

A populáció eloszlása (elméleti eloszlások)



Szűcs Mónika, Griechisch Erika, Rárosi Ferenc
SZTE ÁOK-TTIK Orvosi Fizikai és Orvosi Informatikai Intézet

Utoljára frissítve: 2018. augusztus 31.



1. A leckében előforduló fogalmak

Megfigyelési, mintavételi egység: A vizsgálat alanya vagy tárgya. A mintavételi egységeken végezzük el a vizsgálatokat, méréseket, megfigyeléseket. Orvosi gyakorlatban a leggyakoribb mintavételi egység egy ember, egy szerv, egy vérminta, más vizsgálatok esetén gyakori példa egy család, egy település, egy erdő, egy állat, egy faj, egy iskolai osztály.

Populáció: Az összes lehetséges mintavételi egységet tartalmazó halmaz, amiről szeretnénk információt nyerni a vizsgálat, kísérlet során. A populáció nagyon ritkán egy jól meghatározott véges halmaz, mint például egy ritka trópusi betegségben szenvedő 200 beteg (rajtuk kívül senki mást nem fertőzött meg ez a betegség, és már nem is fog). Gyakoribb azonban hogy a populáció nagysága végtelen, nem lehet az összes múlt, jelen és jövőbeli egyedet beazonosítani. *Biostatistikában, és ebben a leckében is populáció alatt nem magukat a mintavételi egységek sokaságát, hanem az azokon mért értékeket értjük.*

Véletlen változó: egy véletlen jelenség (kísérlet) számmal jellemezhető kimenetele. Például, ha két kockával dobunk, a dobott számok szorzata egy véletlen változó. Másik példa: ha a csoportból az terembe elsőnek belépő hallgató vérnyomását megmérjük, a mért vérnyomás érték egy véletlen változó.

Ebben a leckében a véletlen változók **két nagy csoportjával** foglalkozunk: diszkrét és folytonos.

2. Egy diszkrét véletlen változó eloszlása

Diszkrét véletlen változó véges sok értéket vehet fel.

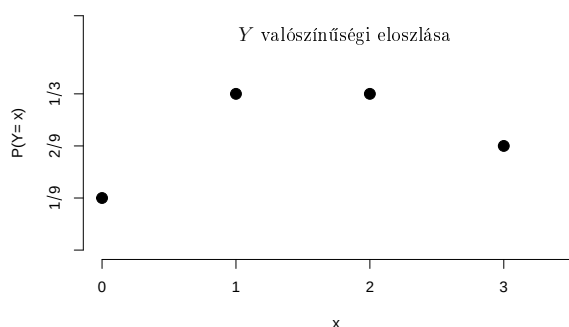
Egy diszkrét véletlen változó valószínűségi eloszlása az egyes kimenetek valószínűségei: p_i .

Tulajdonságok:

- p_i nem negatív bármely i esetén, hiszen valószínűség
- az összes lehetséges érték valószínűségének összege 1. $\sum_i p_i = 1$

Példa: Definíáljuk az Y véletlen változót a következőképpen:

$$Y = \begin{cases} 0 & , \text{ha a vércsoport } 0 \\ 1 & , \text{ha a vércsoport } A \\ 2 & , \text{ha a vércsoport } B \\ 3 & , \text{ha a vércsoport } AB \end{cases}$$



$$P(Y = x) = \begin{cases} 1/9 & , \text{ha } x = 0 \\ 2/9 & , \text{ha } x = 1 \\ 1/3 & , \text{ha } x = 2 \\ 2/9 & , \text{ha } x = 3 \end{cases}$$

SZÉCHENYI 2020

3. Folytonos véletlen változó eloszlása

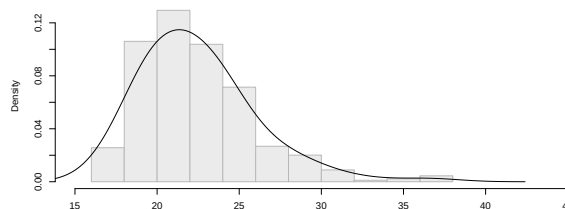
- **Folytonos véletlen változó** végtelen sok értéket vehet fel egy adott intervallumon belül.

- **Példa:** Definiáljuk az Y véletlen változó értékeit a betegek BMI értékeivel.

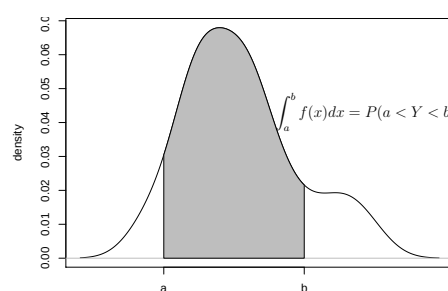
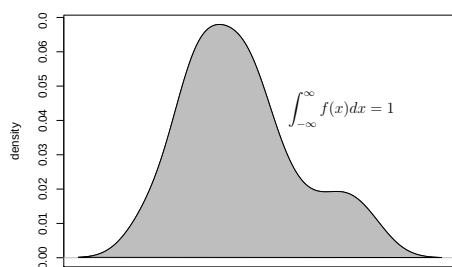
- Folytonos véletlen változó bármely adott értéket 0 valószínűséggel vesz fel.

- Egy folytonos változó eloszlását az $f(x)$ **sűrűségfüggvény** írja le, amely az alábbi tulajdonságokkal rendelkezik:

Hisztogram és sűrűségfüggvény a BMI alapján



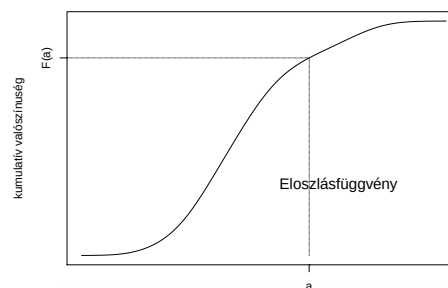
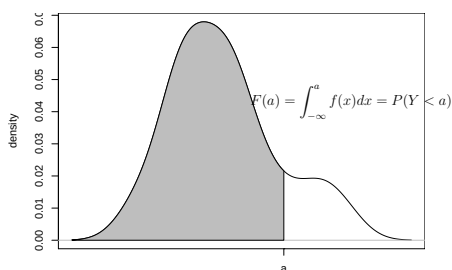
- A sűrűségfüggvény értékei mindig nem negatívak. $f(x) \geq 0$
- Az $f(x)$ alatti teljes terület 1. $\int_{-\infty}^{+\infty} f(x) dx = 1$
- Egy adott intervallumba esés valószínűsége a sűrűségfüggvény görbéje alatti terület az adott intervallumon. $\int_a^b f(x) dx = P(a \leq Y < b)$



- **Eloszlásfüggvény:** az a függvény, amely megadja, hogy a véletlen változó értékei mekkora valószínűséggel kisebbek a -nél. Az Y eloszlásfüggvényét $F_Y(a) = P(Y < a)$ jelöljük.

Egy adott a érték esetén az eloszlásfüggvény megadja az a -tól balra eső sűrűségfüggvény görbéje alatti területet.

$$F_Y(a) = \int_{-\infty}^a f(x) dx = P(Y < a)$$



SZÉCHENYI 2020



MAGYARORSZÁG
KORMÁNYA

Európai Unió
Európai Szociális
Alap

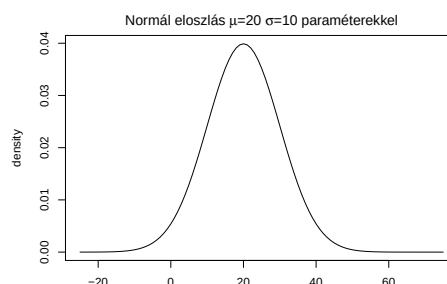


BEFEKTETÉS A JÖVŐBE

4. Normális eloszlás

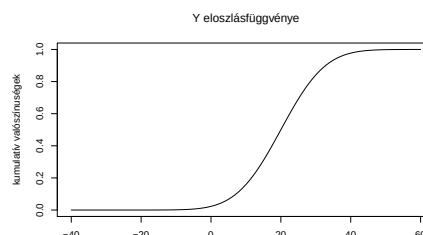
- A normális eloszlás az egyik leggyakrabban használt folytonos eloszlás.
- Azt mondjuk, hogy Y normális eloszlást követ μ populációátlaggal és σ^2 variánciával, ha a sűrűségfüggvénye a haranggörbe.
Jelölés: $Y \sim N(\mu, \sigma^2)$

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



- A normális eloszlás sűrűségfüggvényének integrálja nem írható fel zárt formában. A valószínűségek számolásához a standard normális eloszlás közelítő értékeinek táblázatát használjuk.

$$F(a) = \int_{-\infty}^a \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

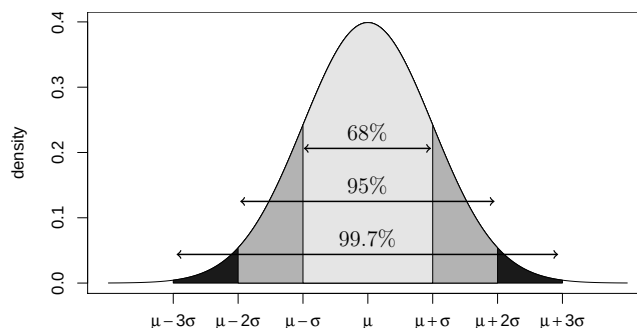


• Példa:

- Sok élettani jellemző normális eloszlást követ: magasság, lábméret, vérnyomás...
- A gyártási folyamatok mérési hibái megközelítően normális eloszlást követnek.

4.1. A 68-95-99,7% -os szabály

- Az értékek 68%-a a populációátlag 1 szórásnyi környezetébe esnek:
 $P(\mu - \sigma \leq Y < \mu + \sigma) = 0,68$
- Az értékek 95%-a a populációátlag 2 szórásnyi környezetébe esnek:
 $P(\mu - 2\sigma \leq Y < \mu + 2\sigma) = 0,95$
- Az értékek 99,7%-a a populációátlag 3 szórásnyi környezetébe esnek:
 $P(\mu - 3\sigma \leq Y < \mu + 3\sigma) = 0,997$

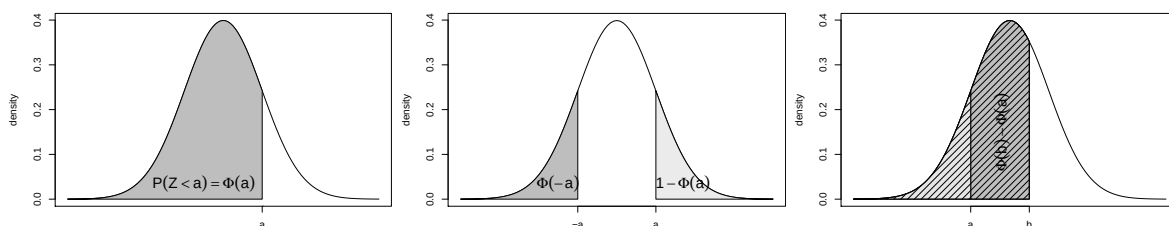


4.2. Standard normális eloszlás

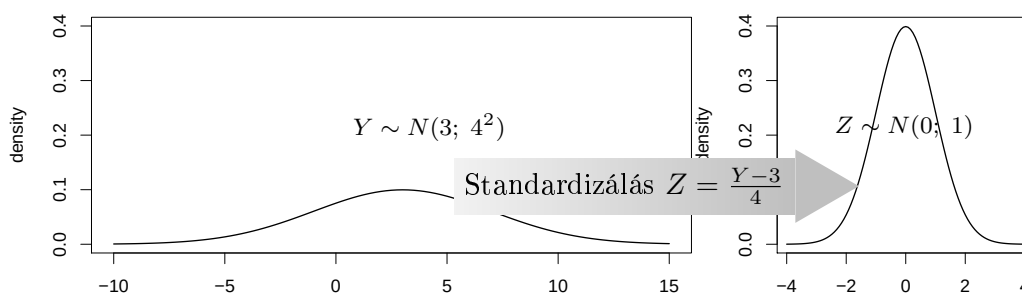
- A 0 várható értékű, 1 szórású normális eloszlást, $N(0; 1)$ standard normális eloszlásnak nevezzük.
- A standard normális eloszlású véletlen változókat általában Z -vel jelöljük.
- Z sűrűségfüggvényét $\varphi(x)$ -vel jelöljük.
- Z eloszlásfüggvénye: $\Phi(x) = P(Z < x)$

- SZABÁLYOK (standard normális eloszlás)

- $P(Z < a) = \Phi(a)$ Az a -tól balra eső sűrűségfüggvény görbéje alatti terület.
- $P(Z \geq a) = 1 - \Phi(a)$ Az a -tól jobbra eső sűrűségfüggvény görbéje alatti terület.
- $P(Z \geq a) = \Phi(-a)$, ugyanis a haranggörbe szimmetrikus.
- $P(a \leq Z < b) = P(Z < b) - P(Z < a) = \Phi(b) - \Phi(a)$



4.3. Standardizálás Legyen $Y \sim N(\mu; \sigma^2)$, akkor $Z = \frac{Y - \mu}{\sigma} \sim N(0; 1)$.



5. A centrális határeloszlás tétele

- Egy μ átlagú és σ szórású populációból vett nagy elemszámú (n) minták átlagai olyan populációból származnak, melyek eloszlása közelítően normális eloszlású μ átlaggal és $\frac{\sigma}{\sqrt{n}}$ szórással.
- $\frac{\sigma}{\sqrt{n}}$ az **átlag szórása** (standard error of mean, SEM). A populációból vett n elemű minták átlagának szóródása az (ismeretlen) populációátlag körül.
- A $(\bar{x} - 2 \cdot \text{sd}, \bar{x} + 2 \cdot \text{sd})$ intervallum tartalmazza az adatok középső 95%-át.
- A $(\bar{x} - 2 \cdot \text{se}, \bar{x} + 2 \cdot \text{se})$ intervallum 95%-os valószínűséggel tartalmazza a populációátlagot.

6. Student féle t -eloszlás

- A t -eloszlás (Student féle t -eloszlás) az a valószínűségi eloszlás, melyet akkor használunk, ha a populáció bizonyos paraméterét úgy akarjuk meghatározni, hogy a populáció szórása ismeretlen.
- Végtelen sok t -eloszlás van, minden szabadságfokhoz (a minta elemszámától függő számhoz) különböző t -eloszlás tartozik.
- Ha veszünk egy n elemű mintát egy μ átlagú és σ szórású populációból, akkor

$$t = \frac{\bar{x} - \mu}{\frac{\text{sd}}{\sqrt{n}}}$$

t -eloszlást követ $n - 1$ szabadságfokkal.

SZÉCHENYI 2020



MAGYARORSZÁG
KORMÁNYA

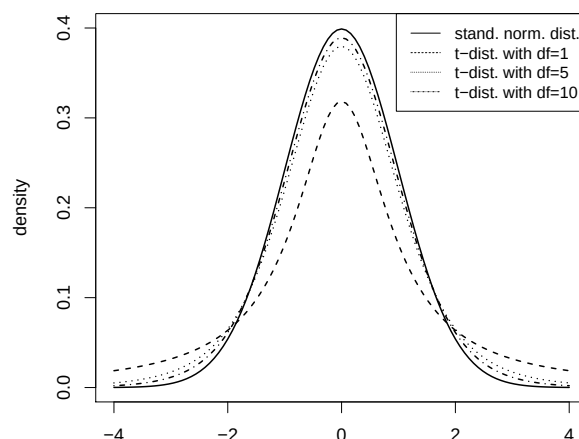
Európai Unió
Európai Szociális
Alap



BEFEKTETÉS A JÖVŐBE

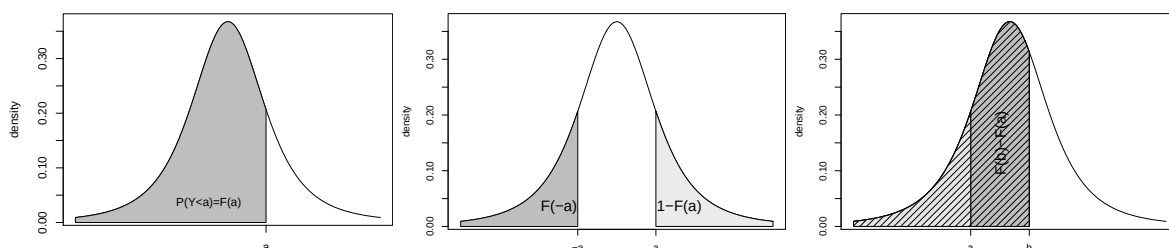
- A t -eloszlás sűrűségfüggvényének tulajdonságai:

- „Harang-alakú” és szimmetrikus görbe, a normális eloszlás sűrűségfüggvényéhez hasonló alakú.
- Szórása nagyobb, mint a standard normális eloszlásé és függ a szabadságfoktól.
- Ahogy növeljük a szabadságfokot, a sűrűségfüggvény alakja egyre jobban közelíti a standard normális eloszlás haranggörbéjét.



- **SZABÁLYOK (t -eloszlás)** Legyen Y egy Student t eloszlású véletlen változó.

- $P(Y < a) = F(a)$ A sűrűségfüggvény a -tól balra eső görbéje alatti terület.
- $P(Y \geq a) = 1 - F(a)$ A sűrűségfüggvény a -tól jobbra eső görbéje alatti terület.
- $P(Y \geq a) = F(-a)$, ugyanis a t eloszlás sűrűségfüggvénye szimmetrikus.
- $P(a \leq Y < b) = P(Z < b) - P(Z < a) = F(b) - F(a)$



Felhasznált irodalom

- Reiczigel Jenő, Harnos Andrea, Solymosi Norbert: Biostatisztika nem statisztikusoknak, Pars Kft. (2014)
- Reiczigel Jenő: Válogatott fejezetek a biostatistikából, SZIE ÁOTK, (2005), <https://docplayer.hu/279567-Reiczigel-jeno-valogatott-fejezetek-a-biostatistikabol.html> (hozzáférés: 2018.08.31.)