

DISTRIBUSI PROBABILITAS

A. Pendahuluan

Kunci aplikasi probabilitas dalam statistic adalah memperkirakan terjadinya peluang/probabilitas yang dihubungkan dengan terjadinya peristiwa tersebut dalam beberapa keadaan. Sebagai contoh, kita ingin mengetahui probabilitas sebuah keluarga untuk memiliki anak laki-laki dan perempuan. Jika kita mengetahui keseluruhan probabilitas dari kemungkinan outcome yang terjadi, seluruh probabilitas kejadian tersebut akan membentuk suatu distribusi probabilitas. Perlu diketahui perbedaan distribusi frekuensi dengan distribusi probabilitas. Distribusi frekuensi hanya mendata semua hasil akhir yang memungkinkan dari suatu eksperimen dan menghitung jumlah frekuensi semua hasil akhir yang muncul. Hitungan tersebut kemudian disusun menjadi grafik yang membuatnya mudah divisualisasikan dan dipahami. Distribusi probabilitas menggambarkan semua hasil akhir yang memungkinkan dari suatu variable, bukan menggambarkan frekuensinya. Walau tampak serupa, keduanya mewakili dua konsep yang berlainan.

Perhatikan perbedaan antara distribusi frekuensi dan distribusi probabilitas berikut ini. Distribusi frekuensi dari permainan lempar dadu 1000 kali dan bandingkan dengan distribusi probabilitas putar dadu tak terbatas. Seperti yang terlihat, distribusi frekuensi memiliki celah dan perubahan yang bergantung pada tampilan sampel. Distribusi probabilitasnya seragam dan tidak pernah berubah. Hubungan antara frekuensi relative dan probabilitas hampir mirip dengan hubungan antara sampel dan populasi. Perbedaannya ada pada perbedaan dasar antara lemparan jangka pendek (100 putaran dadu) dan lemparan jangka panjang (1000, 10.000, 100.000 dan seterusnya lemparan dadu) dan untuk mendefinisikan "jangka panjang" memang agak subyektif.

Mungkin ada yang berpendapat bahwa menjadi perawat tidak perlu melempar dadu untuk mendapatkan probabilitas. Apa pun itu, melempar dadu hanyalah sebuah contoh langsung sebuah eksperimen. Berikut ini adalah beberapa eksperimen sederhana lain dan kemungkinan berhubungan dengan kehidupan dan profesi perawat. Sebagai contoh, jawablah pertanyaan-pertanyaan berikut:

- Berapa frekuensi relative (atau probabilitas) jangka panjang dari lemparan uang logam yang menghasilkan gambar?
- Berapa frekuensi relative (atau probabilitas) jangka panjang dari lempar dadu yang menghasilkan angka 6?

Sekarang, coba jawab pertanyaan yang lebih menarik berikut ini:

- Berapa probabilitas saya akan meninggal karena penyakit kardiovaskular?
- Berapa probabilitas salah satu pasien saya akan dipulangkan sebelum saya kembali ke rumah sakit untuk shift saya berikutnya?

- Berapa probabilitas suatu obat baru akan mematikan tumor kanker hati pasien saya tanpa membunuhnya?

Untuk menjawab pertanyaan-pertanyaan tersebut dengan tingkat keakuratan tertentu membutuhkan banyak kajian, banyak pengulangan sampel, dan menentukan probabilitas.

Ada bermacam-macam distribusi probabilitas (distribusi teoritis) dalam analisis statistic, antara lain Distribusi Binomial (Bernauli), Distribusi Poisson, Distribusi Normal (Gauss), Distribusi Student ('t' W Gosset), Distribusi Chi Square (X²), Distribusi Fisher (F), dan lain-lain.

Dalam bab ini akan dibahas lebih lanjut tiga macam distribusi yang umumnya digunakan, yaitu distribusi Binomial, Poisson, dan Normal. Random variable pada ketiga macam distribusi ini merupakan variable random yang diskrit.

B. Distribusi Binomial

Distribusi binomial menggambarkan fenomena dengan dua hasil atau outcome. Sebagai contoh, peluang sukses dan gagal, sehat dan sakit, dan sebagainya. Penemu distribusi binomial adalah James Bernauli sehingga distribusi binomial dikenal juga sebagai Distribusi Bernauli.

Bernauli trial mempunyai empat syarat sebagai berikut:

1. Jumlah trial merupakan bilangan bulat
2. Setiap eksperimen mempunyai dua outcome (hasil) yaitu sukses dan gagal.
Contoh:
 - Laki-laki / perempuan
 - Sehat / sakit
 - Setuju / tidak setuju
3. Peluang sukses sama setiap eksperimen
4. Setiap eksperimen independen satu sama lain

Di dalam mempelajari distribusi binomial ini, trial independen adalah setiap trial atau peristiwa bebas satu sama lain, misalnya trialnya melemparkan (melambungkan) satu koin 5 kali, antara lambungan pertama, kedua, sampai kelima adalah kejadian yang independen. Setiap eksperimen atau trial mempunyai dua hasil, yaitu sukses dan gagal. Jadi, ketika melambungkan koin mengharapkan akan keluar permukaan H (head) dan pada waktu melambungkan keluar H, maka dikatakan sukses, sedangkan kalau yang keluar sisi T (tail), maka dikatakan gagal. Setiap lambungan mempunyai probabilitas sukses adalah $\frac{1}{2}$ (setengah), pada lambungan seterusnya juga setengah. Jumlah

lambungan adalah bilangan bulat 1, 2, 3, dan seterusnya. Tidak mungkin melambungkan koin $3 \frac{1}{2}$ kali. Bagaimana kalau yang dilambungkan adalah dadu? Dalam hal ini permukaan yang diharapkan keluar misalnya mata lima, maka dikatakan peluang (probabilitas) sukses adalah $\frac{1}{6}$ (seperenam), sedangkan peluang gagal adalah $\frac{5}{6}$ (lima per enam). Untuk itu, probabilitas sukses dilambangkan p, sedangkan peluang gagal adalah $(1-p)$ atau biasa juga dilambangkan q, di mana $q=1-p$.

Kejadian atau peristiwa binomial ini disimbolkan $b(x,n,p)$ di mana b= binomial, x adalah banyaknya sukses yang diinginkan (bilangan random), n = jumlah trail, p=peluang sukses dalam satu kali trial.

Contoh, dadu dilemparkan 5 kali, diharapkan keluar mata 6 dua kali, maka kejadian ini dapat ditulis $b(2, 5, 1/6)$... $x=2, n=5 p =1/6$.

Suatu trial n kali nilai n dapat (1, 2, 3, ... n), sedangkan sukses x dapat merupakan bilangan random mulai dari (0, 1, 2, .. n).

Contoh permasalahan:

Probabilitas seorang bayi tidak diimunisasi polio adalah 0,2 (p). pada suatu hari di Puskesmas 'PQR" ada 4 orang bayi. Hitunglah peluang dari bayi tersebut 2 orang belum imunisasi polio. Jadi, di dalam kejadian binomial ini dikatakan b ($x=2, n=4, p=0,2$) ... $b(2,4, 0,2)$.

Penyelesaian:

Katakanlah bayi tersebut A, B,C, D, dua orang tidak diimunisasi mungkin adalah A&B, A&C, A&D, B&C, B&D, C&D.

Ada enam pasangan yang mungkin merupakan kombinasi dari empat objek yang pasangannya 2 ($x=2 n=4$) kombinasinya

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

Dua orang tidak diimunisasi dan dua orang diimunisasi. Peluangnya adalah $0,2^2 \times 0,8^2$ atau $p^x(1-p)^{n-x}$.

Jadi, rumus untuk $b(x,n,p)$ adalah:

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

Jadi, peluang dua di antara empat bayi yang belum imunisasi polio adalah:

$$\frac{4!}{2!(4-2)!} 0,2^2 \times 0,8^2 = 0,1536 = 0,154$$

C. Distribusi Poisson

Di dalam mempelajari distribusi binomial kita dihadapkan pada probabilitas variable random diskrit yang jumlah trialnya kecil (daftar binomial). Kalau suatu kejadian dengan probabilitas $p \ll \ll$ dan menyangkut kejadian yang luas $n \gg \gg \gg$ maka distribusi binomial tidak mampu lagi menentukan peluang variable diskrit tersebut. Di sini distribusi Poisson dapat dipakai untuk menjelaskannya.

Distribusi Poisson dipakai untuk menentukan peluang suatu kejadian yang jarang terjadi, tetapi mengenai populasi yang luas atau area yang luas dan juga berhubungan dengan waktu.

Contoh:

1. Di suatu gerbang tol yang dilewati ribuan mobil dalam satu hari akan terjadi kecelakaan dari sekian banyak mobil yang lewat.
2. Dikatakan bahwa kejadian seseorang akan meninggal karena shock pada waktu disuntik dengan vaksin meningitis 0,0005. Padahal, vaksinasi tersebut selalu diberikan kalau seseorang ingin pergi haji.

Distribusi Poisson merupakan fungsi probabilitas :

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

$\mu = \lambda = np = E(x)$ -----> nilai rata-rata

$e = \text{konstanta} = 2,71828$

$x = \text{variable random diskrit } (1, 2, \dots, X)$

D. Distribusi Normal

Distribusi ini pertama kali ditemukan oleh matematikawan asal Prancis, Abraham Demoivre pada tahun 1733 dan diaplikasikan dengan lebih baik lagi pada awal abad ke-19 oleh matematikawan asal Prancis Perre Simon de Laplace dan matematikawan sekaligus astronom asal Jerman, Karl Friedrich Gauss. Distribusi normal dikenal juga sebagai distribusi Gauss.

Sir Franciss Allon dan sepupunya Charles Darwin merupakan orang yang pertama kali mengaplikasikan kurva distribusi normal ini ke dalam dunia kedokteran. Beberapa fenomena menunjukkan gambaran distribusi normal terdapat pada variable random kontinu, seperti tinggi badan, serum kolesterol, suhu tubuh orang sehat, dan sebagainya.

Ciri khas Distribusi Distribusi Normal

Distribusi probabilitas untuk variable kontinu dengan puncak distribusi berada pada mean dan bentuk distribusi simetris, yang ditentukan oleh simpang bakunya, memiliki ciri-ciri sebagai berikut:

- simetris
- Seperti lonceng
- Titik belok $\mu \pm \sigma$
- Luas di bawah kurva = probability = 1

Fungsi atau $f(x)$ distribusi kontinu akan selalu dapat dicari dengan persamaan fungsi kurva normal (secara integral), tetapi hal ini tidak praktis. Agar lebih praktis dengan menggunakan table yang menunjukkan luas kurva normal dari suatu nilai yang dibatasi nilai tertentu. Kurva normal standar mempunyai $\mu = 0$ dan $\sigma = 1$, di mana $N(0,1)$.

Untuk suatu sampel yang cukup besar, terutama untuk gejala alam seperti berat badan dan tinggi badan, biasanya kurva yang dibentuk dari distribusi tersebut juga simetris dengan x tertentu dan Sd (simpangan baku) tertentu. Maka kurva simteris yang terjadi disebut kurva normal umum.

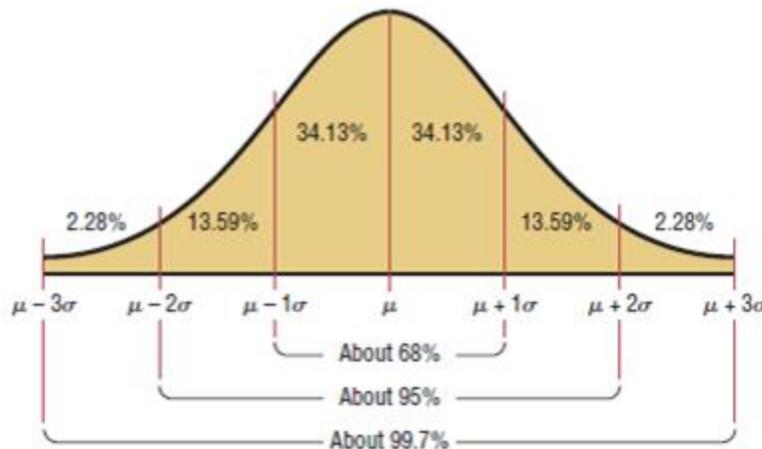
Banyak variable yang menjadi focus kajian memiliki distribusi probabilitas yang sangat menyerupai suatu distribusi yang sangat terkenal yaitu distribusi normal, suatu distribusi probabilitas yang rerata, median, dan modusnya sama. Kenyataannya, salah satu asumsi yang paling seragam dalam melaksanakan riset dasar adalah bahwa variable memiliki distribusi normalnya yang dapat diestimasi berdasarkan distribusi probabilitasnya. Distribusi normal sebenarnya sama dengan yang disebut sebagai "kurva lonceng". Kurva inilah yang terbentuk apabila tidak seorang pun mendapatkan nilai terbaik pada ujian. Banyak orang yang yakin bahwa dengan memberikan soal ujian yang tersusun rapi kepada suatu sampel besar akan menghasilkan distribusi nilai yang dapat diestimasi melalui distribusi normal.

Karena dalam distribusi normal rerata, median, dan modus adalah sama, ada dua hal tentang variable yang memiliki distribusi normal:

- 68% nilai berada dalam satu deviasi standar rerata.
- 95% nilai berada dalam dua deviasi standar rerata

Pada kebanyakan peneliti, distribusi normal merupakan distribusi terpenting dari semua statistik. Dua fakta paling penting yang menjadikan distribusi normal begitu penting adalah:

1. Jika Anda menghitung rerata dari begitu banyak sampel suatu populasi, distribusi rerata sampel (distribusi sampling dari rerata) menjadi normal dalam jangka panjang. Distribusi sampling menempatkan frekuensi actual suatu statistic versus kisaran nilai yang memungkinkan dari statistic tersebut.
2. Ketika Anda terus menambahkan variable acak, distribusi keseluruhannya mendekati distribusi normal.



Di dalam gambar, perhatikan pola distribusi normal ketika Anda menyejajarkan reratanya. Ketika rerata (μ) distribusi menunjukkan peningkatan dalam grafik, kurva kan bergeser ke kanan, jika rerata (μ) menurun kurva akan bergeser ke kiri. Pergeseran ini disebut rerata suatu parameter lokasi.

Di dalam gambar kita lihat apa yang terjadi apabila kita mengubah variansi. Jika kita hanya mengubah variansi (σ) dalam rumus, kita hanya melihat perubahan dalam skala. Ketika kita mengurangi variansi, grafiknya bertambah tinggi dan ramping, dan jika menambah variansinya, grafik akan memendek dan membulat. Itu yang menyebabkan variansi disebut sebagai variansi skala.

Berikut ini adalah beberapa hal penting yang perlu diketahui mengenai distribusi normal:

- 68% area di bawah lengkungan kurva berada dalam 1 deviasi standar rerata
 - 95% area di bawah lengkungan kurva berada dalam 2 deviasi standar rerata

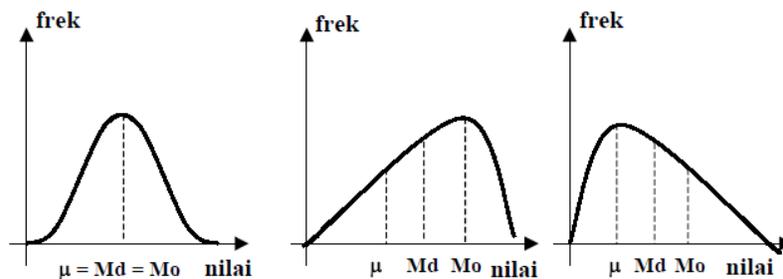
- Meningkatkan rerata menyebabkan urva bergeser ke kanan
- Mengurangi rerata menyebabkan kurva bergeser ke kiri
- Menurunkan variansi menjadikan grafik terlihat lebih tinggi dan lebih ramping
- Meningkatkan variansi menjadikan grafik terlihat lebih pendek dan lebih gemuk

Satu hal penting yang dapat kita lakukan terhadap variable normal adalah mengubah distribusinya menjadi distribusi normal standar. Hal ini membuat semua area di bawah kurva berada pada kurva normal dengan rerata 0 dan deviasi standar 1. Jika Y adalah variable terdistribusi normal, rumus tersebut akan menghasilkan Z yaitu variable normal standar. Variable normal standar sangat membantu karena akan tahu probabilitasnya. Sebagai contoh, 5% probabilitas berada di luar $Z=1,96$

Distribusi Miring

Tidak semua sampel terdistribusi normal. Sebagai contoh, beberapa sampel bentuknya miring, dsitribusi ini memiliki distribusi nilai variable rerata yang asimetris sehingga satu arah lebih panjang dari arah lainnya. Kemiringan biasanya disebabkan oleh sejumlah pencilan (angka ekstrem) yang signifikan. Jika pencilan berada di sisi kanan, kemringannya positif; jika pencilan berada di sisi kiri, kemringannya negative. Dalam distribusi miring, rerata, median, dan modulusnya tidak sama. Ingat hasil ujian seorang siswa yang bernilai 48 sementara tiga siswa lainnya mendapat nilai lebih tinggi? Itulah contoh kemiringan positif akibat pencilan.

1. $\mu = Md = Mo$ \Rightarrow kurvanya simetris
2. $\mu < Md < Mo$ \Rightarrow bentuk kurva miring ke kiri
3. $Mo < Md < \mu$ \Rightarrow bentuk kurva miring ke kanan



Rangkuman

Ketika terdapat data skala- nominal, satu-satunya ukuran kecendrungan tengah yang dapat digunakan adalah modus. Modus adalah nilai/ukuran yang paling sering muncul dalam distribusi. Distribusi bimodal memiliki dua nilai yang paling sering muncul sehingga ada dua modus. Lebih lanjut, distribusi multimodal memiliki lebih dari dua modus.

Median dapat digunakan pada data ordinal, interval, ataupun data rasio dan dapat ditemukan dengan mengurutkan nilai dari yang terkecil sampai terbesar dan menandai nilai yang ditengah. Untuk dapat interval atau rasio, rerata atau rata-rata dapat dihitung juga. Rerata (mean) merupakan ukuran kecendrungan tengah yang paling dikenal dan dapat ditemukan dengan menjumlahkan nilai dibagi dengan total jumlah observasi.

Terkait penyebaran data, ada dua istilah yang penting

1. Kisaran/rentang atau selisih/jarak antara nilai maksimum dan nilai minimum dalam distribusi,
2. Deviasi standar, atau selisih rata-rata nilai dalam distribusi dari titik tengah

Konsep favorit dalam statistic adalah distribusi normal simetris, sering digambarkan dengan kurva lonceng. Jika datanya sesuai untuk jenis distribusi ini, kurva ini akan mempermudah analisis. Beberapa sampel memiliki kemiringan : yaitu sampel-sampel itu memiliki distribusi nilai variable rerata yang asimetris sehingga satu arah lebih besar dari arah lainnya. Jika terdapat jenis distribusi sampel ini, diperlukan penyesuaian sebelum menerapkan beberapa teknik analisis yang sederhana.

Daftar referensi :

Sutanto, P. H., Sabri. L., (2013). Statistik kesehatan. Jakarta: Rajawali Press

Heavey E. (2011). Statistik Keperawatan: Pendekatan Praktik. Jakarta: EGC.