A, diketahui daya hidup rata-rata 1400 jam dengan standar deviasi 120 jam. Sedangkan untuk bola lampu merek B, diketahui daya hidup rata-rata 1200 jam dengan standar deviasi 80 jam. Hitunglah interval konfidensi 95% untuk perbedaan rata-rata daya hidup dari kedua merek bola lampu tersebut.

   Jawab:

$$(\bar{X}_1-\bar{X}_2)-Z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1}{n_1}+\frac{\sigma_2}{n_2}}<(\mu_1-\mu_2)<(\bar{X}_1-\bar{X}_2)+Z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1}{n_1}+\frac{\sigma_2}{n_2}}$$

$$(1400-1200)-1.96\sqrt{\frac{120}{150}+\frac{80}{200}}<\mu_1-\mu_2<(1400-1200)+1.96\sqrt{\frac{120}{150}+\frac{80}{200}}$$

$$177.83 \text{ jam} < P < 222.17 \text{ jam}$$

10. Berdasarkan catatan badan meteorologi diperoleh informasi bahwa selama 15 tahun, rata-rata curah hujan di regioan A adalah 4.93 cm dengan standar deviasi 1.14 cm. Sedangkan di region B, berdasarkan catatan selama 10 tahun, rata-rata curah hujan adalah 2.64 cm dengan standar deviasi 0.66 cm. Hitunglah interval konfidensi 95% untuk perbedaan curah hujan rata-rata yang sesungguhnya dari dua region tersebut. Diasumsikan observasi berasal dari populasi normal dengan varians yang berbeda dan tidak diketahui.

    Jawab:

    $$2.29-2.069\sqrt{\frac{1.14^2}{15}+\frac{0.66^2}{10}}<(\mu_1-\mu_2)<2.29+2.069\sqrt{\frac{1.14^2}{15}+\frac{0.66^2}{10}}$$

    $$1.54 < (\mu_1-\mu_2) < 3.04$$

11. Dari sampel random sebanyak 400 pemirsa dewasa dan 600 pemirsa remaja yang menonton program TV "Cerdas Tangkas", diketahui 100 pemirsa dewasa dan 300 pemirsa remaja menyukai acara tersebut. Hitunglah interval konfidensi 95% untuk perbedaan proporsi pemirsa dewasa dan remaja yang menyukai program tersebut.

    Jawab:

    $$(0.05-0.025)-1.96\sqrt{\frac{(0.5)(0.5)}{600}+\frac{(0.75)(0.25)}{400}}<(P_1-P_2)$$

    $$<(0.05-0.025)+1.96\sqrt{\frac{(0.5)(0.5)}{600}+\frac{(0.75)(0.25)}{400}}$$

    $$1.19 < (P_1-P_2) < 0.31$$

12. Dari seluruh nilai ujian Bahasa Indonesia di sekolah "ACC", diambil sampel random sebanyak 100 nilai. Diketahui nilai rata-ratanya adalah 78.1 dengan standar deviasi 9.5. Hitunglah interval konfidensi 98% untuk deviasi standar seluruh siswa di sekolah tersebut.

    Jawab:

$$\frac{9.5}{1+\frac{2.33}{\sqrt{200}}} < \sigma < \frac{9.5}{1-\frac{2.33}{\sqrt{200}}}$$

$$\frac{9.5}{1.165} < \sigma < \frac{9.5}{0.835}$$

$$8.155 < \sigma < 11.377$$

13. Dari sampel random sebanyak 200 bola lampu listrik diketahui standar deviasi dari daya lampu adalah 100 jam. Hitunglah interval konfidensi 95% untuk standar deviasi daya nyala lampu seluruh bola lampu tersebut.

    Jawab:

$$100 - 1.96\frac{100}{\sqrt{2(200)}} < \sigma < 100 + 1.96\frac{100}{\sqrt{2(200)}}$$

$$100 - 9.8 < \sigma < 100 + 9.8$$

$$90.2 \text{ jam} < \sigma < 109.8 \text{ jam}$$

14. Diambil 12 sampel random dari siswa yang mengikuti pola belajar “case-based learning” dan 10 siswa lainnya yang mengikuti pola belajar “colaborative learning”. Hasil ujian matematika dari siswa yang mengikuti pola belajar “case-based learning” memiliki rata-rata 85 dengan standar deviasi 4. Sedangkan siswa yang mengikuti pola belajar “colaborative learning” memiliki rata-rata 81 dengan standar deviasi 5. Hitunglah interval konfidensi 90% untuk perbedaan antara rata-rata populasi. Asumsi populasi mendekati normal dan memiliki varians sama.

    Jawab:

$$4 - 1.725(1.92) < (\mu_1 - \mu_2) < 4 + 1.725(1.92)$$

$$0.69 < (\mu_1 - \mu_2) < 7.31$$

15. Dari dua area perkebunan, masing-masing diambil 100 sampel bibit. Dari area I, diketahui tinggi rata-rata adalah 9.8 inci dengan standar deviasi 1 inci. Sedangkan dari area II, diketahui tinggi rata-rata adalah 10.5 inci dengan standar deviasi 3 inci. Hitunglah interval konfidensi 90% untuk perbedaan rata-rata tinggi dari kedua populasi tersebut.

    Jawab:

$$0.7-1.64(0.316)<(\mu_2-\mu_1)<0.7-1.64(0.316)$$
$$0.18 \text{ inci}<(\mu_2-\mu_1)<1.22 \text{ inci}$$

16. Buatlah interval konfidensi 95% untuk nilai rata-rata populasi berdasarkan masing-masing random sampel dengan n=8, rata-rata sampel 127.4 dan standar deviasi 26.2.

    Jawab:

    $$127.4-2.365\frac{26.2}{\sqrt{8}}<\mu<127.4+2.365\frac{26.2}{\sqrt{8}}$$
    $$105.5 \text{ ton}<(\mu_2-\mu_1)<149.3 \text{ ton}$$

17. Dari hasil pengukuran 10 sampel diameter balok kayu, rata-rata diameter kayu tersebut adalah 43.8 cm dan standar deviasi 0.6 cm. Hitunglah interval konfidensi 99% rata-rata diameter yang sesungguhnya.

    Jawab:

    $$43.8-3.250\frac{0.6}{\sqrt{10}}<\mu<43.8+3.250\frac{0.6}{\sqrt{10}}$$
    $$43.18 \text{ cm}<\mu<44.2 \text{ cm}$$

18. Rata-rata waktu yang dibutuhkan alat mekanik tertentu berdasarkan 6 sampel perakitan adalah 13 menit, 14 menit, 12 menit, 16 menit, 12 menit dan 11 menit. Hitunglah interval konfidensi 95% untuk waktu rata-rata yang sesungguhnya dibutuhkan untuk merakit alat tersebut.

    Jawab:

    $$13-2.571\frac{1.79}{\sqrt{6}}<\mu<13+2.571\frac{1.79}{\sqrt{6}}$$
    $$11.12 \text{ menit}<\mu<14.88 \text{ menit}$$

19. Dari 200 nilai statistik, diambil 50 nilai secara random. Dari sampel tersebut, diketahui rata-rata 75 dan standar deviasi 10. Hitunglah interval konfidensi 95% untuk rata-rata dari 200 nilai matematika tersebut.

    Jawab:

$$75 - 1.96\frac{10}{\sqrt{50}}\sqrt{\frac{200-50}{200-1}} < \mu < 75 + 1.96\frac{10}{\sqrt{50}}\sqrt{\frac{200-50}{200-1}}$$

75-2.4 $< \mu <$ 75+2.4

72.6 $< \mu <$ 77.4

20. Berdasarkan soal (19), apabila dikatakan bahwa rata-rata dari 200 nilai matematika tersebut adalah 75 $\pm$ 2, tentukan tingkat konfidensinya.

    Jawab:

    $$75 - 1.23Z_{\frac{\alpha}{2}} < \mu < 75 + 1.23Z_{\frac{\alpha}{2}}$$

    75-2 $< \mu <$ 75+2

    $$1.23Z_{\frac{\alpha}{2}} = 2$$

    $$Z_{\frac{\alpha}{2}} = 1.63$$

    Jadi, daerah kurva normal dari Z=0 sampai Z=1.63 adalah 0.4484.

    Tingkat konfidensi adalah 2(0.44984)=0.8968 atau 90%

MODUL 10
STATISTIK INFERENSI: PENGUJIAN HIPOTESIS I-II

Pokok Bahasan:

- Prosedur umum dari Pengujian Hipotesis
- Pengujian Hipotesis Rerata Populasi
- Pengujian Hipotesis Selisih Dua Rerata Populasi
- Pengujian Hipotesis Proporsi Populasi
- Pengujian Hipotesis Selisih Dua Proporsi Populasi

10.1 Prosedur umum dari Pengujian Hipotesis

Dalam konteks pendugaan, informasi dari sampel digunakan untuk menduga nilai parameter populasi. Sedangkan dalam konteks pengujian, nilai parameter populasi diasumsikan terlebih dahulu. Selanjutnya, digunakan informasi dari sampel untuk menerima atau menolak nilai yang sudah diasumsikan tersebut.

Prosedur pengujian hipotesis mencakup:

1. Menetapkan *null hypothesis* (H_O) dan *alternative hypothesis* (H_1).
2. Menentukan "nilai kritis" atau daerah untuk menolak/menerima H_O.
3. Menghitung nilai tes statistik (sesuai distribusi yang digunakan).
4. Membuat keputusan secara statistik untuk menolak atau menerima H_O dengan membandingkan nilai tes statistik dengan nilai kritis.

10.2 Pengujian Hipotesis Rerata Populasi

Rata-Rata Populasi. Jika nilai rata-rata populasi yang dihipotesiskan adalah μ dan distribusi sampling rata-rata mendekati normal atau normal, rumus tes statistik untuk pengujian rata-rata populasi adalah: $Z = \dfrac{\bar{x} - \mu}{\sigma_{\bar{x}}}$.

Keterangan:

Z memiliki distribusi normal standar

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

$\sigma_{\bar{x}}$ = standar deviasi rata-rata

Jika standar deviasi populasi (σ) tidak diketahui, maka standar deviasi populasi dapat diduga dengan standar deviasi sampel (S). Standar deviasi rata-rata $(S_{\bar{X}})$ diduga dengan $S_{\bar{X}} = S/\sqrt{n}$.

Jika populasi normal, statistika $\bar{x} - \mu/S_{\bar{x}}$ memiliki distribusi t dengan derajat bebas n-1. Rumus tes statistik untuk pengujian rata-rata populasi (dengan distribusi t) adalah:

$$t = \frac{\bar{x} - \mu}{S_{\bar{x}}}$$

10.3 Pengujian Hipotesis Selisih Dua Rerata Populasi

Selisih Rata-Rata Populasi. Jika ada dua populasi dan masing-masing populasi diambil sampel lebih besar dari 30, maka distribusi sampling selisih rata-rata akan memiliki distribusi normal atau mendekati normal. Rumus tes statistik untuk pengujian selisih rata-rata populasi adalah:

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}}.$$

Keterangan:

Z memiliki distribusi normal standar

$(\bar{x}_1 - \bar{x}_2)$ = selisih rata-rata sampel

$(\mu_1 - \mu_2)$ = selisih rata-rata populasi

$\sigma_{\bar{x}_1 - \bar{x}_2}$ = standar deviasi rata-rata populasi

Rumus standar deviasi yang digunakan adalah:

(i) Jika standar deviasi populasi diketahui.

$$\sigma_{\bar{x}1-\bar{x}2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

(ii) Jika standar deviasi populasi tidak diketahui (σ), standar deviasi diduga dengan standar deviasi sampel (S). Standar deviasi rata-rata populasi $\sigma_{\bar{x}1-\bar{x}2}$ diduga dengan standar deviasi rata-rata sampel $(S_{\bar{x}1-\bar{x}2})$.

$$S_{\bar{x}1-\bar{x}2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

(iii) Jika standar deviasi populasi tidak diketahui tetapi nilainya dianggap sama $(\sigma_1 = \sigma_2)$, rumus standar deviasi adalah: (dengan derajat bebas $n_1 + n_2 - 2$)

$$S_{\bar{x}1-\bar{x}2} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Berdasarkan point (iii), nilai tes statistiknya adalah:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{S_{\bar{x}_1 - \bar{x}_2}}$$

10.4 Pengujian Hipotesis Proporsi Populasi

Proporsi Populasi. Jika nilai proporsi populasi yang dihipotesiskan adalah P dan distribusi sampling proporsi mendekati normal, rumus tes statistik untuk pengujian proporsi populasi adalah:

$$Z = \frac{p - P}{\sqrt{P(1 - P)/n}} \quad \text{atau} \quad Z = \frac{x - nP}{\sqrt{nP(1 - P)}} .$$

Keterangan:

$p =$ proporsi sampel

$x =$ jumlah sukses dalam sampel

Note:

Suatu distribusi sampling proporsi akan mendekati normal jika np dan $n(1 - p)$ keduanya lebih besar dari 15.

10.5 Pengujian Hipotesis Selisih Dua Proporsi Populasi

Selisih Proporsi Populasi. Jika nilai proporsi populasi yang dihipotesiskan adalah P dan distribusi sampling proporsi mendekati normal, rumus tes statistik untuk pengujian proporsi populasi adalah:

$$Z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sigma_{p1-p2}} .$$

Keterangan:

$(p_1 - p_2)$ = selisih proporsi sampel

$(P_1 - P_2)$ = selisih proporsi populasi

σ_{p1-p2} = standar deviasi selisih proporsi populasi

Rumus standar deviasi yang digunakan adalah:

(i) Jika standar deviasi populasi diketahui.

$$\sigma_{p1-p2} = \sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}$$

(ii) Jika nilai P tidak diketahui, dan nilai P diduga dengan p maka standar deviasi selisih proporsi populasi adalah

$$\sigma_{p1-p2} = \sqrt{p(1-p)\left(\frac{1}{n_1-1} + \frac{1}{n_2-1}\right)} \text{ dimana } p = \frac{x_1 + x_2}{n_1 + n_2}$$

x_1 dan x_2 adalah jumlah sukses pada sampel pertama dan kedua.

(iii) Jika nilai P tidak diketahui dan diasumsikan P_1 dan P_2 memiliki nilai sama sehingga $P_1 = P_2 = P$, maka

$$\sigma_{p1-p2} = \sqrt{P(1-P)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Note:

Pada umumnya, pengujian selisih proporsi menggunakan ukuran sampel yang besar agar distribusi sampling selisih proporsi mendekati normal. Sehingga disyaratkan $n_1 p_1, n_1(1-p_1), n_2 p_2$ dan $n_2(1-p_2)$ keduanya lebih besar dari 15.

10.6 Latihan: Soal & Jawab

1. Di sebuah kompleks perumahan "Sederhana dan Bahagia", petugas PLN mencatat perubahan konsumsi/penggunaan listrik sebagai dampak dari perubahan tegangan (dari 110 V menjadi 220 V). Sebelum ada perubahan tegangan, konsumsi listrik rata-rata untuk setiap pelanggan per bulan adalah 84 Kwh. Setelah terjadi perubahan tegangan menjadi 220 V, diadakan survei ke 100 pelanggan di kompleks tersebut. Hasilnya menunjukkan bahwa konsumsi listrik rata-

rata mengalami peningkatan menjadi 86,5 Kwh dengan standar deviasi 14 Kwh. Berdasarkan data tersebut, ujilah pendapat yang menyatakan bahwa perubahan tegangan tersebut mempunyai pengaruh kuat terhadap peningkatan konsumsi listrik di kompleks tersebut. Asumsi $\alpha = 5\%$.

Jawab:

a. H0:μ=84 Kwh

H1:μ>84 Kwh

b. Nilai $Z_{0.05} = 1.64$

c. H0 diterima jika $Z \leq 1.64$

H0 ditolak jika $Z > 1.64$

d. $Z = \dfrac{86.5\text{-}84}{\dfrac{14}{\sqrt{100}}} = 1.79$

e. Oleh karena nilai Z hitung (1.79) lebih besar daripada Z tabel (1.64) maka dapat disimpulkan bahwa perubahan tegangan dari 110 V menjadi 220 V mempunyai pengaruh kuat dalam konsumsi listrik

2. Di sebuah area perkebunan hortikultura, dibuat uji coba penanaman melon. Ada enam area yang masing-masing seluas ½ ha. Produksi di masing-masing area sebesar 1.4 ton, 1.8 ton, 1.1 ton, 1.9 ton, 2.2 ton, dan 1.2 ton. Dengan $\alpha = 5\%$, apakah angka-angka tersebut mendukung hipotesis bahwa rata-rata produksi melon per ½ ha adalah 1.5 ton.

Jawab:

a. H0:μ=1.5 ton

H1:$\mu \neq 1.5$ ton

b. Nilai $t_{0.025;5} = 2.571$

c. H0 diterima jika $-2.571 \leq t \leq 2.571$

H0 ditolak jika t>2.571 atau t<-2.571

d. $t = \dfrac{1.6\text{-}1.5}{\dfrac{0.4336}{\sqrt{6}}} = \dfrac{0.1}{0.177} = 0.565$

e. Oleh karena nilai t hitung (0.565) lebih kecil daripada t tabel (2.571) maka dapat disimpulkan bahwa hipotesis awal diterima

3. Petugas YLKI melakukan sampel random untuk produk bubuk puding "Enak" untuk melihat ketepatan antara berat yang tercantum pada label dengan berat yang sebenarnya. Dari 30 sampel random, tercantum di label "berat bersih 400 g" memiliki berat bersih rata-rata 398 g dengan deviasi standar 6 g. Apakah hasil pengukuran tersebut memungkinkan untuk menolak hipotesis $\mu = 400$ g; jika pengujian dilakukan dengan hipotesis alternatif dimana μ p 400 g dengan $\alpha = 1\%$.

   Jawab:

   a. H0: $\mu \geq 400$ g

      H1: $\mu < 400$ g

   b. Nilai $Z_{0.01} = -2.33$

   c. H0 diterima jika $Z \geq -2.33$

      H0 ditolak jika $Z < -2.33$

   d. $Z = \dfrac{398\text{-}400}{\dfrac{6}{\sqrt{30}}} = \dfrac{-2}{1.095} = -1.826$

   e. Oleh karena nilai Z hitung (-1.826) lebih besar daripada t tabel (-2.33) maka dapat disimpulkan bahwa hipotesis awal diterima

4. Di perusahaan tekstil "Tampak Hebat", mesin lama menghasilkan produksi rata-rata 12 unit/jam. Manajer dieri masukan untuk mengganti dengan mesin baru meskipun biayanya agak mahal. Mesin baru tersebut baru menguntungkan jika produksi rata-rata lebih besar dari 15 unit/jam. Untuk memutuskan apakah menggunakan mesin baru atau tidak, diadakan uji coba dengan 9 mesin baru. Dari uji coba tersebut, produksi rata-rata 16.5 unit/jam dengan standar deviasi 2.8 unit. Apa keputusan yang harus diambil jika $\alpha = 5\%$.

   Jawab:

a. H0: $\mu = 15$ menit

H1: $\mu > 15$ menit

b. Nilai $t_{0.05;9\text{-}1} = 1.860$

c. H0 diterima jika $t \leq 1.860$

H0 ditolak jika $t > 1.860$

d. $t = \dfrac{16.5\text{-}15}{\frac{2.8}{\sqrt{9}}} = 1.607$

e. Oleh karena nilai t hitung (1.607) lebih kecil daripada t tabel (1.860) maka dapat disimpulkan bahwa hipotesis awal diterima

5. Perusahaan penyalur tenaga kerja "Sangat Cepat" memberikan pelatihan mengetik kepada anggotanya. Berdasarkan hasil uji sampel sebanyak 12 orang, rata-rata kecepatan mengetik adalah 73.8 kata/menit dengan standar deviasi 7.9 kata. Dengan $\alpha = 1\%$, anda diminta untuk menguji pendapat bahwa rata-rata kecepatan mengetik kurang dari 75 kata/menit.

Jawab:

a. H0: $\mu = 75$ kata menit

H1: $\mu < 75$ kata menit

b. Nilai $-t_{0.01;12\text{-}1} = -2.718$

c. H0 diterima jika $t \geq -2.718$

H0 ditolak jika $t < -2.718$

d. $t = \dfrac{73.8\text{-}7.5}{\frac{7.9}{\sqrt{12}}} = -0.526$

e. Oleh karena nilai t hitung (-0.526) lebih besar daripada t tabel (-2.718) maka dapat disimpulkan bahwa hipotesis awal diterima

6. Perusahaan cat "Warna-Warni" mampu menghasilkan cat tembok super. Setiap kaleng cukup untuk memoles tembok rata-rata seluas 10 m^2. Untuk meyakinkan bahwa pendapat tersebut benar, ada 6 kaleng cat yang diuji coba. Tembok yang berhasil dicat untuk setiap kaleng masing-masing seluas 9.8 m^2, 10.2 m^2, 10 m^2, 9.4 m^2, 9.6 m^2, dan 10.5 m^2. Bagaimana kesimpulan yang bisa diambil dengan dengan $\alpha = 5\%$.

Jawab:

a. H0: $\mu = 10\ \text{m}^2$

H1: $\mu \neq 10\ \text{m}^2$

b. Nilai $-t_{0.025;6\text{-}1} = 2.571$

c. H0 diterima jika $-2.571 \leq t \leq 2.571$

H0 ditolak jika $t < -2.571$ atau $t > 2.571$

d. $t = \dfrac{9.92\text{-}10}{\dfrac{0.402}{\sqrt{6}}} = \dfrac{-0.08}{0.164} = -0.488$

e. Oleh karena nilai t hitung (-0.488) lebih besar daripada t tabel (-2.571) maka dapat disimpulkan bahwa hipotesis awal diterima

7. Perusahaan Farmasi "Mujarab" melakukan pendataan hasil penjualan di kota A dan B. Diambil sampel random 9 salesman dari kota A dan 6 salesman dari kota B pada periode tertenrtu. Data di bawah menunjukkan angka penjualan tersebut. (dalam unit)

Kota A: 41, 47, 62, 39, 56, 64, 37, 61 dan 52

Kota B: 34, 63, 45, 55, 24, 43

Dengan $\alpha = 1\%$, ujilah apakah perbedaan antara rata-rata kedua sampel tersebut signifikan.

Jawab:

a. H0: $\mu 1 \neq \mu 2$

H1: $\mu 1 = \mu 2$

b. Nilai $-t_{0.005;6\text{-}2} = 3.012$

c. H0 diterima jika $-3.012 \leq t \leq 3.012$

H0 ditolak jika $t < -3.012$ atau $t > 3.012$

d. $t = \dfrac{51\text{-}44}{\sqrt{\dfrac{(8)(10.44)^2 + (5)(14.03)^2}{9+6-2}\left(\dfrac{1}{9} + \dfrac{1}{6}\right)}} = \dfrac{7}{6.298} = 1.11$

e. Oleh karena nilai t hitung (1.11) lebih kecil daripada t tabel (3.012) maka dapat disimpulkan bahwa hipotesis awal diterima

8. Pemerintah ingin melihat dampak permberian kredit dengan angsuran ringan kepada pengusaha kecil. Oleh karena itu, dilihat kondisi pengusaha kecil sebelum dan sesudah pemberian kredit. Hasil survei dari 8 sampel random pengusaha kecil adalah sebagai berikut.

Pengusaha kecil	Profit sebelum kredit	Profit sesudah kredit
A	200	250
B	400	390
C	350	400
D	250	350
E	250	300
F	150	200
G	450	600
H	500	600

Tunjukkan apakah kredit tersebut dapat meningkatkan profit pengusaha kecil dengan $\alpha = 5\%$.

Jawab:

a. H0: $\mu 2 = \mu 1$

H1: $\mu 2 > \mu 1$

b. Nilai $-t_{0.005;8-1} = 1.895$

c. H0 diterima jika $t \leq 1.895$

H0 ditolak jika $t > 1.895$

d. $t = \dfrac{\overline{d}}{S_{\bar{d}}\sqrt{n}} = \dfrac{67.5}{48.033/\sqrt{8}} = 3.975$

e. Oleh karena nilai t hitung (3.975) lebih besar daripada t tabel (1.895) maka dapat disimpulkan bahwa hipotesis awal ditolak

9. Dinas perkebunan ingin melihat pertumbuhan dari dua bibit jeruk unggulan, yaitu “Super” dan “Heboh”. Untuk kedua jenis bibit unggul tersebut masing-masing diambil sampel random sebanyak 12 dan 15. Bibit jeruk “Super” memiliki tinggi rata-rata 13.6 kaki dan standar deviasi 1.2 kaki. Sedangkan bibit jeruk “Heboh” memiliki tinggi rata-rata 12.7 kaki dan standar deviasi 1.5 kaki. Apakah perbedaan rata-rata kedua sampel tersebut signifikan dengan $\alpha = 1\%$.

Jawab:

a. H0: $\mu_1 = \mu_2$

H1: $\mu_1 \neq \mu_2$

b. Nilai $-t_{0.005;12+15-2} = 2.787$

c. H0 diterima jika $-2.787 \leq t \leq 2.787$

H0 ditolak jika $t > 2.787$ atau $t < -2.787$

d. $$t = \frac{13.6\text{-}12.7}{\sqrt{\left(\frac{(11)(1.2)^2 + (14)(1.5)^2}{12+15-2}\right)\left(\frac{1}{12} + \frac{1}{15}\right)}} = \frac{0.9}{2.66} = 0.338$$

e. Oleh karena nilai t hitung (0.338) lebih kecildaripada t tabel (2.787) maka dapat disimpulkan bahwa hipotesis awal diterima

10. Dari sampel random turis yang berkunjung ke Gili Trawangan, diketahui 84 dari 250 wisatawan laki-laki dan 156 dari 300 wisatawan perempuan membeli kain tenun Lombok. Dengan $\alpha = 5\%$, uji hipotesis nol bahwa tidak ada perbedaan antara kedua proporsi populasi tersebut. Hipotesis alternatifnya adalah $P_2 > P_1$.

Jawab:

a. H0: $P_2 = P_1$

H1: $P_2 > P_1$

b. Nilai $Z_{0.05} = 1.64$

c. H0 diterima jika $Z \leq 1.64$

H0 ditolak jika $Z > 1.64$

d. $$P = \frac{84+156}{250+300} = 0.44$$

$$Z = \frac{\frac{156}{310} - \frac{84}{250}}{\sqrt{0.44(0.56)\left(\frac{1}{250} + \frac{1}{300}\right)}} = \frac{0.184}{0.043} = 4.28$$

e. Oleh karena nilai t hitung (4.28) lebih besar daripada t tabel (1.64) maka dapat disimpulkan bahwa hipotesis awal ditolak

11. Dalam pertandingan memasak, diambil sampel random sebanyak 18 peserta untuk melihat waktu yang dibutuhkan untuk memasak sop iga; standar deviasi sebesar 2.1 menit. Ujilah hipotesis nol bahwa $\sigma = 2.5$ menit dan hipotesis alternatif $\sigma \neq 2.5$ menit pada $\alpha = 5\%$.

Jawab:

a. H0: $\sigma = 25$ menit

H1: $\sigma \neq 25$ menit

b. Nilai $\chi^2{}_{0.025;\ 17} = 30.191$

c. $\chi^2{}_{(1-\frac{\alpha}{2});n-1} = \chi^2{}_{0.975;\ 17} = 7.564$

H0 diterima jika $7.564 \leq \chi^2 \leq 30.191$

H0 ditolak jika $\chi^2 > 30.191$ atau $\chi^2 < 7.564$

d. $\chi^2 = \dfrac{\text{(n-1)S}^2}{\sigma_0^2} = 0.44$

$$= \frac{(18\text{-}1)(2.1)^2}{(2.5)^2} = \frac{74.97}{6.25} = 11.995$$

e. Oleh karena nilai t hitung (11.995) lebih kecil daripada t tabel (30.191) maka dapat disimpulkan bahwa hipotesis awal diterima

12. Seorang pelatih lari ingin mengetahui apakah rata-rata waktu yang dibutuhkan untuk mengelilingi stadion dalam satu putaran adalah 23 menit. Dari sampel random sebanyak 32 putaran, rata-rata waktu yang dibutuhkan adalah 25.8 menit setiap putaran dengan standar deviasi 1.5 menit. Dengan $\alpha = 5\%$, anda diminta untuk menguji hipotesis awal bahwa $\sigma = 2.0$ menit dan hipotesis alternatif $\sigma < 2.0$ menit.

Jawab:

a. H0: $\sigma = 2$ menit

H1: $\sigma < 2$ menit

b. Nilai $Z_{0.05} = 1.64$

c. H0 diterima jika $Z \geq -1.64$

H0 ditolak jika $Z < -1.64$

$$\text{d.Z} = \frac{1.5\text{-}2}{(2)/\sqrt{64}} = \frac{-0.5}{0.25} = -2.0$$

e. Oleh karena nilai Z hitung (-2.0) lebih kecil daripada Z tabel (-1.64) maka dapat disimpulkan bahwa hipotesis awal ditolak

13. Seorang guru bahasa Indonesia mengajar di dua kelas, yaitu kelas A dan B yang masing-masing berjumlah 16 dan 25 siswa. Dari hasil nilai ujian, rata-rata di kedua kelas tersebut tidak menunjukkan perbedaan yang signifikan; dilihat dari standar deviasi, standar deviasi kelas A dan B adalah 9 dan 12. Apakah dapat diambil kesimpulan bahwa pada $\alpha = 5\%$, variabilitas kelas B lebih besar dari kelas A.

Jawab:

$$\hat{S}_A^2 = \frac{n_A}{n_A - 1} S_A^2 = \frac{16}{15}(9)^2 = 86.4$$

$$\hat{S}_B^2 = \frac{n_B}{n_B - 1} S_B^2 = \frac{25}{24}(12)^2 = 150$$

a. H0: $\sigma_A = \sigma_B$

H1: $\sigma_B > \sigma_A$

b. Nilai $F_{0.01;24;15} = 3.29$

c. H0 diterima jika $F \leq 3.29$

H0 ditolak jika $F > 3.29$

$$\text{d.F} = \frac{\hat{S}_B^2}{\hat{S}_A^2} = \frac{150}{86.4} = 1.74$$

e. Oleh karena nilai F hitung (1.74) lebih kecil daripada F tabel (3.29) maka dapat disimpulkan bahwa hipotesis awal diterima

14. Seorang pengusaha percetakan ingin meneliti mesin fotokopi yang dimilikinya. Ada dua jenis mesin fotokopi, yaitu “COPYKU” dan “COPYMU”. Untuk mesin fotokopi “COPYKU” dari 60 kerusakan yang terjadi diperlukan waktu rata-rata 91.6 menit untuk memperbaiki dengan standar deviasi 18.8 menit. Sedangkan mesin fotokopi "COPYMU" dari 60 kerusakan yang terjadi diperlukan waktu rata-rata 84.2 menit untuk memperbaiki dengan standar deviasi 19.4 menit. Dengan $\alpha = 1\%$ anda diminta untuk menguji apakah cukup alasan untuk menganggap varians kedua populasi tersebut adalah sama.

Jawab:

a. H0: $\sigma_1 = \sigma_2$

H1: $\sigma_1 \neq \sigma_2$

b. Nilai $F_{0.01;59;59} = 1.84$

c. H0 diterima jika $F \leq 1.84$

H0 ditolak jika $F > 1.84$

d. $F = \dfrac{\hat{S}_2^2}{\hat{S}_1^2} = \dfrac{19.4}{18.8} = 1.03$

e. Oleh karena nilai F hitung (1.03) lebih kecil daripada F tabel (1.84) maka tidak cukup alasan untuk mengatakan bahwa deviasi standar kedua populasi berbeda nyata

15. Dari hasil riset wisata kuliner yang dilakukan di kota Surabaya dan Malang dapat ditunjukkan preferensi makanan favorit ibu rumah tangga. Di Surabaya, dari 100 ibu RT, 68 diantaranya menyukai "Rujak Ulek Madura" dibandingkan dengan "Gudeg Yogja". Sedangkan di kota Malang, dari 300 ibu RT, 213 diantaranya menyukai "Gudeg Yogja" dibandingkan dengan "Rujak Ulek Madura".

Jawab:

a. H0: $P_1 = P_2$

H1: $P_1 \neq P_2$

b. Nilai $Z_{0.005} = 2.58$

c. H0 diterima jika $-2.58 \leq Z \leq 2.58$

H0 ditolak jika $Z > 2.58$ atau $Z < -2.58$

d. $P = \dfrac{68+213}{100+300} = \dfrac{281}{400} = 0.7025$

$$Z = \frac{(0.68-0.71)}{\sqrt{0.7025(0.2975)\left(\frac{1}{100}+\frac{1}{300}\right)}} = \frac{-0.03}{0.0528} = -0.568$$

e. Oleh karena nilai Z hitung (-0.568) lebih besar daripada Z tabel (-2.58) maka hipotesis awal diterima.

16. Seorang ahli nutrisi berpendapat bahwa 75% murid SD di daerah pedalaman X menderita kekuarangan gizi. Berdasarkan sampel random sebanyak 300 murid SD, diketahui 206 anak menderita kekeurangan gizi. Dengan $\alpha = 1\%$, anda diminta menguji hipotesis awal yang menyatakan bahwa *p=0.75* dan hipotesis alternatif *p<0.75.*

Jawab:

a. H0: $P = 0.75$

H1: $P < 0.75$

b. Nilai $Z_{0.001} = 2.33$ (pengujian satu sisi ke kiri)

c. H0 diterima jika $Z \geq -2.33$

H0 ditolak jika $Z < -2.33$

d. $\dfrac{X}{n} = \dfrac{206}{300} = 0.687$

$$Z = \frac{(0.687-0.75)}{\sqrt{0.75(0.25)/300}} = \frac{-0.063}{0.25} = -2.52$$

e. Oleh karena nilai Z hitung (-2.52) lebih kecil daripada Z tabel (-2.33) maka hipotesis awal ditolak.

17. Berdasarkan informasi, 30% siswa yang mengambil kelas montir dikeluarkan (DO) sesudah tahun pertama. Dengan $\alpha = 5\%$, anda diminta untuk menguji informasi tersebut dengan alternatif angka DO kurang dari 30% apabila dari sampel random sebanyak 500 siswa, ada 124 siswa yang DO sesudah tahun pertama.

Jawab:

a. H0:$P = 30\%$

H1:$P < 30\%$

b. Nilai $Z_{0.005} = 1.64$ (pengujian satu sisi ke kiri)

c. H0 diterima jika $Z \geq -1.64$

H0 ditolak jika $Z < -1.64$

d.$\frac{X}{n} = \frac{124}{500} = 0.248$

$$Z = \frac{(0.248 - 0.300)}{\sqrt{0.30(0.70)/500}} = \frac{-0.052}{0.0205} = -2.537$$

e. Oleh karena nilai Z hitung (-2.537) lebih kecil daripada Z tabel (-1.64) maka hipotesis awal ditolak.

18. Pabrik mobil "Selalu Bertahan" menyatakan bahwa 35% dari seluruh mobil keluaran tahun 1980 masih dalam keadaan baik di tahun 1990. Untuk menguji pernyataan tersebut, diambil sampel random sebanyak 800 mobil keluaran tahun 1980. Dari 800 mobil tersebut, ada 257 mobil yang masih berjalan baik. Dengan hipotesis alternatif $p \neq 0.35$, apa yang dapat anda simpulkan dengan $\alpha = 1\%$.

Jawab:

a. H0:$P = 35\%$

H1:$P \neq 35\%$

b. Nilai $Z_{0.005} = 2.58$ (pengujian dua sisi)

c. H0 diterima jika -2.58 $\leq Z \leq 2.58$

H0 ditolak jika $Z < -2.58$ atau $Z > 2.58$

d.$\frac{X}{n} = \frac{257}{800} = 0.32125$

$$Z = \frac{(0.32125 - 0.35)}{\sqrt{0.35(0.65)/800}} = \frac{-0.02875}{0.016886} = -1.705$$

e. Oleh karena nilai Z hitung (-1.705) lebih besar daripada Z tabel (-2.58) maka hipotesis awal diterima.

19. Seorang dosen dari universitas "Sangat Terbuka" menyatakan bahwa 75% mahasiswa lebih menyukai paper dibandingkan dengan ujian final. Untuk menguji pernyataan tersebut diambil sampel random sebanyak 320 mahasiswa dari dosen tersebut. Dari 320 mahasiswa, ada 267 mahasiswa yang lebih menyukai paper. Dengan $\alpha = 5\%$, anda diminta menguji hipotesis awal *p=0.75* dengan hipotesis altenatif *p> 0.75.*

Jawab:

a. $H0: P = 0.75$

$H1: P > 0.75$

b. Nilai $Z_{0.05} = 1.64$ (pengujian satu sisi kanan)

c. H0 diterima jika $Z \leq 1.64$

H0 ditolak jika $Z > 1.64$

d. $\frac{X}{n} = \frac{267}{320} = 0.8343$

$$Z = \frac{0.8343 - 0.75}{\sqrt{\frac{0.75(0.25)}{320}}} = 3.486$$

e. Oleh karena nilai Z hitung (3.486) lebih besar daripada Z tabel (1.64) maka hipotesis awal ditolak.

20. Dilakukan ujian bagi dua kategori siswa, yaitu kelompok A dan B dimana kelompk A sedikit lebih cerdas dibandingkan dengan kelompok B. Dari 250 siswa kelompok A, ada 205 siswa yang menjawab dengan benar. Sedangkan dari 250 siswa kelompok B, ada 137 siswa yang mejawab dengan benar. Dengan $\alpha = 5\%$, anda diminta menguji pendapat bahwa kelompok A bisa memberikan jawaban benar sebanyak 20% lebih tinggi dibandingkan dengan kelompok B.

Jawab:

a. H0: $P1 - P2 = 20\%$

H1: $P1 - P2 \neq 20\%$

b. Nilai $Z_{0.025} = 1.96$ (pengujian dua sisi)

c. H0 diterima jika -1.96 $\leq Z \leq 1.96$

H0 ditolak jika $Z < -1.96$ atau $Z > 1.96$

$$\text{d.} Z = \frac{\frac{205}{250} - \frac{137}{250} - 0.2}{\sqrt{\frac{\frac{205}{250}\left(\frac{45}{250}\right)}{250} + \frac{\frac{137}{250}\left(\frac{113}{250}\right)}{250}}} = \frac{0.072}{0.03976} = 1.811$$

e. Oleh karena nilai Z hitung (1.811) lebih kecil daripada Z tabel (1.96) maka hipotesis awal diterima.

Referensi:

1. **Levin**, Richard I. dan David S. Rubin, (1996), ***Statistics for Management***, Prentice Hall, Inc. **(LR)**
2. **Newbold,** Paul , (1995), ***Statistics for Business and Economics***, Prentice Hall, Inc. **(PN)**
3. **Lind**, Douglas A., William G. Marchal and Samuel A. Wathen, (2004), ***Statistical Techniques in Business & Economics***, McGraw-Hill. **(LMW)**

4. **Mulyon**o, Sri, (2003***), Statistika untuk Ekonomi***, LPFEUI. **(SM)**
5. Wikipedia (Ensiklopedia Statistik)