Normális eloszlás

Normális eloszlás

A statisztikai tesztek előfeltétele többnyire unimodális eloszlás (csak egy módusz).

Parametrikus tesztek előfeltétele gyakran normális eloszlás.

- Folytonos változók,
- módusz nagyjából az eloszlás közepén található, megegyezik a mediánnal és az átlaggal,
- onnan mindkét irányban szimmetrikus csökkenés,
- megközelítőleg harang formájú (Gauß-görbe).
- aszimptotikus (0-hoz közelít).

Hol fordul elő a normális eloszlás?

A legtöbb adat normális eloszlású, ha a minta elég nagy.

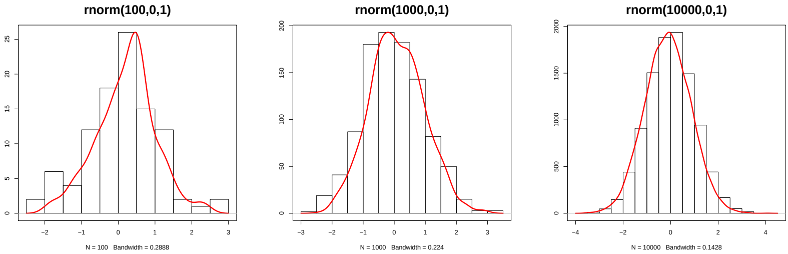
- IQ – az átlag önkényesen rögzített érték! (Mentális életkor/valós életkor*100.)
- Emberek magassága és súlya nemenként.
- Egy adott tóból kifogott halak súlya és hossza.
- Egy vizsgán szerzett osztályzatok, ha a minta elég nagy, és a tanár szigorú ☺.
- Klíma, pl. minimum és maximum hőmérsékletek adott időn át.

Hol nem találunk normális eloszlást?

- Pénzügyi mutatók és gazdasági adatok.
- Árváltozások, hozamok, tőzsdei értékek, részvények, árfolyamok.
- Emberek élettartama.
- Műszaki és elektronikus termékek élettartama.
- Várakozási idő sorban állva.
- Biztonsági adatok (pl. autóbalesetek).

Normális eloszlás (N)

Függ n-től – az elemszámtól, és k-tól – az osztályok/kategóriák számától. Itt: $n = 100, 1000, 10000$.



`a = rnorm(100,0,1)`: 100 random szám húzása normális eloszlásból, $\bar{x} = 0$ átlaggal, $s = 1$ szórással.
`hist(a)`: hisztogramm. `plot(density(a))`: sűrűségfüggvény.
Újabb ábra egy már meglévő ábrára: a két függvény között `par(new=T)` futtatása.

N paraméterei

μ: populáció feltételezett átlaga.

σ: populáció feltételezett szórása.

$\bar{x}$: minta átlaga.

s: minta szórása.

minta átlaga:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

N paraméterei

μ: populáció feltételezett átlaga.

σ: populáció feltételezett szórása.

$\bar{x}$: minta átlaga.

s: minta szórása.

minta átlaga:

$$\bar{x} = \frac{x_1+x_2+x_3+\dots+x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

minta varianciája:

$$v = s^2 = \frac{\sum_{i=1}^{n} (x_i-\bar{x})^2}{n-1}$$

N paraméterei

μ: populáció feltételezett átlaga.
σ: populáció feltételezett szórása.
$\bar{x}$: minta átlaga.
s: minta szórása.

minta átlaga:

$$\bar{x} = \frac{x_1+x_2+x_3+\cdots+x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

minta varianciája:

$$v = s^2 = \frac{\sum_{i=1}^{n} (x_i-\bar{x})^2}{n-1}$$

minta szórása:

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i-\bar{x})^2}{n-1}}$$

z-transzformáció

Szórás függ az elemszámtól és az átlagtól $\Rightarrow$ eltérő átlagok eloszlása nem összehasonlítható.
Megoldás: standardizálás z-értékre. Minden egyes elemre:

$$z_i = \frac{x_i - \mu}{\sigma}$$

azaz minden egyes elem mérőszámából kivonjuk a populáció (itt: minta) átlagát, és elosztjuk a **populáció** feltételezett szórásával. (Ekkor a nevezőben n van, nem $n-1$, mint a **minta** szórásánál.)

Ez az eljárás a z-transzformáció.

Standard normális eloszlás

Normális eloszlás jellemzői: $N(\mu, \sigma)$.

átlag: $x_i = \mu = \bar{x}$
z-transzformáció:

$$z_i = \frac{\mu - \mu}{\sigma} = 0$$

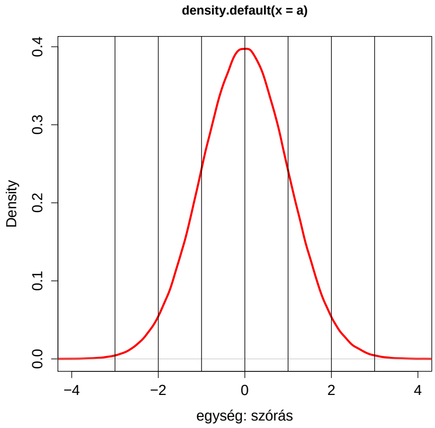
szórás: $\sigma = \mu + \sigma$

$$z_i = \frac{(\mu + \sigma) - \mu}{\sigma} = 1$$

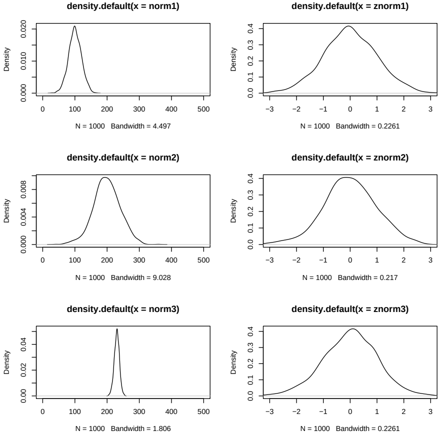
standard normális eloszlás jellemzése: $N(0, 1)$

Standard normális eloszlás

Sűrűségfüggvény:
x-tengely: egységnyi szórás, tartomány: $\sigma = -\infty \cdots + \infty$.



Hozzunk létre normális eloszlású vektorokat eltérő átlagokkal és szórásokkal, majd standardizáljuk őket z-transzformácóval. Szükséges parancssor: 5_statisztika_R.txt



Függvény jellemzői

- az x-tengely és a sűrűségfüggvény által bezárt terület összege $= 1$.
- Az esetek 50%-a az átlagtól balra helyezkedik el.
- $\sigma = -1 \cdots + 1$ közötti tartomány az esetek 68,27%-át tartalmazza.
- $\sigma = -2 \cdots + 2$ közötti tartomány az esetek 95,45%-át tartalmazza.
- $\sigma = -3 \cdots + 3$ közötti tartomány az esetek 99,73%-át tartalmazza.

Standard normális eloszlás