Matematikai alapok és valószínűségszámítás

Normál eloszlás

A normál eloszlás

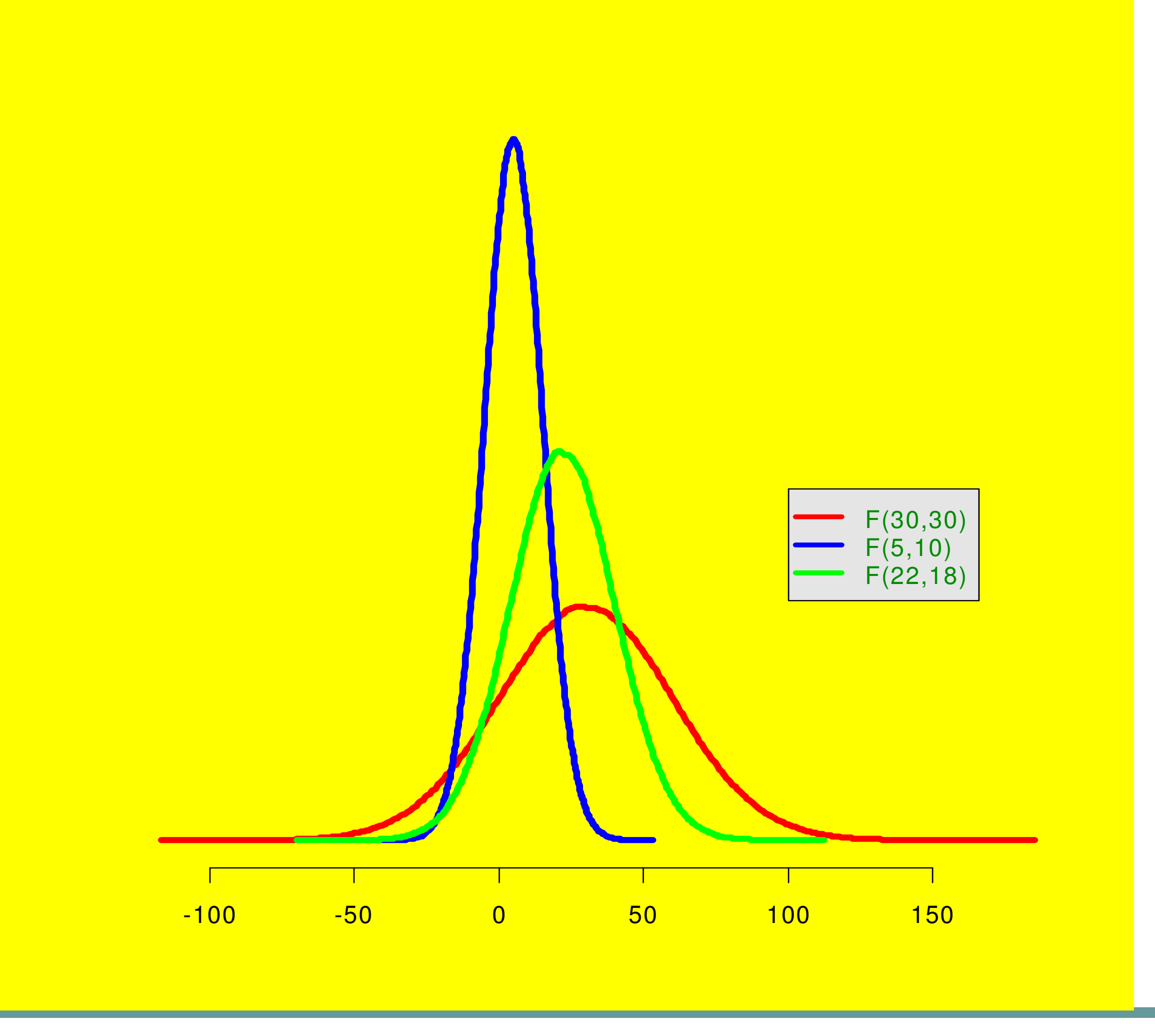
Folytonos változók esetén az eloszlás meghatározása nehezebb, mint diszkrét változók esetén. A változó értékei nem sorolhatóak fel, mert végtelen sok érték lehet, ezért az egyes konkrét értékek valószínűsége rendkívül csekély.

Folytonos változók esetén nem a konkrét értékek valószínűségét, hanem egy adott szakaszra esés valószínűségét adjuk meg, tehát például annak a valószínűségét, hogy az IQ 109.5 és 110.5 közé esik.

Folytonos változók esetén is végtelen sokféle eloszlás képzelhető el, de vannak bizonyos kitüntetett eloszlások, melyek közül a legjelentősebb a normál eloszlás.

A normál eloszlást szokás Gauss-görbének, vagy haranggörbének is nevezni.

Néhány normál eloszlás sűrűségfüggvénye:



A Normál eloszlás

A normál eloszlásnak, két jellemző paramétere van:

az átlag, $\boldsymbol{\mu}$, (az eloszlás helye)

és a szórás, $\boldsymbol{\sigma}$ (az eloszlás skála paramétere).

A normál eloszlás jelölése:

N(μ, σ)

A normál eloszlás sűrűségfüggvénye a következőképpen írható fel:

$$p(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

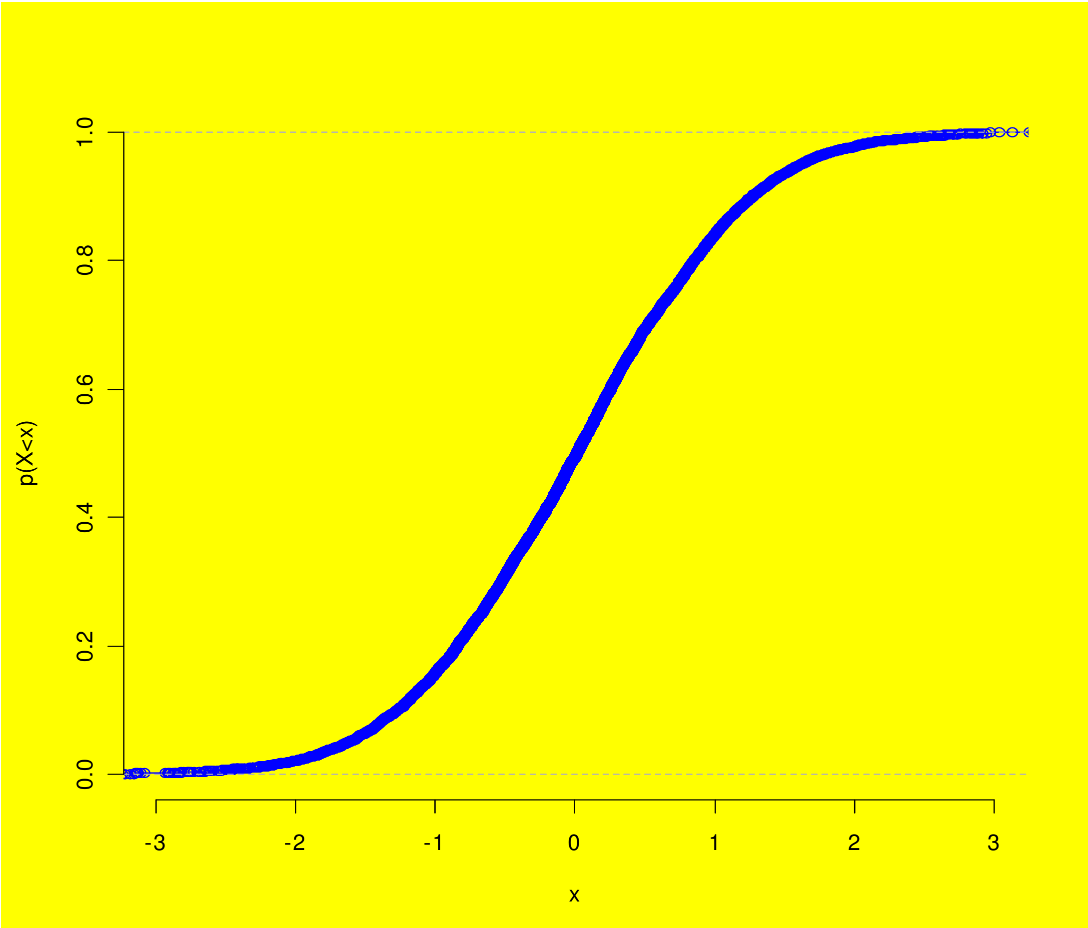
A Normál eloszlás

Normál eloszlás esetén is igaz az, hogy az egyes értékek diszjunkt eseményeknek tekinthetők, és uniójuk a biztos eseményt adja, azaz az értékek valószínűségének összege 1. Tehát a sűrűségfüggvény alatti terület 1 lesz, mert ez adja meg a biztos esemény bekövetkezésének valószínűségét.

Annak valószínűségét, hogy egy érték egy bizonyos szakaszra esik, a sűrűségfüggvény adott szakasz fölé eső része alatti területként definiálhatjuk.

Folytonos változók eloszlásai, így a normál eloszlás esetén is definiálható a kumulatív valószínűség, ami tehát egy adott érték vagy annál kisebb érték bekövetkezésének valószínűsége.

Normál eloszlás kumulatív eloszlásfüggvénye:



Normál eloszlás

A természetben nagyon sok változó normál eloszlást követ. Ennek oka, hogy számtalan változó nagyszámú faktor átlagos hatásának eredményeként alakul ki.

Ha például a testmagasságot, mint véletlen változót tekintjük, elmondható, hogy sok gén működése befolyásolja az értékét, de egyéb, például környezeti hatások, vagy táplálkozási szokások, stb. is hatással vannak rá.

Tudományos alapjául a *centrális határeloszlás elmélet* szolgál, mely szerint ha sokszor veszünk megfelelően nagy, azonos elemszámú mintát, akkor a minták átlagai mindig normál eloszlást követnek, függetlenül az eredeti eloszlástól.

Normál eloszlás

Az átlag körüli értékek valószínűsége nagyobb, mint extrémebb értékek előfordulásáé, mert átlagos, vagy ahhoz közeli értékek a befolyásoló faktorok értékeinek lényegesen többféle kombinációjának eredményeképpen kialakulhatnak, mint extrémebb értékek.

(Extrémebb értékek bekövetkezéséhez az szükséges, hogy az egyes befolyásoló faktorok értékei is extrémek legyenek, aminek kisebb a valószínűsége, mint annak, hogy a faktorok értékei vegyesen extrémek, kevésbé extrémek, vagy éppen a másik szélsőséget képviselik.)

Standardizálás, z-értékek

Bár sok változó követ normál eloszlást, a különböző változók legtöbbször nem azonos, vagy összehasonlítható egységekben vannak kifejezve. Bizonyára mindenki számára nyilvánvaló, hogy a 71 kg-os testsúly közvetlenül nem hasonlítható össze a Beck Depression Invetory-n (BDI; egy gyakran használt depresszió kérdőív) elért 16-os pontszámmal.

Tegyük fel, hogy egy diák a statisztika vizsgán 60 pontot szerez, az anatómia vizsgán pedig 70-et. A statisztika pontszámok átlaga 50, szórása 5, az anatómia pontszámok átlaga 55, szórása 10. Melyik vizsgán teljesített jobban a diák?

Ahhoz, hogy ezt a kérdést meg tudjuk válaszolni, a két eloszlást "közös nevezőre" kell hozni. Ennek módja, ha a kérdéses értékeket standardizáljuk, azaz a nyers pontértékekből kiszámoljuk a z-értéket.

Standardizálás, z-értékek

Tehát, ha adott egy normál eloszlású véletlen változó, X, (X ~ N(μ, σ)) akkor ezen X változó egy x értékéhez tartozó z érték a következőképpen számolható:

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

Ha egy normál eloszlást követő változó minden értékét standardizáljuk, akkor az így kapott z értékek normál eloszlást fognak követni 0 átlaggal, és 1 szórással, függetlenül az eredeti normál eloszlás paramétereitől.

A 0 átlagú, 1 szórású normál eloszlást ***standard normál eloszlás***nak nevezzük, és ***N(0,1)*** –el jelöljük.

Standardizálás, z-értékek

Visszatérve kiinduló példánkhoz, a következőket ismerjük:

	statisztika	anatómia
nyers pontszám	$x_s = 60$	$x_a = 70$
átlag	$\mu_s = 50$	$\mu_a = 55$
szórás	$\sigma_s = 5$	$\sigma_a = 10$

$$z_{x_s} = \frac{x_s - \mu_s}{\sigma_s} = \frac{60 - 50}{5} = 2$$

$$z_{x_a} = \frac{x_a - \mu_a}{\sigma_a} = \frac{70 - 55}{10} = 1.5$$

Tehát a példában szereplő diák statisztikából teljesített jobban.

Kumulatív valószínűségi táblázat és használata

A standard normál eloszlás (z) értékeihez tartozó kumulatív valószínűségek is megtalálhatók táblázatba foglalva, illetve kiszámíthatóak például az R statisztikai szoftver használatával.

Normál eloszlások esetén elegendő a standard normál eloszlást alkotó z értékek kumulatív valószínűségi értékeit táblázatba foglalni, mivel bármely normál eloszlás átalakítható standard normál eloszlásúra.

Az R szoftver esetén a pnorm(érték,μ,σ) parancs alkalmazásával kaphatjuk meg az 'érték'-hez tartozó kumulatív valószínűséget, az μ és σ paraméterekkel leírható Normál eloszlás esetén.