

Bentuk Distribusi Data

Setiap nilai dalam sekumpulan data memiliki frekuensi kemunculan yang berbeda-beda; nilai ekstrem biasanya muncul lebih jarang, sedangkan nilai yang mendekati *mean* lebih sering muncul. Pola ini membentuk **distribusi data**, yang jika digambarkan dalam grafik seperti poligon frekuensi, akan membentuk kurva. Pada distribusi yang ideal, yaitu **distribusi normal**, kurva berbentuk lonceng (*bell-shaped*) dan **simetris** di sekitar nilai ukuran pemusatan data seperti *mean* atau median. Namun, pada kenyataannya, data bisa saja miring (*skewed*) jika nilai-nilai terkonsentrasi di satu sisi, atau memiliki kurtosis jika kurvanya lebih runcing atau lebih datar dari distribusi normal. Pemahaman tentang bentuk distribusi ini penting karena banyak analisis statistik, terutama yang bersifat inferensial, mengasumsikan bahwa data terdistribusi normal. Untuk itu, bagian-bagian berikut akan membahas lebih lanjut mengenai **kurva normal**, **skewness**, **kurtosis**, dan **z-score** sebagai alat untuk menstandarkan data.

Kurva Normal

Seperti telah disampaikan sebelumnya, distribusi data dapat dikatakan normal jika kurvanya membentuk lonceng dan simetris. Pada distribusi data yang demikian, frekuensi dari setiap nilai **data tersebar secara simetris** antara yang terletak di atas tendensi sentral dan yang di bawah tendensi sentral. Data tersebar 50% di atas dan 50% di bawah tendensi sentral, dengan frekuensi yang semakin mengecil pada nilai data yang semakin ekstrem, baik yang kecil maupun yang besar (lihat Figure 1).

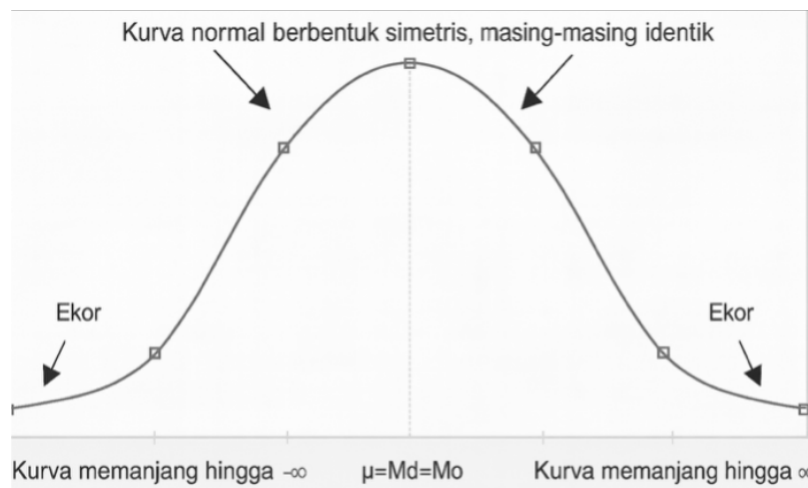


Figure 1: Bentuk kurva normal

Pada kurva normal, tiga pengukuran data terpusat (*mean*, median, dan modus) berada pada satu titik atau nilai yang relatif sama. Dari Figure 1 dapat lebih terlihat jelas bahwa 50% data

tersebar di atas tendensi sentral dan 50% data tersebar di bawah tendensi sentral, di mana semakin jauh dari tendensi sentral maka semakin sedikit frekuensinya.

Untuk menjawab masalah-masalah tertentu, kita perlu mencari besaran frekuensi sebuah nilai atau cakupan nilai berdasarkan sebaran data yang ada. Misalnya, kita ingin memperkirakan ada berapa jumlah mahasiswa yang nilai ujiannya antara 70 dan 85, di mana nilai $mean = 80$ $SD = 5$, dan $N = 250$. Dengan memanfaatkan informasi standar deviasi (s), proporsi frekuensi data pada distribusi normal dapat kita hitung berdasarkan luas area dalam kurva. Figure 2 menunjukkan besaran proporsi frekuensi pada data dilihat dari luas area berdasarkan nilai standar deviasi.

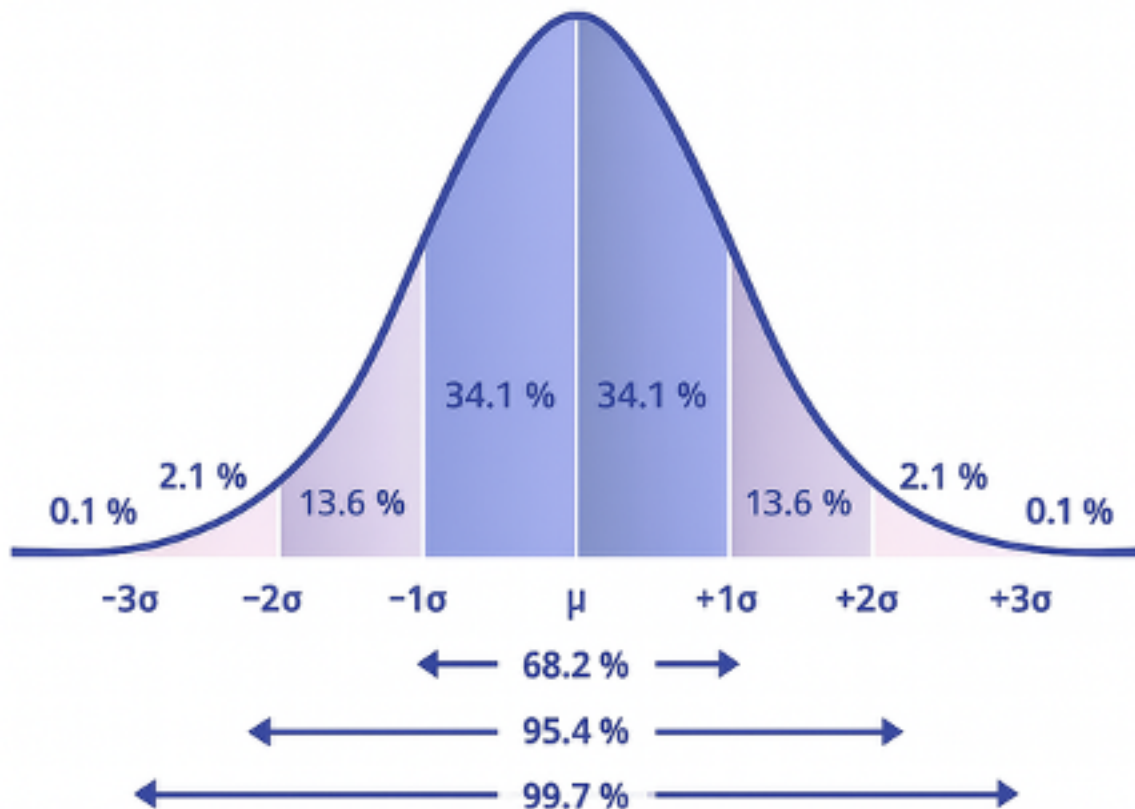


Figure 2: Proporsi sebaran data pada kurva normal

Dari kurva normal yang ditampilkan pada Gambar 3.3, kita dapat mengetahui bahwa setiap belahan kurva dibagi menjadi 4 area berdasarkan standar deviasi, yaitu (1) $mean - +/- 1 SD$ (34,1%), 1 $SD - 2 SD$ (13,6%), 2 $SD - 3 SD$ (2,1%), dan $> 3 SD$ (0,1%). Artinya, proporsi terbanyak dari data adalah nilai-nilai yang mendekati $mean$, dan sebaliknya, semakin jauh dari $mean$ maka proporsinya semakin kecil, terutama untuk nilai-nilai ekstrem ($> +/- 3 SD$).

Proporsi frekuensi data dengan nilai antara $mean$ hingga 1 SD di bawah $mean$ adalah sebesar

34,1% dari keseluruhan data, begitu juga dengan yang 1 *SD* di atas *mean*. Selanjutnya, jika kita ingin mengetahui proporsi frekuensi data antara *mean* hingga 2 *SD* di atas *mean*, maka kita menjumlahkan proporsi (*mean* - 1 *SD*) + (1 *SD* - 2 *SD*) = 34,1% + 13,6% = 47,7%.

Dengan demikian, untuk menjawab pertanyaan sebelumnya (proporsi mahasiswa yang memperoleh nilai ujian 70 (X_1) - 85 (X_2) dengan *mean* = 80, *SD* = 5, dan N = 250), maka kita menghitungnya dengan mencari besaran *SD* dari batas bawah dan atas rentang skor:

$$X_1 - \text{mean} = 70 - 80 = -10; \text{ karena } SD = 5, \text{ maka skor } 70 = -2SD$$

$$X_2 - \text{mean} = 85 - 80 = 5; \text{ karena } SD = 5, \text{ maka skor } 85 = 1SD$$

Maka, proporsi $X_1 - X_2$ = (proporsi -2SD - mean) + (proporsi mean - 1SD)

$$= (13,6\% + 34,1\%) + (34,1\%)$$

$$= 81,8\%$$

Sehingga, jumlah mahasiswa dengan skor antara 70 - 85 = 81,8% \times 250 orang = 204 orang (pembulatan).

Skewness

Data pada sampel yang diambil dari populasi penelitian tidak selalu terdistribusi secara normal. Sebaran yang tidak normal ini seringkali terjadi pada sekumpulan data yang memiliki nilai ekstrem (*outlier*) yang proporsinya cukup jauh melebihi data yang tersebar secara normal. Besarnya *outlier* ini membuat sebaran data seolah-olah terpusat di bawah nilai tengah, atau sebaliknya terpusat di atas nilai tengah.

Jika data yang tersebar secara normal membentuk kurva berbentuk lonceng yang simetris, maka **data yang tidak normal** tersebut bisa membentuk **kurva yang miring** atau juling (*skewed*). Arah kemiringan (*skewness*) kurva bergantung pada besaran frekuensi data yang menyimpang dari nilai tengah. Jika frekuensi nilai yang berada di bawah *mean* jauh lebih kecil daripada frekuensi nilai di atas *mean*, maka puncak kurva akan condong ke kanan, yang disebut sebagai *skew* negatif (Figure 3a). Sebaliknya, jika frekuensi nilai yang berada di bawah *mean* jauh lebih besar daripada frekuensi nilai di atas *mean*, maka puncak kurva akan condong ke kiri, yang disebut sebagai *skew* positif (Figure 3b).

Terdapat dua cara yang umum digunakan untuk mengetahui arah kemiringan (*skewness*) dari sebuah distribusi data, yaitu dengan (a) menggunakan rumus Pearson, dan (b) membandingkan nilai-nilai ukuran pemusatan data.

1. Rumus *Pearson*:

$$\text{Skewness} = 3 \times \frac{\text{Mean} - \text{Median}}{SD}$$

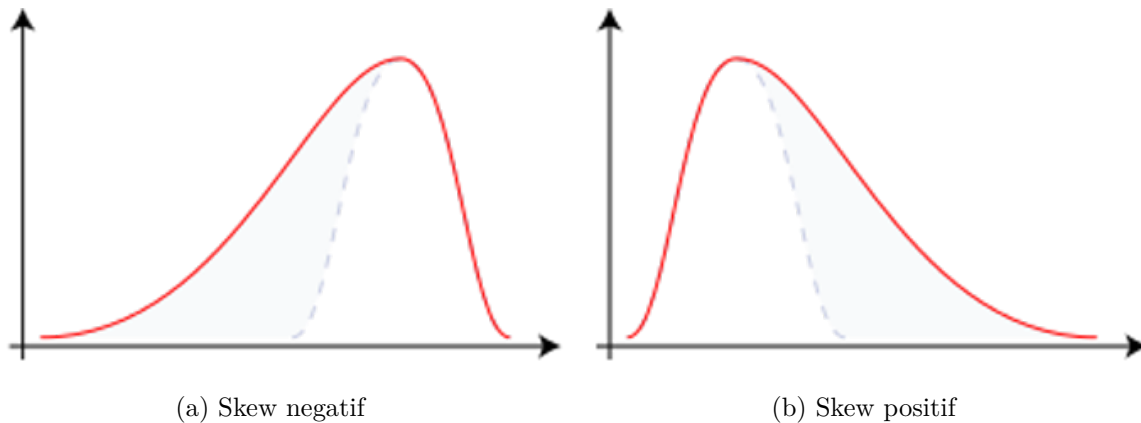


Figure 3: Distribusi skewed

Jika hasil penghitungan nilai *skewness* adalah positif, maka disebut dengan kurva *skewed* positif. Jika hasil hitungannya negatif, maka disebut dengan kurva *skewed* negatif. Apabila hasil hitungannya (mendekati) nol, maka disebut dengan kurva normal. Tidak ada angka pasti, seberapa dekat dengan nol dapat disebut sebagai kurva normal, tetapi kesepakatan umum adalah $\pm 0,5$.

2. Membandingkan nilai ukuran pemusatan data

Jika *mean* lebih kecil dari median, maka bentuk distribusinya adalah *skewed* negatif. Sebaliknya, jika *mean* lebih lebih besar dari median, maka bentuk distribusinya adalah *skewed* positif. Namun, jika *mean*, median, dan modus berada pada satu titik yang relatif sama atau berdekatan, maka bentuk distribusi datanya adalah normal (Figure 4).

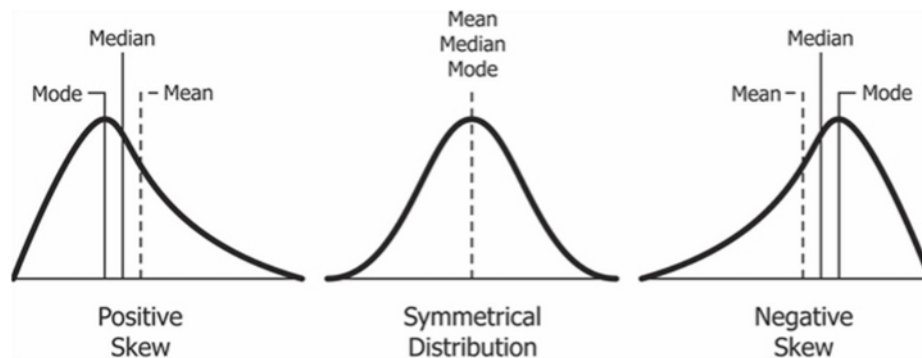


Figure 4: Skewness data berdasarkan lokasi tendensi sentral

Kurtosis

Kurtosis adalah seberapa **berbeda bentuk kurva dari kurva normal**, dilihat dari seberapa tebal-tipis bagian ekor dari kurva tersebut. Selain dilihat dari ekor kurva, kurtosis sebenarnya juga menunjukkan seberapa tajam bentuk puncak dari kurva. Terdapat tiga bentuk kurtosis, yaitu normal, ekor panjang (*heavy-tailed*) dan ekor pendek (*light-tailed*) (Figure 5).

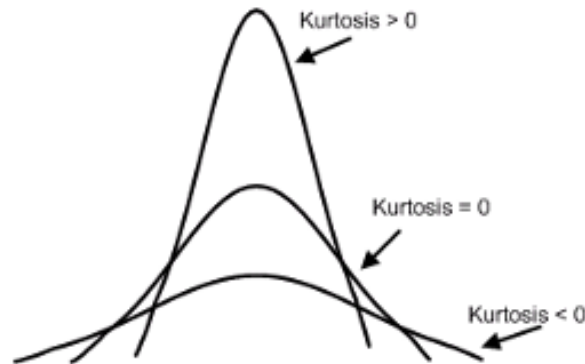


Figure 5: Visualisai kurtosis dalam distribusi data

Kurtosis pada distribusi data yang normal adalah ketika ekor kurva tersebar secara proporsional dan puncak kurva tidak terlalu curam maupun terlalu landai. Jika nilai kurtosis > 0 , maka kurva akan memiliki ekor yang pendek dan puncak yang curam, menandakan data menumpuk di nilai-nilai sekitar nilai tengah. Sebaliknya, jika nilai kurtosis < 0 maka ekor kurva akan memanjang dan puncaknya melandai, yang artinya hampir seluruh nilai memiliki frekuensi atau jumlah kejadian yang serupa. Kurva yang menunjukkan baik ekor panjang maupun ekor pendek menandakan bahwa data tidak terdistribusi secara normal.

Z-Score

z-score, yang dikenal juga dengan nama **skor baku**, menunjukkan **posisi sebuah skor** dibandingkan dengan rata-rata dalam satuan standar deviasi. Dengan kata lain, *z-score* menyatakan posisi relatif suatu nilai (X) dalam distribusi data. $z = \text{nol}$ berarti nilai tersebut tepat di *mean*, *z-score* negatif menandakan bahwa nilainya lebih rendah dari *mean*, dan *z-score* positif berarti nilainya di atas *mean*. *z-score* dapat dihitung dengan cara:

$$z = \frac{X - \bar{x}}{\text{SD}}$$

Tujuan utama penggunaan *z-score* adalah untuk menstandarkan data, sehingga memungkinkan perbandingan antar nilai dari distribusi yang berbeda. Sebagai analogi, kita tidak

bisa menilai mana yang memiliki nilai yang lebih tinggi antara nilai dua mata uang yang berbeda (misalnya, 10 Rupee India dengan 10 Baht Thailand) tanpa mengkonversinya ke dalam satuan mata uang yang sama atau standar.

Dalam hal data, dua nilai dari distribusi yang berbeda (misalnya, nilai 8 pada subtes aritmatika dan nilai 8 pada subtest logika numerikal) bisa jadi memiliki posisi yang berbeda dalam distribusi data masing-masing, sehingga tidak dapat dibandingkan secara langsung. Oleh karena itu, kedua nilai tersebut perlu ditransformasi ke *z-score* agar memiliki satuan yang sama untuk dapat dibandingkan.

z-score juga bermanfaat dalam mendeteksi *outlier*, serta dalam berbagai analisis statistik lanjutan seperti uji hipotesis dan analisis distribusi normal. Karena *z-score* mengubah data ke skala yang seragam, ia menjadi alat penting dalam interpretasi dan generalisasi hasil penelitian.