# Mi az adat?

- **Az adat elemi ismeret.**
- Tények, fogalmak olyan megjelenési formája, amely alkalmas emberi eszközökkel történő értelmezésre, feldolgozásra, továbbításra.
- **Az adatokból** gondolkodás vagy gépi feldolgozás útján **információkat**, azaz **új ismereteket nyerünk.**
- **A számítástechnikai eszközökkel rögzített, azokkal feldolgozható és megjeleníthető információt adatnak nevezzük.**
- **Az információ tehát értelmezett adat.**

# Adat osztályozása

- Az adatok ***jellegűk szerint*** lehetnek:

- minőségi / megállapítható / kvalitatív,
- mennyiségi / mérhető / kvantitatív adatok.

- Az adatok ***értékük / értékkészletük*** szerint lehetnek:
    - bináris,
    - diszkrét,
    - folytonos adatok.

# Az adat típusa (skálája)

- **<u>Nominális adat</u>**: amelynek lehetséges értékei között csak az azonos vagy nem azonos reláció van értelmezve (=, ≠).
- **<u>Ordinális adat</u>**: amelynek lehetséges értékei között a kisebb v. nagyobb (<,>) reláció is megengedett az előőbi relációk mellett.
- **<u>Intervallum adat</u>**: amelynél fentieken kívül az +, -, * műveleteket is tudjuk értelmezni.
- **<u>Arány skála adat:</u>** mind a négy alapművelet értelmezve van (/).

# Kategórikus adatok

| Adatfajta | Az adatokon értelmezett relációk | Példa |
|---|---|---|
| Nominális (nevesítő) | =, ≠ | Nem, faj, betegségfajta |
| Ordinális (rendező) | =, ≠, <, > | Betegség súlyossága, dohányzás súlyossága |

Azokat a kategorikus adatokat, amelyek csak két osztály valamelyikébe sorolhatók, *dichotómikus* vagy *bináris* adatoknak nevezzük.

# Metrikus (folytonos) adatok

| Adatfajta | Az adatokon értelmezett relációk | Példa | Fontos |
|---|---|---|---|
| Intervallum | =, ≠, <, >, +, -, * | Potenciál, Celsius fokban mért hõmérséklet | Nincs fix 0 pont, az önkényes. |
| Arány | =, ≠, <, >, +, -, *, ÷ | Tömeg, Abszolut hõmérséklet, vérnyomás | Létezik fix 0 pont. |

A *kvantitatív adatok lehetnek folytonos vagy diszkrét (mérhető vagy leszámlálható,* gyakran *metrikusnak nevezettek) adatok.*

# Sokaság (n, populáció)

- Mindazon elemek halmaza (a vizsgálat tárgya), amelyre statisztikai következtetés irányul.
- A sokaság nem emberek (egyedek), hanem egy őket jellemző ismérv (változó) által felvett vagy felvehető értékek halmaza.
- Két típusa van: időpontban **(álló sokaság)** vagy időintervallumban **(mozgó sokaság)** vizsgáljuk-e.

# Minta (N)

- Egy adott véges számú sokaságból kiválasztott – véges számú – egységek összessége.
- A minta elemszáma mindig kisebb, mint maga az alapsokaság.
- Ha e kettő megegyezik, akkor cenzusról beszélünk (például népszámlálás).
- <u>Mintaelem: </u>a minta egyes elemei.

# Véletlen minta

- A minták egy speciális esete.
- A véletlen kiválasztás esetében az alapsokaság minden egységéről megmondható, hogy milyen valószínűséggel kerülhet be a mintába.
- A társadalomkutatás módszertana, azaz a matematikai statisztika a véletlen mintavételre támaszkodik.

# Reprezentativitás

- Egy minta bizonyos változók mentén akkor reprezentatív, ha a mintába került elemek (emberek) ugyan olyan arányban vannak jelen, mint az alapsokaságban.
- Tehát egy minta csak bizonyos szempontok alapján nevezhető reprezentatívnak.
- Ha más szempontokat veszünk figyelembe, akkor mintánk lehet, hogy nem reprezentatív.

# Mintavételi hiba

- A mintavétel hibája abból adódik, hogy nem a teljes populációt kérdezzzük meg, hanem annak csak egy részét.
- Így információink részlegesek lesznek a teljes alapsokaságról.
- A mintavételi hiba mértéke szoros összefüggésben van a minta elemszámával, valamint a mintavételi módszerrel.
- A mintavételi hiba mértéke akkor számolható ki érvényesen, ha véletlen mintát vettünk.

# Nem mintavételi hiba

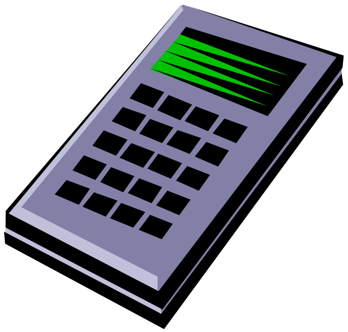
- A nem mintavételi hibák az adatfelvételhez kapcsolódnak, nincsenek kapcsolatban sem a mintavétel módszerével, sem a minta elemszámával. Nem mintavételi hibák összetevői az alábbiak lehetnek:

  - a kérdőív hibás megszerkesztése;
  - kérdezőbiztosok munkájának hibái;
  - hibás rögzítés;
  - a mintába bekerült válaszadó félreérthető vagy valós véleményét elfedő válasza, stb.

# Mit kell tenni?

## A klinikai tervezés előtt:

- Mintaszám meghatározás
- Erő (Power) analízis



# Valószínűség

- Azt a számot, amely körül egy esemény relatív gyakorisága ingadozik, az illető esemény valószínűségének nevezzük.
- Egy esemény bekövetkezésének vagy be nem következésének mértékbeli megadása.
- A valószínűség, mint mérték 0 és 1 közötti szám.
- A jelölés a latin probabilitas, valószínűség szó kezdőbetűjéből ered.

- **Eseménynek** nevezünk mindent, amiről a kísérlet elvégzése után eldönthető, hogy bekövetkezett-e, vagy sem (**elemi esemény).**
- Valamely kísérlet összes kimenetele egy halmazt alkot. Ezt nevezzük **eseménytérnek** ($\Omega$).
- **Biztos eseményről** akkor beszélünk, ha a kísérlet során biztosan (minden kimenetelnél) bekövetkezik.
- Azt az eseményt, mely akkor és csak akkor következik be, ha az *A* esemény nem következik be, az *A* esemény **ellentett eseményének** nevezzük ($\overline{\mathrm{A}}$) .

# Alapfogalmak

***1. Valószínűség:***

Eseményeken értelmezett számértékű függvény (mérték).
Jelölésben $P(A)=p$

**<u>Kolmogorov axiómák:</u>**

- $0 \leq P(A) \leq 1$
- $P(0)=0$ és $P(I)=1$
- Ha $AB = 0 \Rightarrow P(A+B) = P(A) + P(B)$

***2. Valószínűségszámítás és a statisztika kapcsolata***
*(klasszikus modell)*

$$p = \frac{k}{n} = \frac{\text{kedvező esetek száma}}{\text{összes eset száma}}$$

***3. Statisztikai próba (teszt):***
A mért adatokon értelmezett függvény.

***4. Szignifikancia értelmezése :*** $p < 0.05$

## Feltételes valószínűség

$$P(A|B) = \frac{P(AB)}{P(B)}$$

## Teljes valószínűség tétele

Ha $B_1, B_2, B_3, \ldots\ldots, B_n$ események teljes eseményrendszert alkotnak és $P(B_I) \neq 0$, akkor az A esemény valószínűsége

$$P(A) = \sum_{i=1}^{N} P(A|B_i) \cdot P(B_i)$$

## Bayes elmélet

Ha a $B_1, B_2, B_3, \ldots\ldots, B_n$ események teljes eseményrendszert alkotnak és $P(Bi) \neq 0$, akkor tetszőleges A esemény mellett ( $P(A) \neq 0$), igaz

$$P(B_i|A) = \frac{P(A|B_i) \cdot P(B_i)}{\sum_{k=1}^{N} P(A|B_k) \cdot P(B_k)}$$

**Mennyi a valószínűsége, hogy érmével fejet dobunk?**

Egy dobásnak 2 kimenetel van, tehát n = 2.
Kedvező számunkra, hogy fejre esik az érme, tehát
k = 1
P(fejet dobunk) = ½ = 0,5 = 50%

**Mennyi a valószínűsége, kockával páros számot dobunk?**

Egy dobásnak 6 kimenetele van, n = 6.
Kedvező számunkra 3 eset, így k = 3
P(páros számot dobunk kockával) =3/6 = ½=0.5=50%

**Mennyi a valószínűsége, hogy kockával nem 6-ost dobunk?**

n = 6
k = 5
P(nem 6-ost dobunk) = 5/6 = 0,833

**Egy dobozban 5 piros golyó van. Hány fehér golyót tegyünk hozzá, hogy a fehér golyó húzásának valószínűsége 80% legyen?**

Jelöljük x-szel a fehér golyók számát!

n = 5 + x

k = x

P = 0,8

$$\frac{x}{5+x} = 0{,}8$$
$$x = 0{,}8(5+x)$$
$$x = 4 + 0{,}8x$$
$$0{,}2x = 4$$
$$x = 20$$

Tehát 20 fehér golyót kell a dobozba tenni. Ekkor az össze golyók száma 25 (ez az n), ebből 20 fehér (ez a k), így a fehér golyó húzásának valószínűsége

$$\frac{20}{25} = \frac{4}{5} = \frac{8}{10} = 0{,}8$$

## Mennyi a valószínűsége, hogy ötösöd lesz a lottón?

n = $\dfrac{90 \cdot 89 \cdot 88 \cdot 87 \cdot 86}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}$ = 43.949.268

k = 1

P = házi feladat ☺

# Változó

- Több értéket képes fölvenni, így nem kell egyenként behelyettesítenünk őket például egy függvénybe, hanem elegendő csak a változó.
- A változónk értékét külön tárolhatjuk, vagy más módon is megadhatjuk – például intervallummal.
- Természetesen nem csak olyan változók léteznek, melyek számokat vehetnek fel.

# Valószínűségi változó

- A valószínűségi változó a változók speciális esete.
- Ebben az esetben meg lehet mondani, hogy a változó milyen valószínűséggel veheti fel értékeit, tehát minden egyes általa fölvehető értékhez hozzá tudunk rendelni egy valószínűséget.

- A statisztikában gyakran előfordul még a ***<u>függő és független változók</u>*** megkülönböztetése. A gyakorlatban ez azt jelenti, hogy egyik tulajdonság függvényében miként változik egy másik tulajdonság, ami értelemszerűen többváltozós esetekre is értelmezhető.

- **Gyakoriság: e**gy vagy több változó által felvehető értékre, értékekre jutó megfigyelések száma.
- **Relatív gyakoriság:** ha egy változó által felvehető értékekre jutó megfigyelések számát elosztjuk a teljes mintanagysággal, akkor a relatív gyakorisághoz jutunk.

# Eloszlás

- A statisztika központi fogalma: valami vagy valamik hol vannak, hogyan oszlanak el, hogyan helyezkednek el.
- A sokaság eloszlását – a változó típusától függően – jellemezhetjük:

  - függvénnyel (valószínűségsűrűség-függvény, kumulatív eloszlásfüggvény),

  - vagy paraméteresen (elméleti szórás, várható érték stb.).

# Valószínűségeloszlás

- Egy teljes eseményrendszer valószínűségeinek sorozatát valószínűségeloszlásnak, vagy röviden eloszlásnak nevezzük.
- Olyan függvény, mely leírja, hogy egy valószínűségi változó milyen valószínűséggel vehet fel egy bizonyos értéket.
- Eloszlásfüggvénye minden valószínűségi változónak létezik.

# Diszkrét eloszlások

Csak diszkrét, definiált értékeket vehet fel.

**Néhány véges tartományú diszkrét eloszlás:**

Binomiális eloszlás

Poisson-féle binomiális eloszlás

Bernoulli-eloszlás

Diszkrét egyenletes eloszlás

Hipergeometrikus eloszlás

# Folytonos eloszlások

A folytonos eloszlású valószínűségi változó végtelen sok értéket vehet fel.

**1. Néhány zárt intervallumú folytonos eloszlás:**

- Arkuszszinusz-eloszlás
- Logit normális eloszlás

**2. Félig végtelen intervallumú folytonos eloszlás [0,∞):**

- Khí-eloszlás
- Khí-négyzet eloszlás
- Exponenciális eloszlás
- F-eloszlás

**3. A teljes valós tartományra érvényes eloszlások [-∞, ∞)**

- Normális eloszlás
- T-eloszlás

# Valószínűségi változók

**Diszkrét**

**Folytonos**

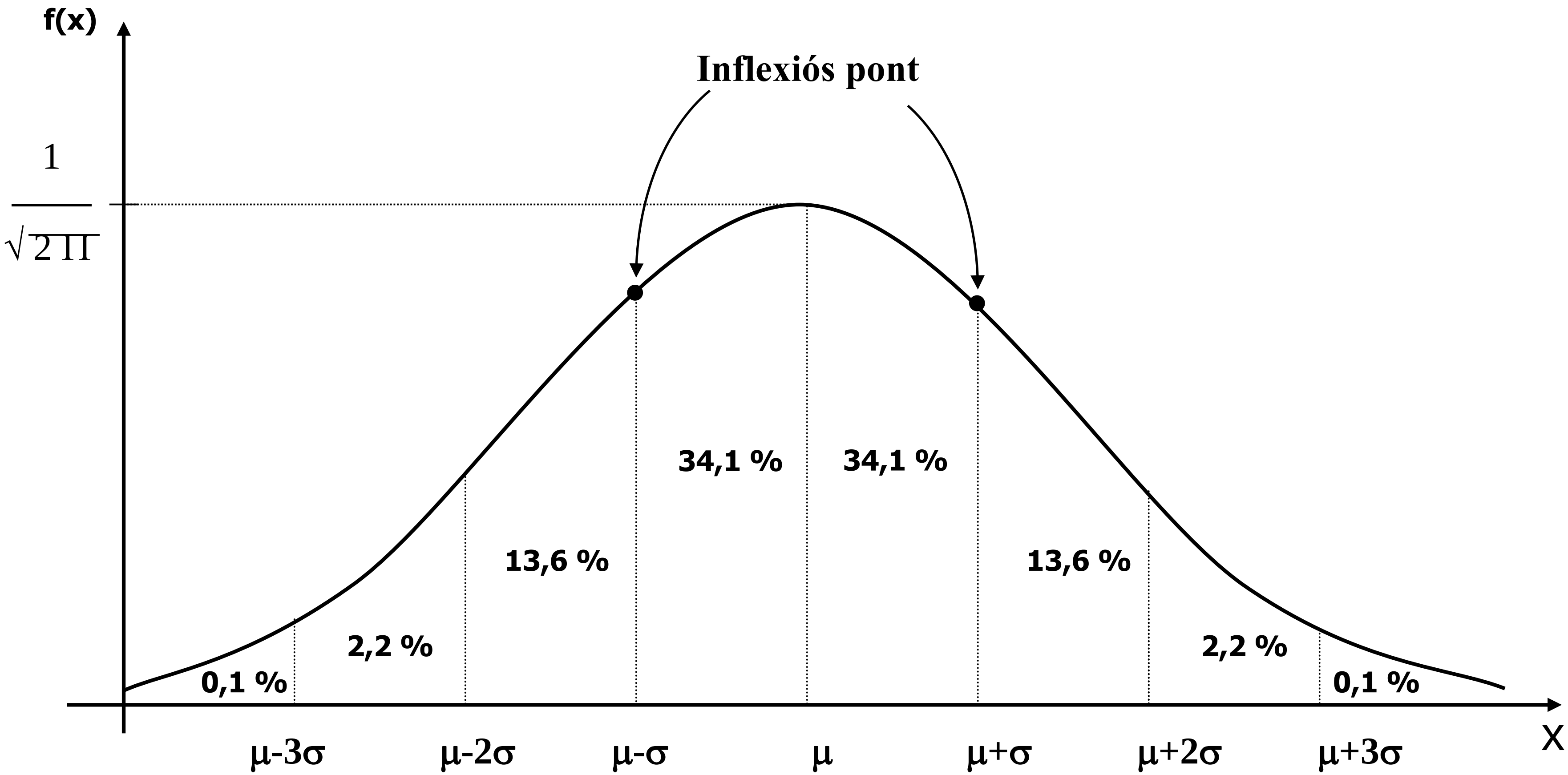
$$p_k = P(A_k) = P(\xi < x)$$

# Eloszlás függvény

$$F(x) = P(\xi < x)$$

$$F(x)' = f(x)$$

# Legfontosabb folytonos eloszlások

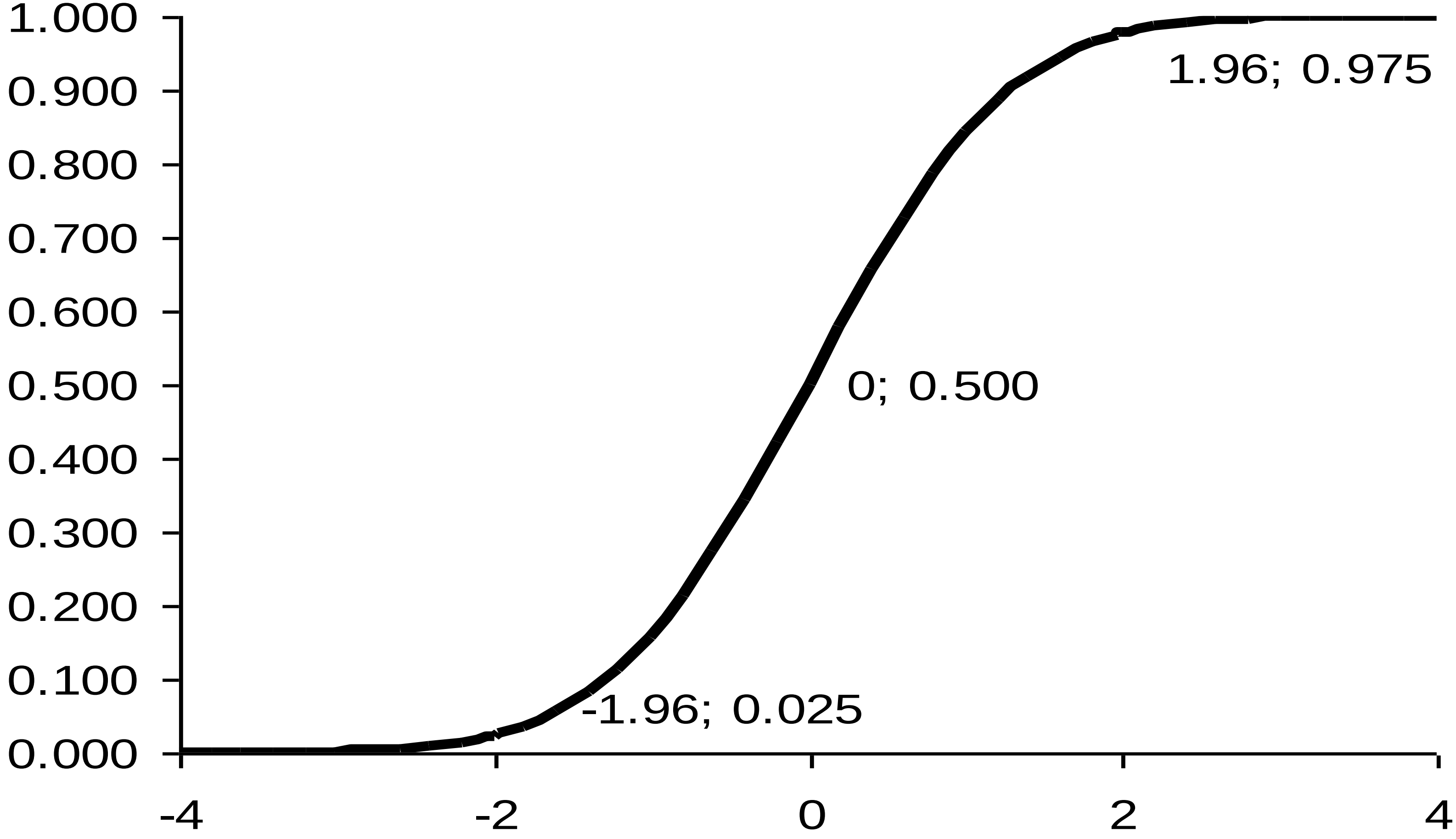



# Normális eloszlás tulajdonságai

# Sűrűségfüggvény

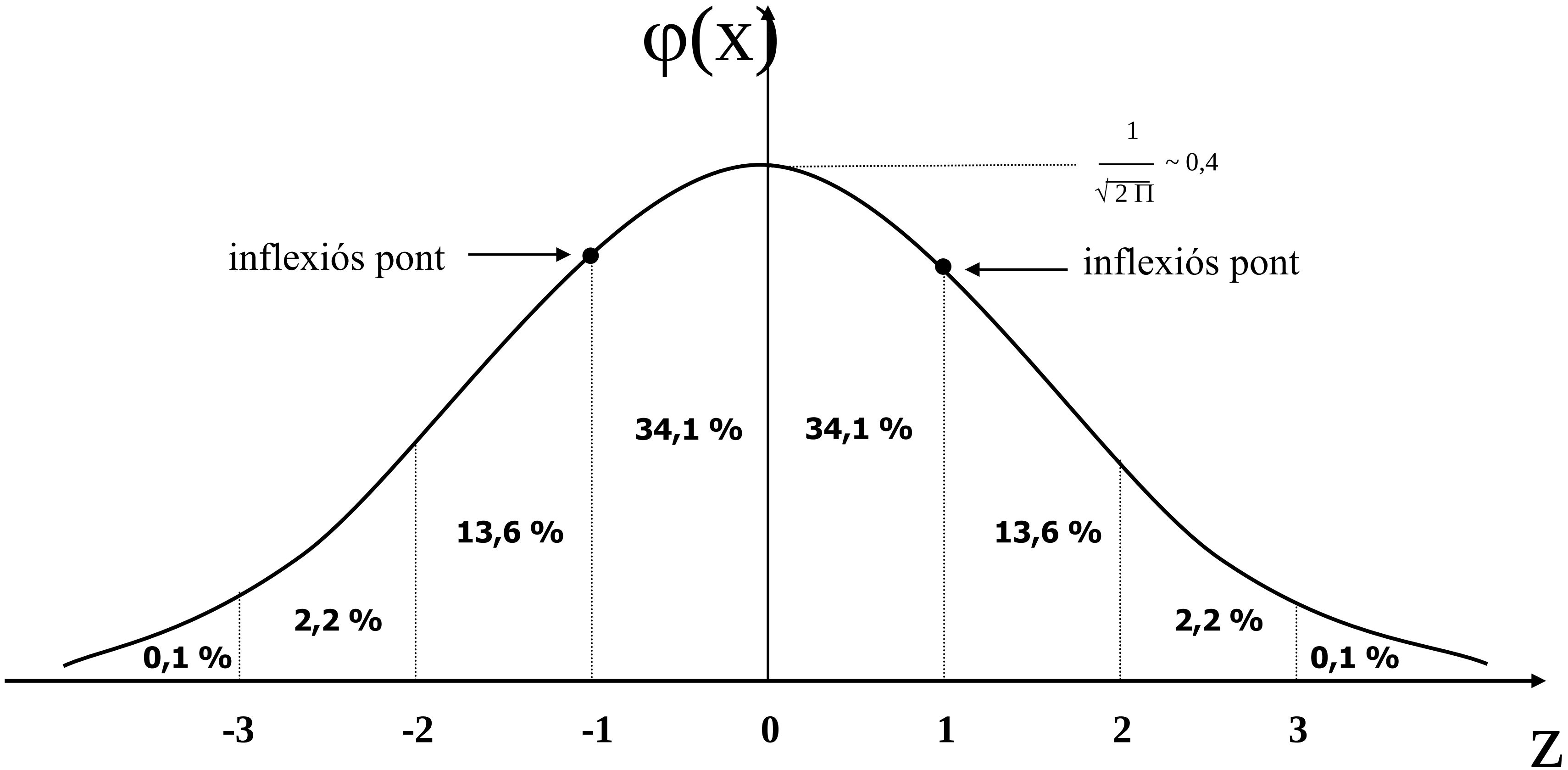
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

# Normáleloszlás eloszlásfüggvénye



# Eloszlásfüggvény

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



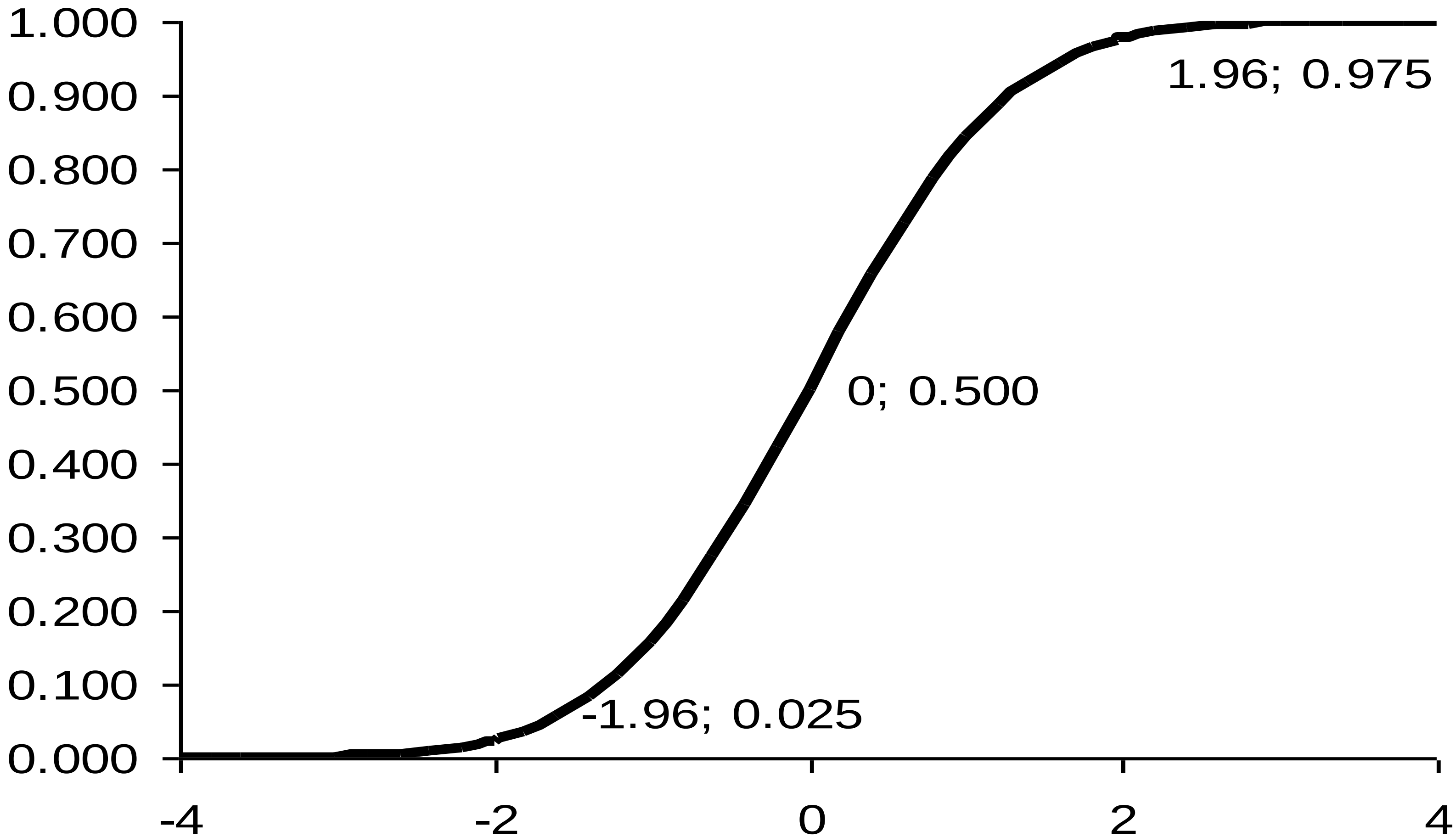

**Standard normális eloszlás**

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

# Standard normális eloszlás sűrűségfüggvénye

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

# Standard normáleloszlás eloszlásfüggvénye



# Standard normális eloszlás eloszlásfüggvénye

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{x^2}{2}} dx$$

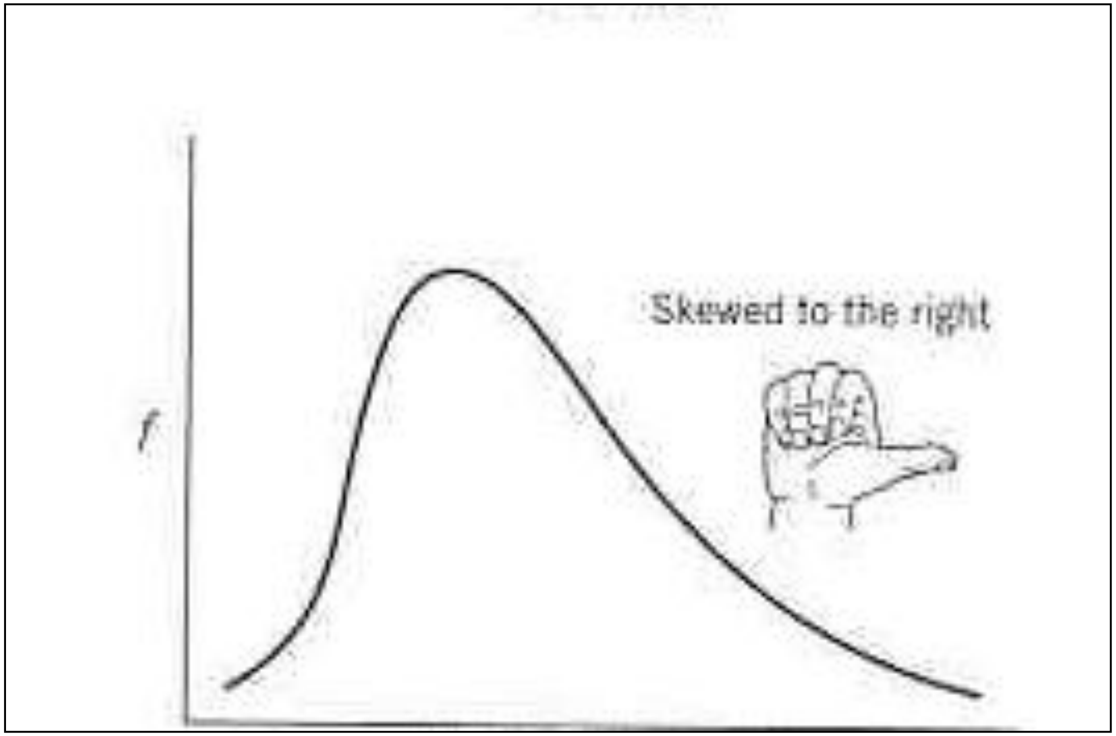
# A normál eloszlás nevezetes értékei

| $\alpha$% | $\mu \pm \sigma$ |
|---|---|
| 5 | 1,96 |
| 1 | 2,58 |
| 0,1 | 3,29 |

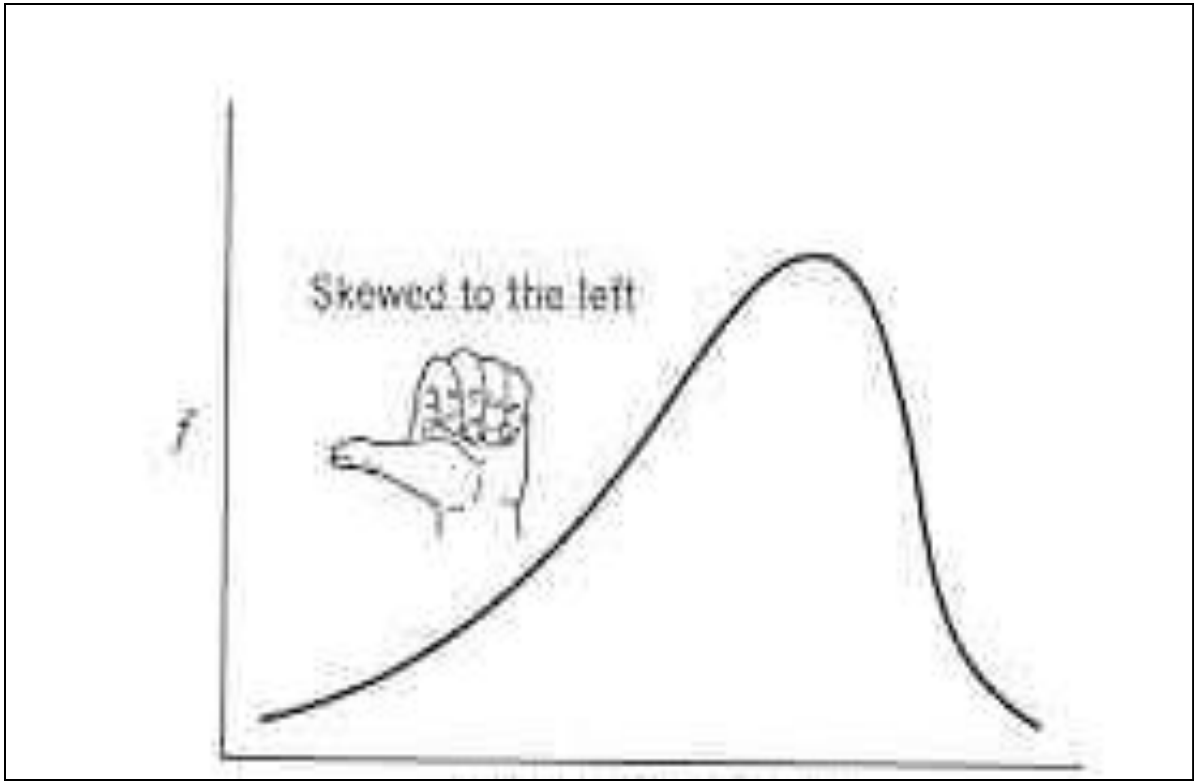
# Az eloszlás alakjának jellemzése

- Ferdeség (skewness, normális eloszlás=0 körüli érték)
- Csúcsosság (kurtosis, normális eloszlás=0 körüli érték)

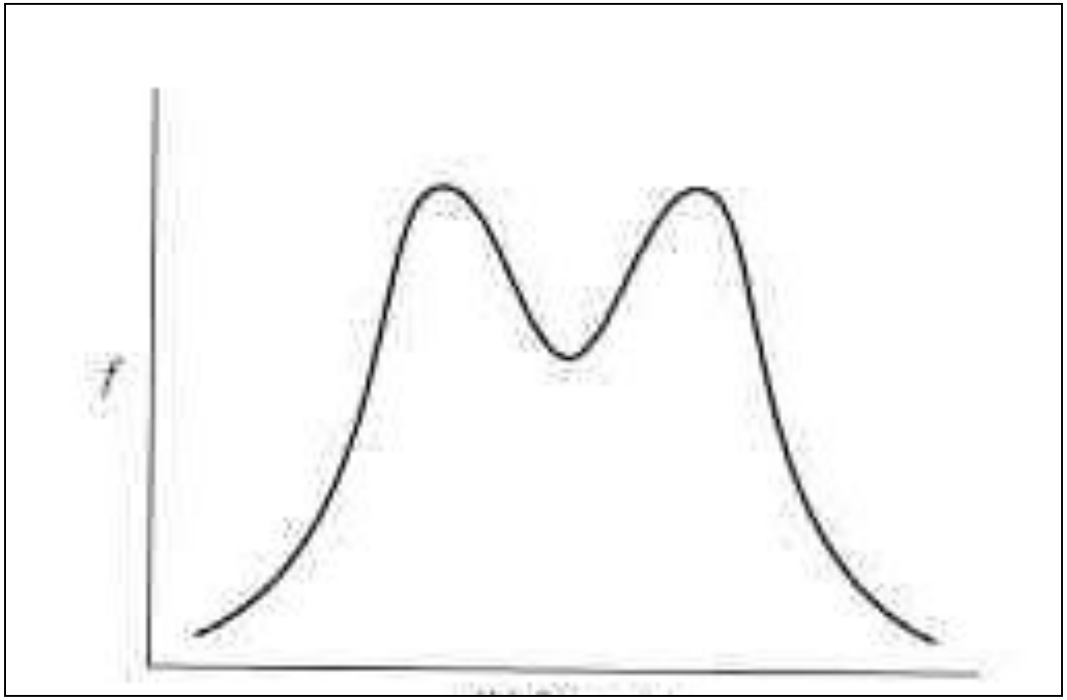
# POSITIVELY SKEWED



# NEGATIVELY SKEWED



# BI-MODAL



# További folytonos eloszlások

- t-eloszlás
- Exponenciális eloszlás
- Egyenletes eloszlás
- F-eloszlás
- Gamma