

Distribusi Peluang

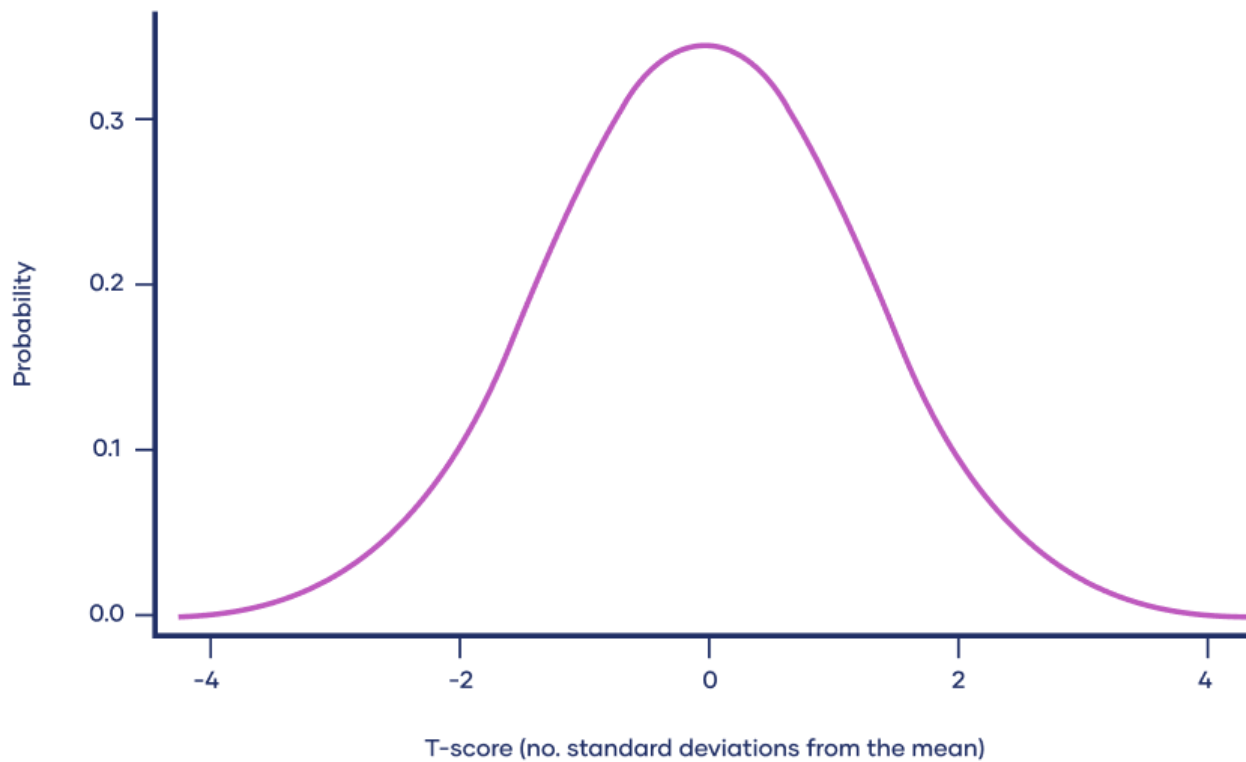
EGI SAFITRI, S.MAT., M.SI

Distribusi-t / Student

Distribusi t /Student

Distribusi t , juga dikenal sebagai distribusi t Student, adalah cara mendeskripsikan data yang mengikuti kurva lonceng ketika diplot pada grafik, dengan jumlah observasi terbanyak mendekati [rata-rata](#) dan lebih sedikit observasi di bagian ekor.

Ini adalah jenis [distribusi normal](#) yang digunakan untuk ukuran sampel yang lebih kecil, dimana [varians](#) datanya tidak diketahui.



Dalam statistik, distribusi t paling sering digunakan untuk:

- Temukan nilai kritis untuk interval kepercayaan ketika data mendekati distribusi normal.
- Temukan [nilai \$p\$](#) yang sesuai dari [uji statistik](#) yang menggunakan distribusi $-t$ ([uji- \$t\$](#) , [analisis regresi](#)).

Apa itu Distribusi-t?

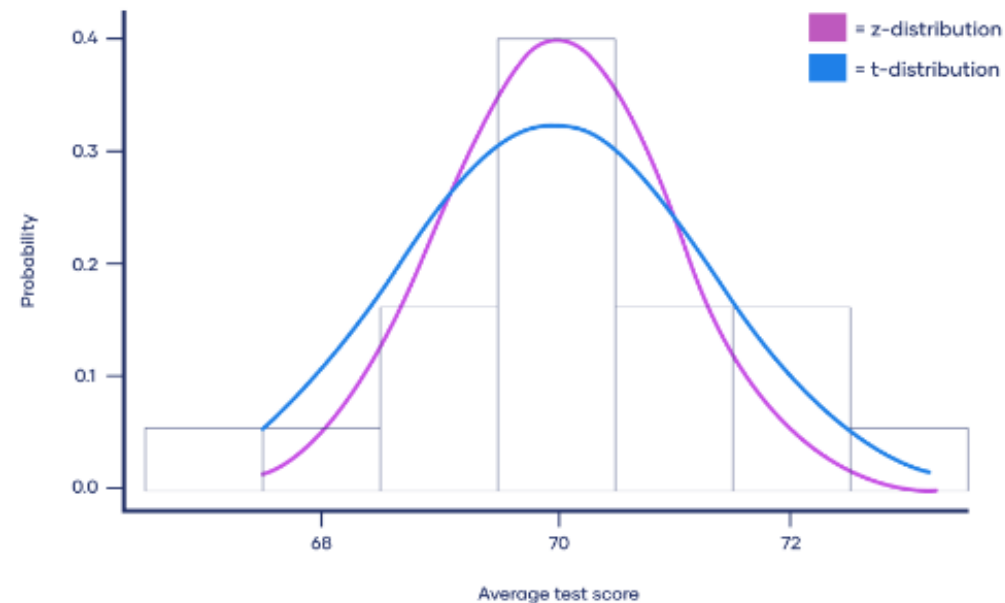
Distribusi t adalah jenis distribusi normal yang digunakan untuk ukuran sampel yang lebih kecil. Data yang berdistribusi normal membentuk bentuk lonceng ketika diplot pada grafik, dengan lebih banyak observasi di dekat mean dan lebih sedikit observasi di bagian ekor.

Distribusi t digunakan jika data berdistribusi *mendekati* normal, artinya data mengikuti bentuk lonceng tetapi varians populasinya tidak diketahui. Varians dalam distribusi $-t$ diperkirakan berdasarkan **derajat kebebasan** kumpulan data (jumlah total observasi dikurangi 1).

Ini adalah bentuk **distribusi normal standar** yang lebih konservatif, juga dikenal sebagai distribusi z . Artinya distribusi tersebut memberikan probabilitas yang lebih rendah pada pusat dan probabilitas yang lebih tinggi pada bagian ekor dibandingkan dengan distribusi normal standar.

Contoh: distribusi-t vs distribusi-z

Jika Anda mengukur rata-rata nilai tes dari sampel yang hanya terdiri dari 20 siswa, Anda harus menggunakan **distribusi t** untuk memperkirakan interval kepercayaan di sekitar mean. Jika Anda menggunakan **z-distribution**, interval kepercayaan Anda akan menjadi tepat secara artifisial.

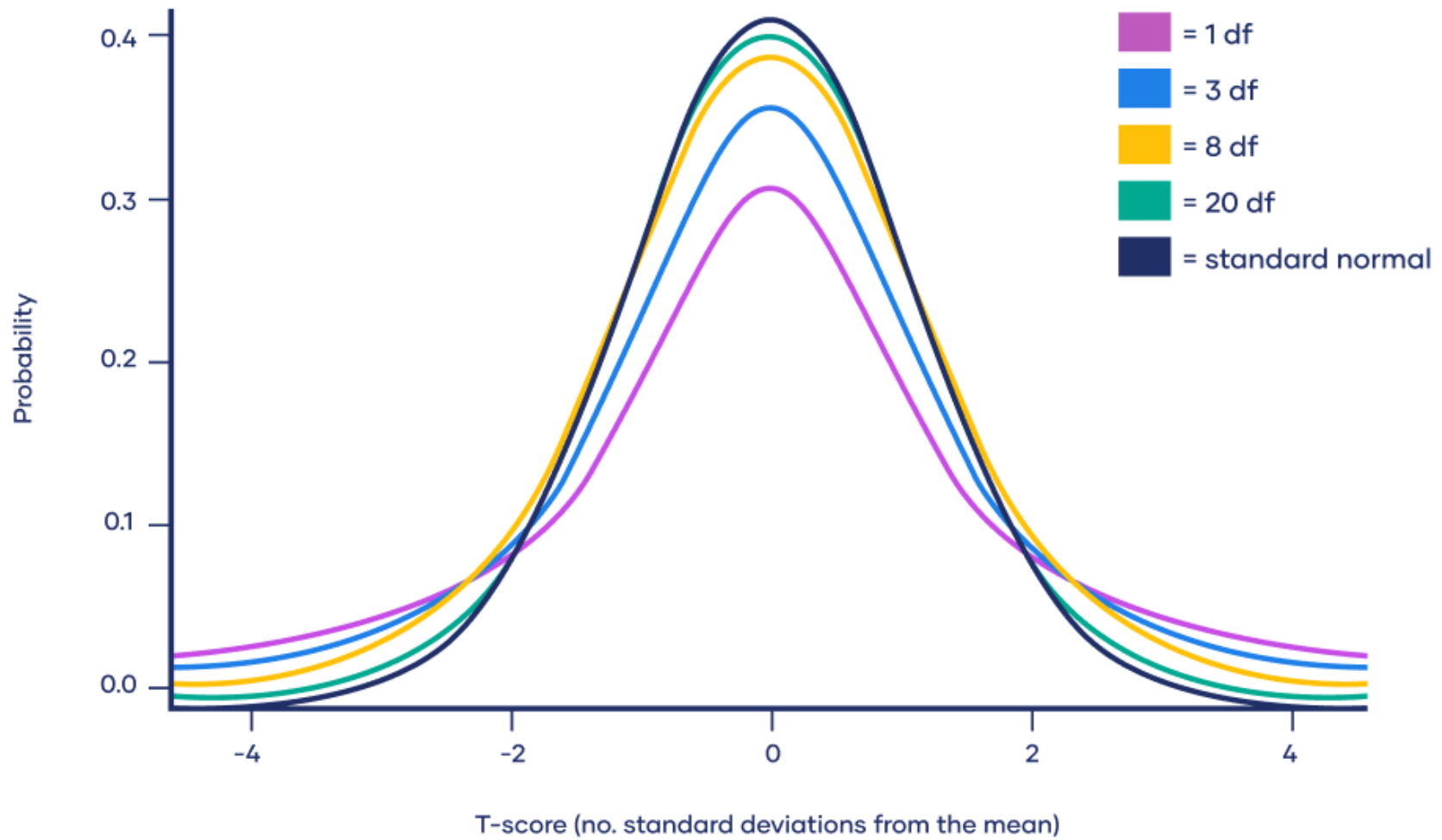


T-distribusi dan normal standar

Dengan bertambahnya **derajat kebebasan** (jumlah observasi dikurangi 1), maka distribusi t akan semakin mendekati pencocokan dengan distribusi normal standar alias distribusi z hingga hampir identik.

Di atas 30 derajat kebebasan, distribusi t kira-kira sama dengan distribusi z . Oleh karena itu, distribusi- z dapat digunakan sebagai pengganti distribusi $-t$ dengan ukuran sampel yang besar.

Distribusi z lebih disukai daripada distribusi t ketika membuat estimasi statistik karena memiliki varian yang diketahui. Ini dapat membuat perkiraan yang lebih tepat daripada distribusi $-t$, yang variansnya diperkirakan menggunakan derajat kebebasan data.



t-distribusi dan t-skor

Skor $-t$ adalah jumlah simpangan baku dari mean dalam distribusi $-t$. Anda biasanya dapat mencari t -score di t -table , atau dengan menggunakan kalkulator t -score online.

Dalam statistik, t -score terutama digunakan untuk menemukan dua hal:

1. Batas atas dan bawah interval kepercayaan ketika data mendekati distribusi normal.
2. Nilai p dari statistik uji untuk uji t dan uji regresi.

t-skor dan interval kepercayaan

Interval kepercayaan menggunakan skor- t untuk menghitung batas atas dan bawah interval prediksi. Skor t yang digunakan untuk menghasilkan batas atas dan bawah juga dikenal sebagai **nilai kritis t** , atau t^* .

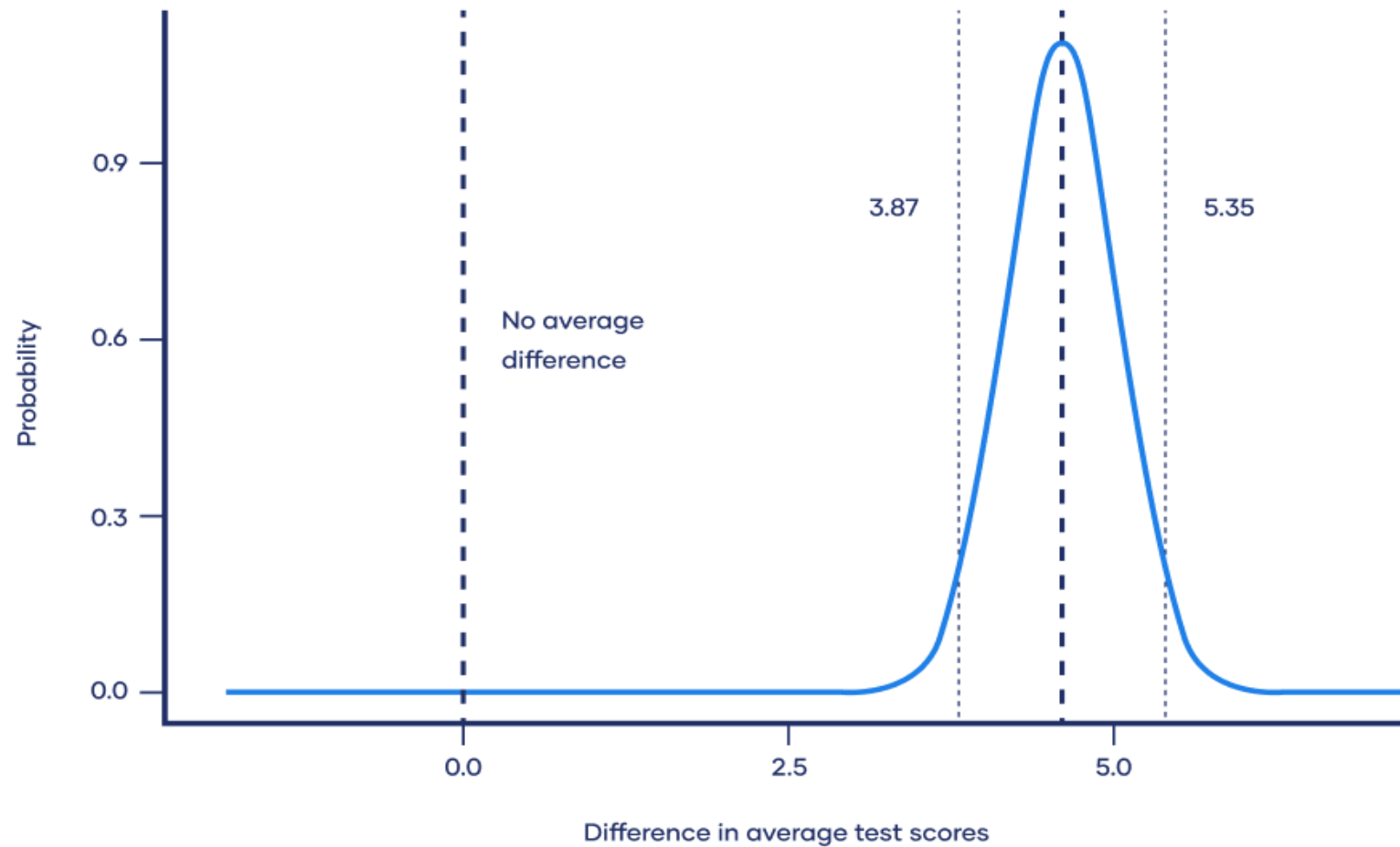
Contoh: Interval Kepercayaan

Anda telah mengambil sampel 20 siswa dari dua kelas berbeda untuk memperkirakan nilai rata-rata tes standar dan ingin mengetahui apakah ada perbedaan antara kedua kelompok.

Dengan menggunakan uji -t dua sisi , Anda menghasilkan perkiraan perbedaan antara dua kelas dan interval kepercayaan di sekitar perkiraan tersebut. Dari uji- t diketahui selisih rata-rata skor antara kelas 1 dan kelas 2 adalah 4,61, dengan interval kepercayaan 95% sebesar 3,87 hingga 5,35.

Karena interval kepercayaan tidak melampaui nol, dan bahkan cukup jauh dari nol, kecil kemungkinannya bahwa perbedaan nilai tes ini terjadi berdasarkan **hipotesis nol** yang menyatakan tidak ada perbedaan antar kelompok.

T-distribution of the difference between average test scores



t-skor dan p-nilai

Uji statistik menghasilkan **statistik uji** yang menunjukkan seberapa jauh data Anda dari hipotesis nol uji statistik. Mereka kemudian menghitung **nilai p** yang menggambarkan kemungkinan terjadinya data Anda jika hipotesis nol benar.

Statistik uji untuk **uji t** dan **uji regresi** adalah **t -score**. Meskipun sebagian besar program statistik akan secara otomatis menghitung nilai **p** yang sesuai untuk skor **t** , Anda juga dapat mencari nilai dalam tabel **t** , menggunakan derajat kebebasan dan skor **t** untuk menemukan nilai **p** .

Skor **t** yang menghasilkan nilai **p** di bawah ambang batas **signifikansi statistik** dikenal sebagai nilai kritis **t** , atau **t ***.

Contoh: nilai-p

Uji t dua sisi terhadap selisih nilai tes menghasilkan nilai *t* sebesar 12,79. Artinya selisih mean kelompok adalah 12,79 standar deviasi jauh dari mean distribusi hipotesis nol.

Derajat kebebasannya adalah 38 ($n-1$ untuk setiap kelompok). Mencarinya di tabel *-t* (atau menghitungnya di program statistik favorit Anda) Anda menemukan nilai $p < 0,001$.

Temuan ini, seperti temuan dari interval kepercayaan, menunjukkan bahwa Anda tidak mungkin menemukan perbedaan sebesar ini jika perbedaan sebenarnya dalam nilai rata-rata tes adalah nol.

FAQ

Apa itu distribusi-t?



Distribusi - t adalah cara untuk mendeskripsikan sekumpulan pengamatan yang sebagian besar pengamatannya mendekati **nilai rata-rata** , dan sisa pengamatan membentuk ekor di kedua sisinya. Ini adalah jenis **distribusi normal** yang digunakan untuk ukuran sampel yang lebih kecil, dimana **varians** datanya tidak diketahui.

Distribusi t membentuk kurva lonceng ketika diplot pada grafik. Hal ini dapat dijelaskan secara matematis menggunakan mean dan **deviasi standar** .

FAQ

Apa perbedaan antara distribusi t dan distribusi normal standar?

Distribusi *t* memberikan lebih banyak kemungkinan untuk observasi pada bagian ekor distribusi dibandingkan [distribusi normal standar](#) (alias [distribusi z](#)).

Dengan cara ini, distribusi *t* lebih konservatif daripada distribusi normal standar: untuk mencapai tingkat kepercayaan atau [signifikansi statistik](#) yang sama , Anda perlu menyertakan data yang lebih luas.

FAQ

Apa itu skor-t?



Skor $-t$ (alias nilai $-t$) setara dengan jumlah deviasi standar yang menjauhi rata-rata distribusi $-t$.

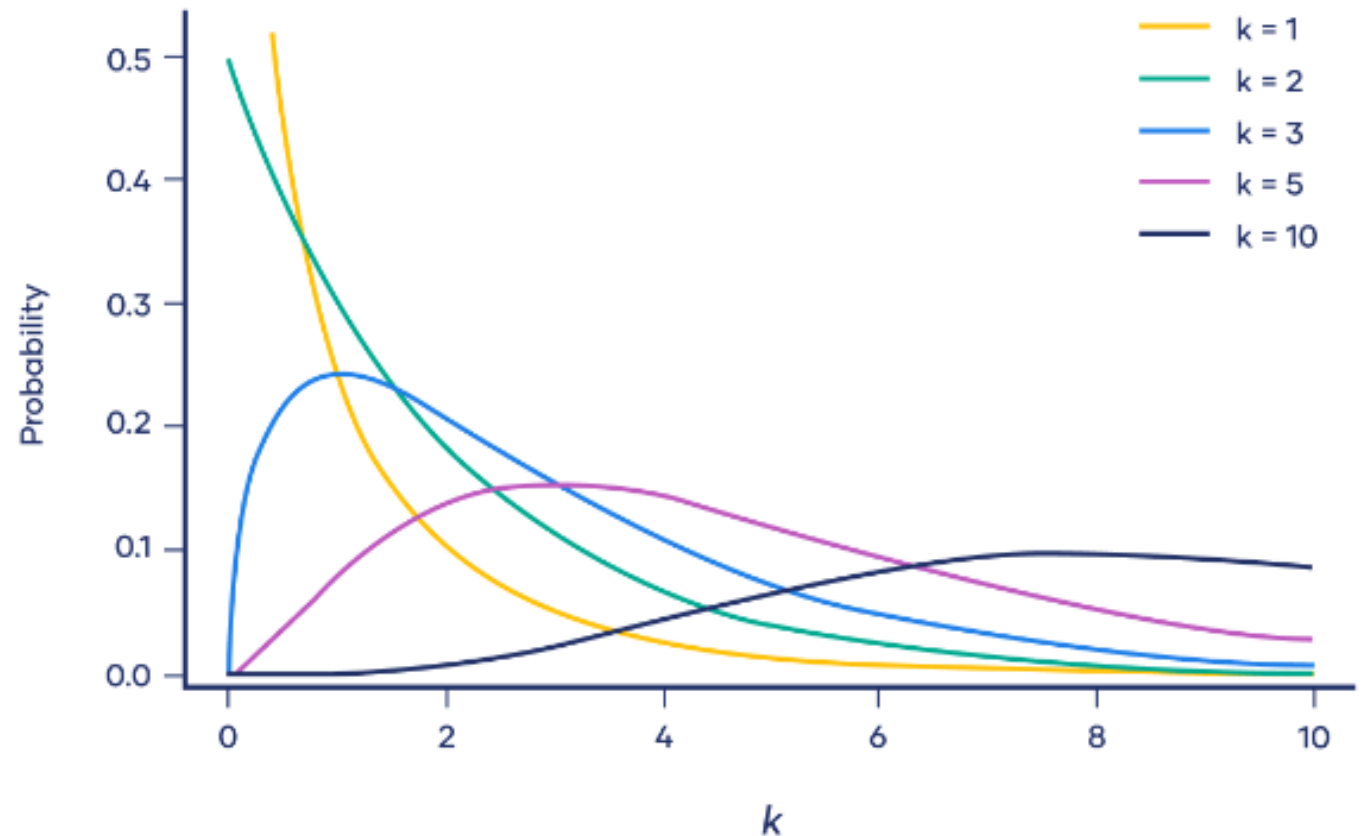
T -score adalah statistik uji yang digunakan dalam uji t dan uji regresi. Ini juga dapat digunakan untuk menggambarkan seberapa jauh dari mean suatu observasi ketika data mengikuti distribusi $-t$.

Distribusi Chi-square

Pengertian

Distribusi chi-kuadrat (χ^2) adalah [distribusi probabilitas](#) kontinu yang digunakan dalam banyak uji hipotesis.

Bentuk distribusi chi-kuadrat ditentukan oleh parameter k . Grafik di samping menunjukkan contoh distribusi chi-kuadrat dengan nilai k yang berbeda



Distribusi Chi-square?

Distribusi chi-kuadrat (χ^2) adalah kelompok distribusi probabilitas kontinu. Mereka banyak digunakan dalam [uji hipotesis](#), termasuk uji kesesuaian chi-kuadrat dan uji independensi chi-kuadrat.

Bentuk distribusi chi-kuadrat ditentukan oleh [parameter \$k\$](#) yang mewakili [derajat kebebasan](#).

Sangat sedikit pengamatan di dunia nyata yang mengikuti distribusi chi-kuadrat. Tujuan utama distribusi chi-kuadrat adalah pengujian hipotesis, bukan menggambarkan distribusi dunia nyata.

Sebaliknya, sebagian besar distribusi lain yang banyak digunakan, seperti [distribusi normal](#) atau [distribusi Poisson](#), dapat menggambarkan hal-hal berguna seperti berat lahir bayi baru lahir atau kasus penyakit per tahun.

Chi-Square vs Normal Standar

Distribusi chi-kuadrat berguna untuk pengujian hipotesis karena hubungannya yang erat dengan [distribusi normal standar](#). Distribusi normal standar, yaitu distribusi normal dengan rata-rata nol dan varians satu, merupakan inti dari banyak uji dan teori [statistik penting](#).

Bayangkan mengambil [sampel acak](#) dari distribusi normal standar (Z). Jika Anda mengkuadratkan semua nilai dalam sampel, Anda akan mendapatkan distribusi chi-kuadrat dengan $k = 1$.

$$X^2_1 = (Z)^2$$

Chi-Square vs Normal Standar

Sekarang bayangkan mengambil sampel dari dua distribusi normal standar (Z_1 dan Z_2). Jika setiap kali Anda mengambil sampel sepasang nilai, Anda mengkuadratkannya dan menjumlahkannya, Anda akan mendapatkan distribusi chi-kuadrat dengan $k = 2$.

$$X^2_2 = (Z_1)^2 + (Z_2)^2$$

Secara umum, jika Anda mengambil sampel dari k distribusi normal standar independen lalu mengkuadratkan dan menjumlahkan nilainya, Anda akan menghasilkan distribusi chi-kuadrat dengan k derajat kebebasan .

$$X^2_k = (Z_1)^2 + (Z_2)^2 + \dots + (Z_k)^2$$

Statistik Uji

Uji chi-kuadrat adalah uji hipotesis dengan uji statistik yang mengikuti distribusi chi-kuadrat dengan hipotesis nol . Uji chi-kuadrat Pearson merupakan uji chi-kuadrat pertama yang ditemukan dan paling banyak digunakan.

Statistik uji chi-kuadrat Pearson adalah:

Rumus	Penjelasan
$X^2 = \sum \frac{(O-E)^2}{E}$	<p>Di mana</p> <ul style="list-style-type: none">• X^2 adalah statistik uji chi-kuadrat• \sum adalah operator penjumlahan (artinya “ambil jumlah”)• O adalah frekuensi yang diamati• E adalah frekuensi yang diharapkan

Note:

Jika Anda mengambil sampel suatu [populasi](#) berkali-kali dan menghitung statistik uji chi-kuadrat Pearson untuk setiap sampel, statistik uji akan mengikuti distribusi chi-kuadrat jika hipotesis nolnya benar.

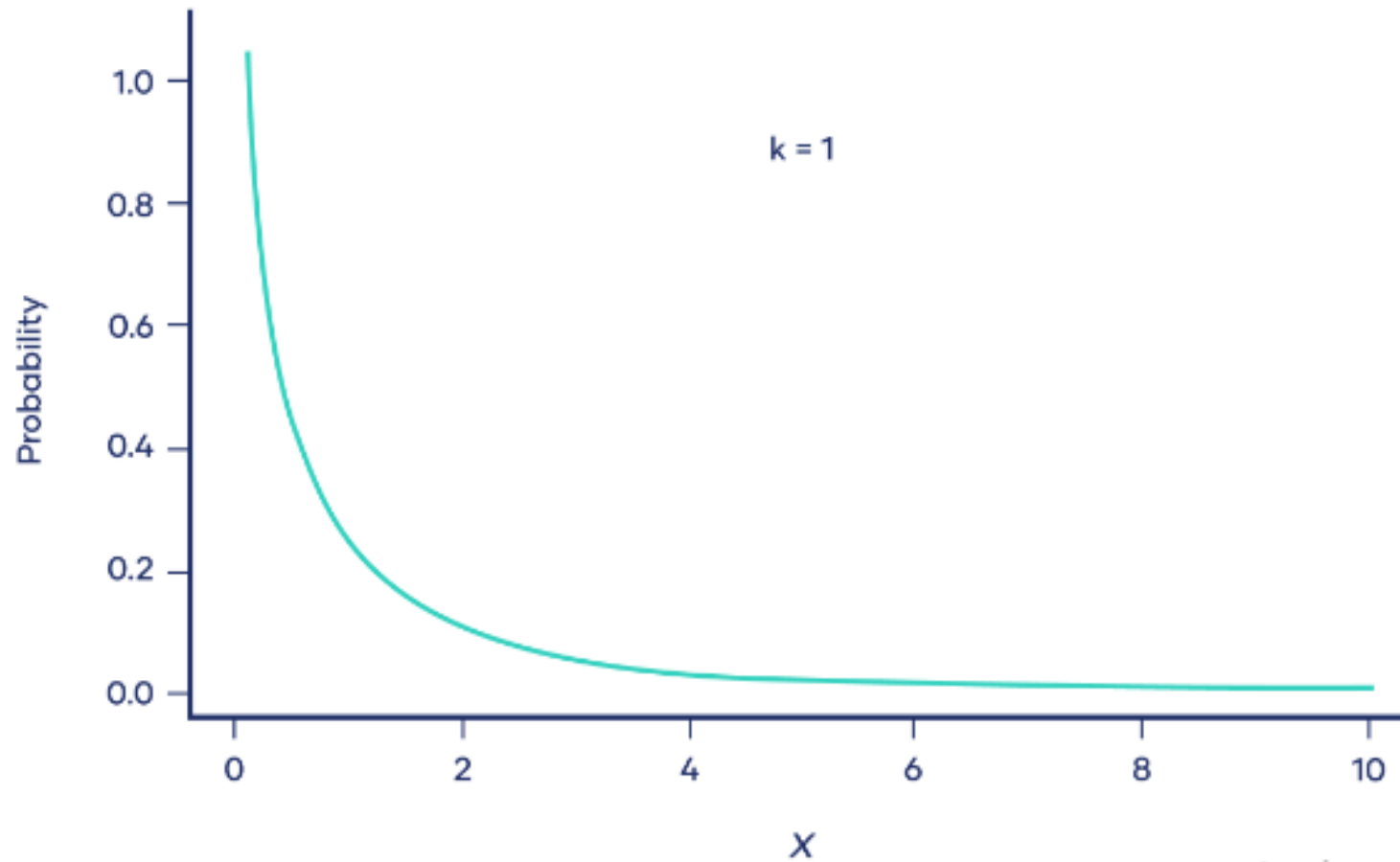
Bentuk Distribusi

Kita dapat melihat bagaimana bentuk distribusi chi-kuadrat berubah seiring dengan meningkatnya derajat kebebasan (k) dengan melihat grafik fungsi kepadatan probabilitas chi-kuadrat. Fungsi kepadatan probabilitas adalah fungsi yang menggambarkan **distribusi probabilitas kontinu**.

Ketika k adalah satu atau dua

Jika k bernilai satu atau dua, distribusi chi-kuadrat berbentuk kurva seperti huruf “J” terbalik. Kurva awalnya tinggi dan kemudian turun, artinya ada kemungkinan besar bahwa X^2 mendekati nol.

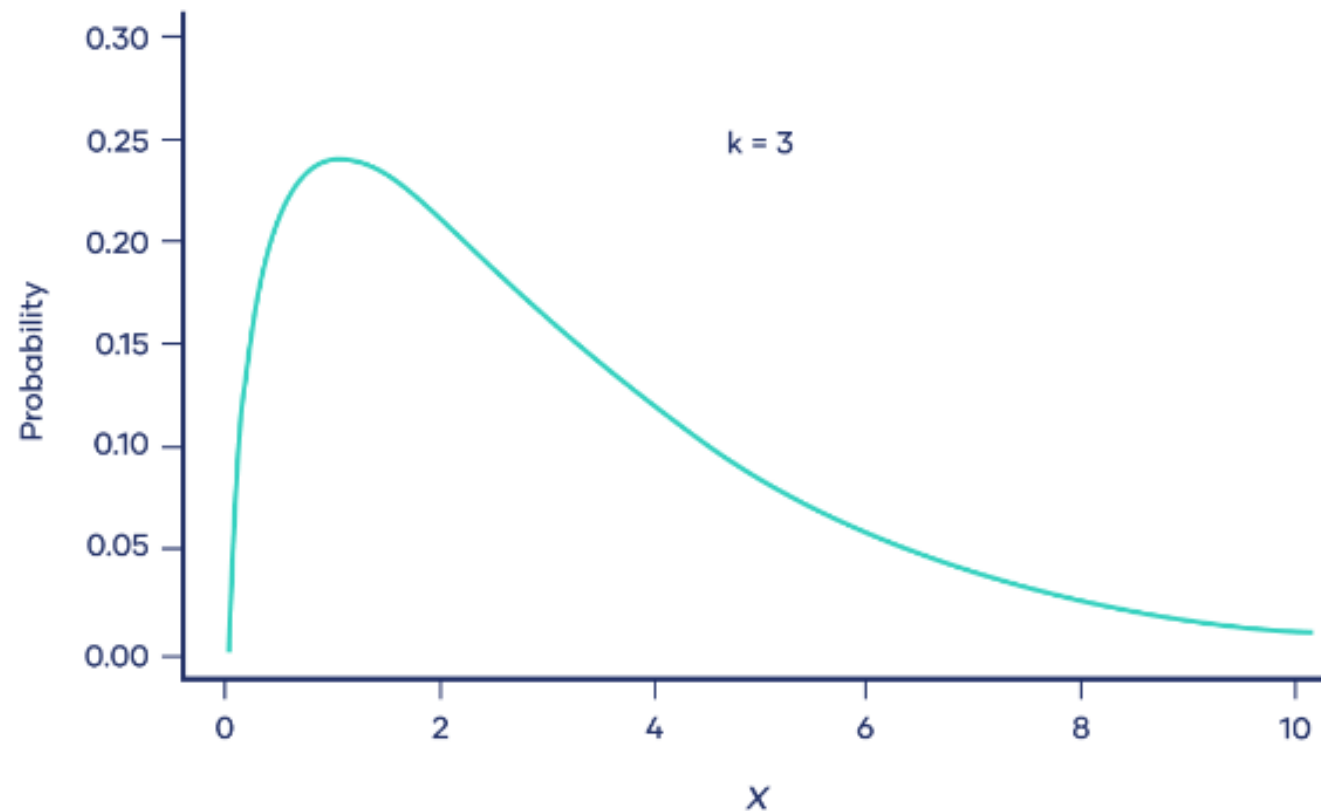
Bentuk Distribusi



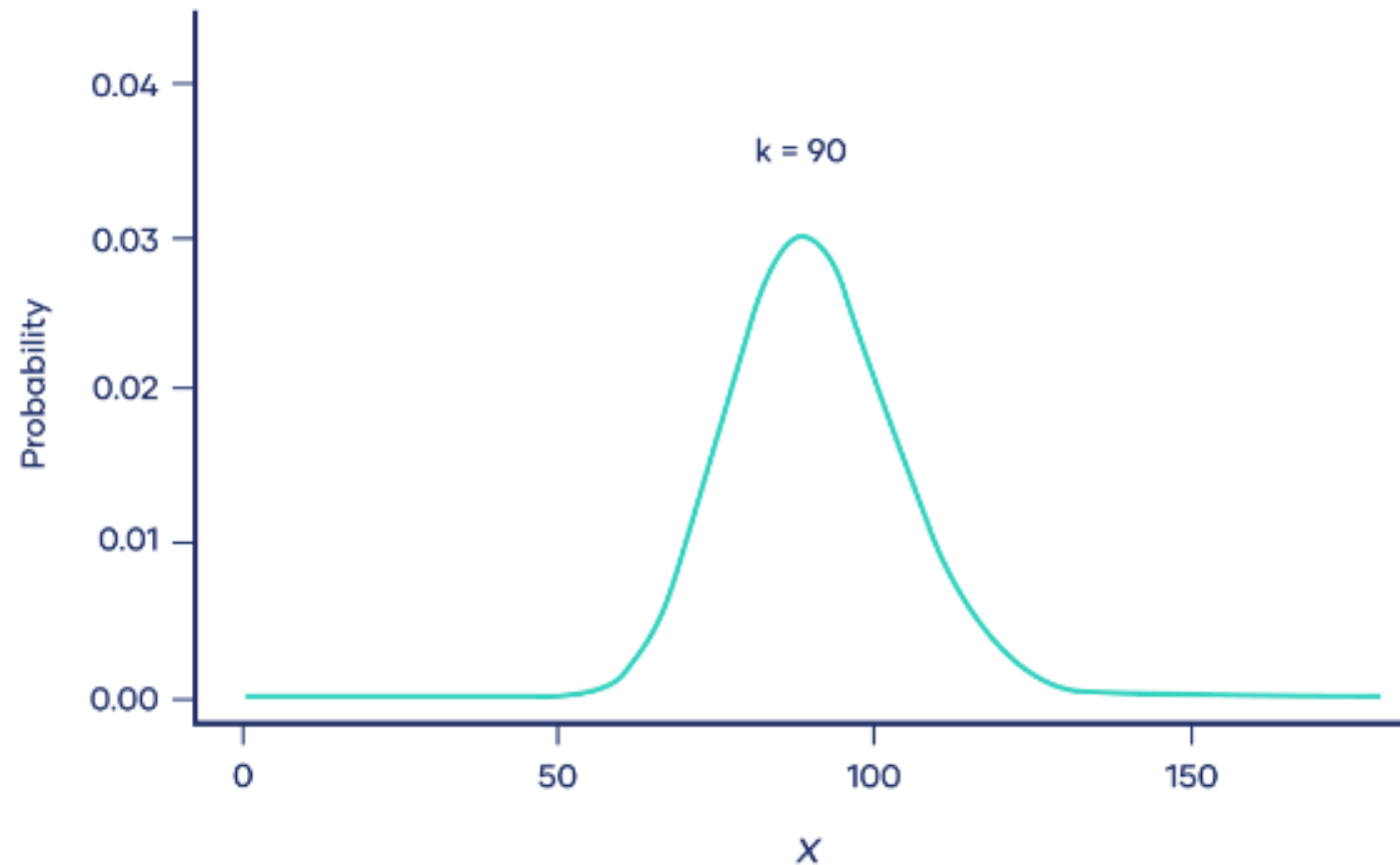
Ketika k lebih besar dari dua

Jika k lebih besar dari dua, distribusi chi-kuadrat berbentuk punuk. Kurvanya dimulai dari rendah, meningkat, dan kemudian menurun lagi. Ada kemungkinan kecil bahwa X^2 sangat mendekati atau sangat jauh dari nol. Nilai X^2 yang paling mungkin adalah $X^2 - 2$.

Ketika k hanya sedikit lebih besar dari dua, distribusi pada sisi kanan puncaknya jauh lebih panjang dibandingkan sisi kirinya (yaitu, distribusinya sangat **condong ke kanan**).



Ketika k bertambah, distribusinya terlihat semakin mirip dengan [distribusi normal](#). Faktanya, jika k bernilai 90 atau lebih besar, distribusi normal merupakan perkiraan yang baik untuk distribusi chi-kuadrat.



Sifat Distribusi Chi-Square

Distribusi chi-kuadrat dimulai dari nol dan berlanjut hingga tak terhingga. Distribusi chi-kuadrat dimulai dari nol karena menggambarkan **jumlah variabel acak yang dikuadratkan**, dan bilangan kuadrat tidak boleh negatif.

Rata-rata (μ) dari distribusi chi-kuadrat adalah derajat kebebasannya, k . Karena distribusi chi-kuadratnya miring ke kanan, maka meannya lebih besar daripada **median** dan modulusnya. Varians distribusi chi-kuadrat adalah $2k$.

Sifat Distribusi Chi-Square

Properti	Nilai
Kontinu atau diskrit	Kontinu
Berarti	k
Mode	$k - 2$ (bila $k > 2$)
Perbedaan	$2k$
Deviasi standar	$\sqrt{2k}$
Jangkauan	0 hingga ∞
Simetri	Asimetris (condong ke kanan), tetapi semakin simetris seiring bertambahnya k .

Contoh penerapan distribusi chi-kuadrat

Distribusi chi-kuadrat muncul dalam banyak uji statistik dan teori. Berikut adalah beberapa penerapan distribusi chi-kuadrat yang paling umum.

Uji chi-kuadrat Pearson

Salah satu penerapan distribusi chi-kuadrat yang paling umum adalah uji chi-kuadrat Pearson.

Uji chi-kuadrat Pearson adalah uji statistik untuk **data kategorikal**. Mereka digunakan untuk menentukan apakah data Anda berbeda secara signifikan dari yang Anda harapkan. Ada dua jenis uji chi-kuadrat Pearson:

- Uji kesesuaian chi-kuadrat
- Uji independensi chi-kuadrat

Contoh

Sebuah perusahaan yang menjual kaos ingin mengetahui apakah semua warna kaosnya sama-sama populer, maka mereka mencatat jumlah penjualan per warna kaos selama satu minggu.

Warna	Frekuensi
Merah	30
Abu-abu	29
Kuning	26
Merah Jambu	33
Hitam	56
Putih	90
Biru	86

Karena total ada 350 penjualan kaos, maka 50 penjualan per warna akan sama persis. Jelas sekali bahwa tidak ada 50 penjualan per warna. Namun, ini hanya sampel satu minggu, jadi diperkirakan jumlahnya akan sedikit tidak seimbang hanya karena kebetulan.

Apakah sampel tersebut memberikan cukup bukti untuk menyimpulkan bahwa frekuensi penjualan kaos memang berbeda antar warna kaos?

Uji kesesuaian chi-kuadrat dapat menguji apakah frekuensi yang diamati berbeda **secara signifikan** dari frekuensi yang sama. Dengan membandingkan statistik uji chi-kuadrat Pearson dengan distribusi chi-kuadrat yang sesuai, perusahaan dapat menghitung probabilitas nilai penjualan kaos ini (atau nilai yang lebih ekstrim) terjadi secara kebetulan.

Varians

Distribusi chi-kuadrat juga dapat digunakan untuk membuat kesimpulan tentang varians suatu populasi (σ^2) atau deviasi standar (σ). Dengan menggunakan distribusi chi-kuadrat, Anda dapat menguji hipotesis bahwa suatu varians populasi sama dengan nilai tertentu dengan menggunakan uji **varians tunggal** atau menghitung interval kepercayaan untuk varians suatu populasi.

Distribusi F

Distribusi chi-kuadrat penting dalam menentukan distribusi F , yang digunakan dalam [ANOVA](#).

Bayangkan Anda mengambil sampel acak dari distribusi chi-kuadrat, lalu membagi sampel tersebut dengan k distribusi tersebut. Selanjutnya, Anda mengulangi proses tersebut dengan distribusi chi-kuadrat yang berbeda. Jika Anda mengambil rasio nilai dari kedua distribusi tersebut, Anda akan mendapatkan distribusi F .