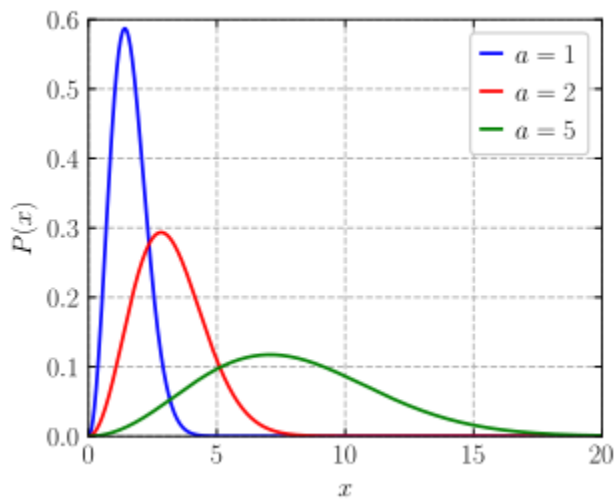
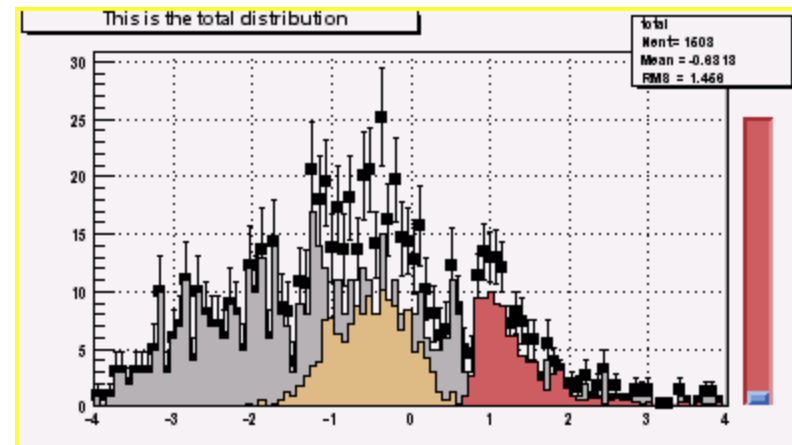
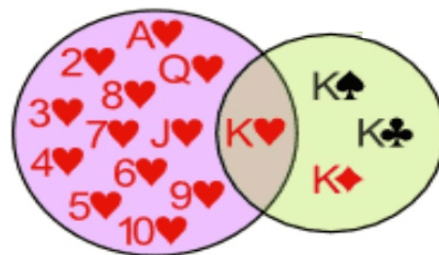
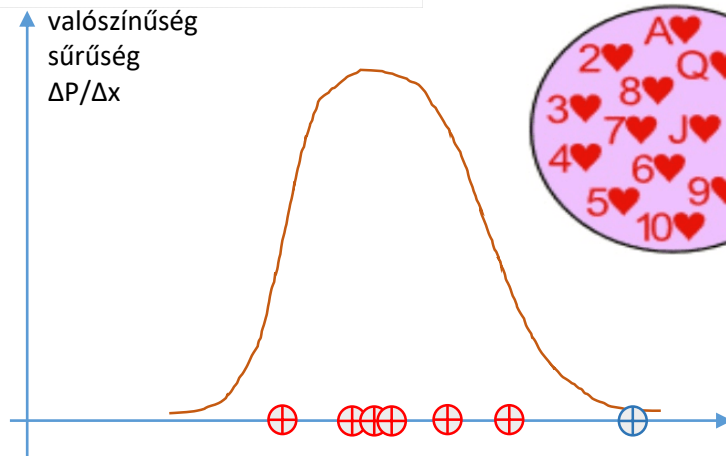
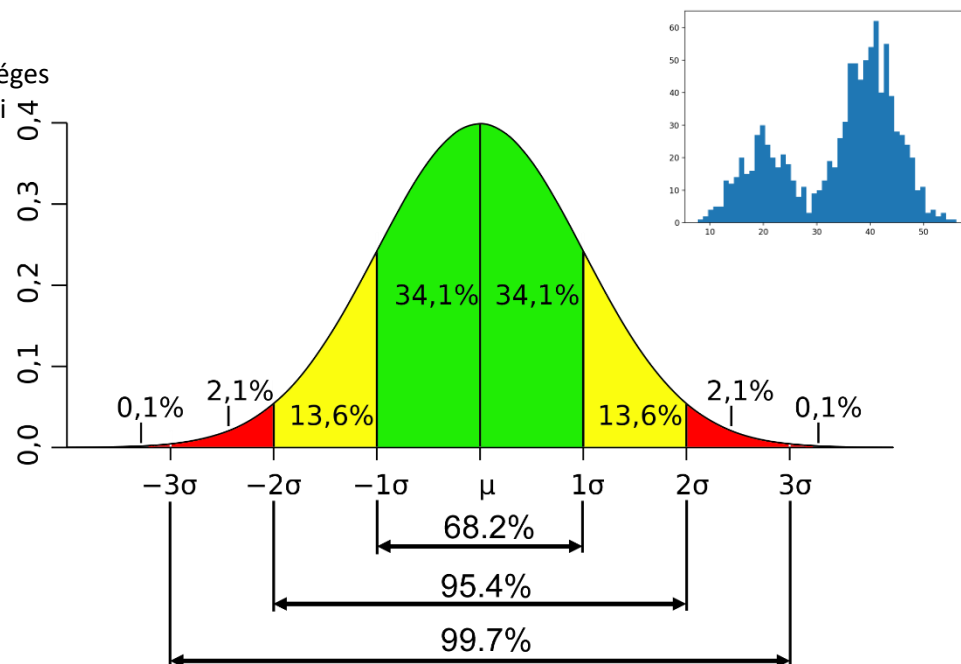


# Esemény, valószínűség, eloszlás.

valószínűség  
sűrűség  
 $\Delta P / \Delta x$



x  
lehetséges  
értékei



## Esemény:

Egy véletlen helyzet („kísérlet”) kimenetele  
Tehát valaminek a *megfigyelhető* bekövetkezése

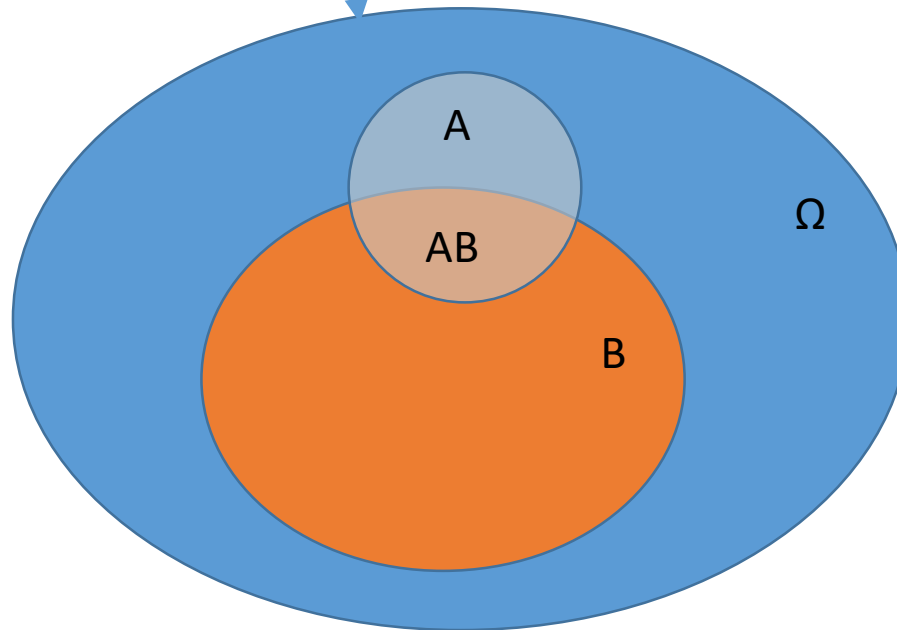


Egy eseményről egyértelműen meg  
tudjuk állapítani, hogy **bekövetkezett**,  
vagy **nem következett** be.

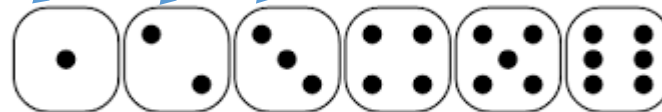


## Eseménytér és események

Vannak elemi események  
és összetett események



Ha minden **elemi eseményt** figyelembe vettünk, akkor meghatároztuk a teljes **eseményteret**.



$=\Omega$

## Esemény **gyakorisága**

Elvégzünk N db kísérletet, és megszámloljuk hányszor következett be az adott esemény.

Pl.:

- > a társasjátékban 5 dobásból hányszor dobunk 6-ost.
- > mennyiszer kell felkelni éjszaka az ügyeletben.
- > hány sürgősségi császármetszés fordul elő egy hónapban.

*Az esemény bekövetkezésének darabszámát hívjuk az esemény gyakoriságának.*

Gyakran k-val jelöljük, de lehet n-el, stb is. pl.  $k_A$  az A esemény megfigyelt gyakorisága.

## **Relatív** gyakoriság

Megadjuk az esemény gyakoriságát az elvégzett kísérletek *arányában*, tehát  $k_A/N$ -t adjuk meg. (lehet %-ban is megadni)

## Nagy számok törvénye

A relatív gyakoriság általában csak *körülbelül* ugyanaz ha több kísérlet-sorozatot hasonlítunk össze.

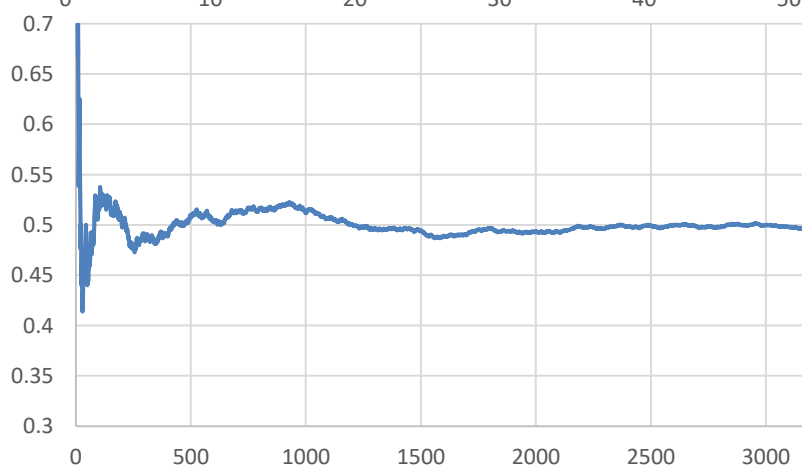
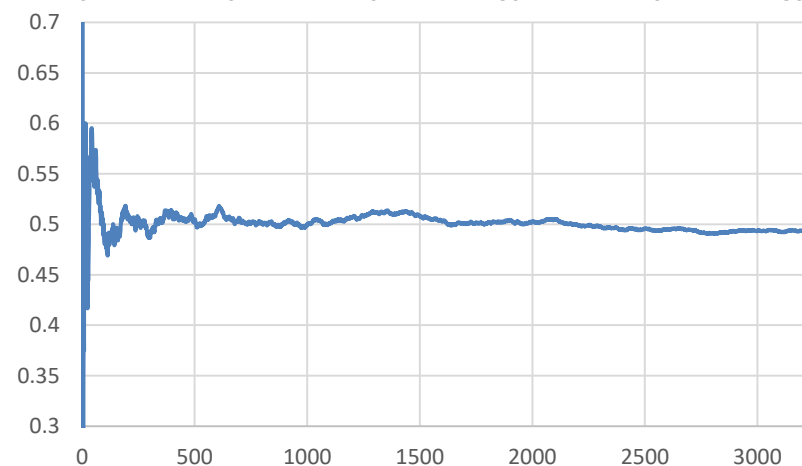
Azonban általában *stabilizálódik* ha a elég sok kísérletünk van.

**Az a szám ami körül a relatív gyakoriság stabilizálódik\* a valószínűség (P)**

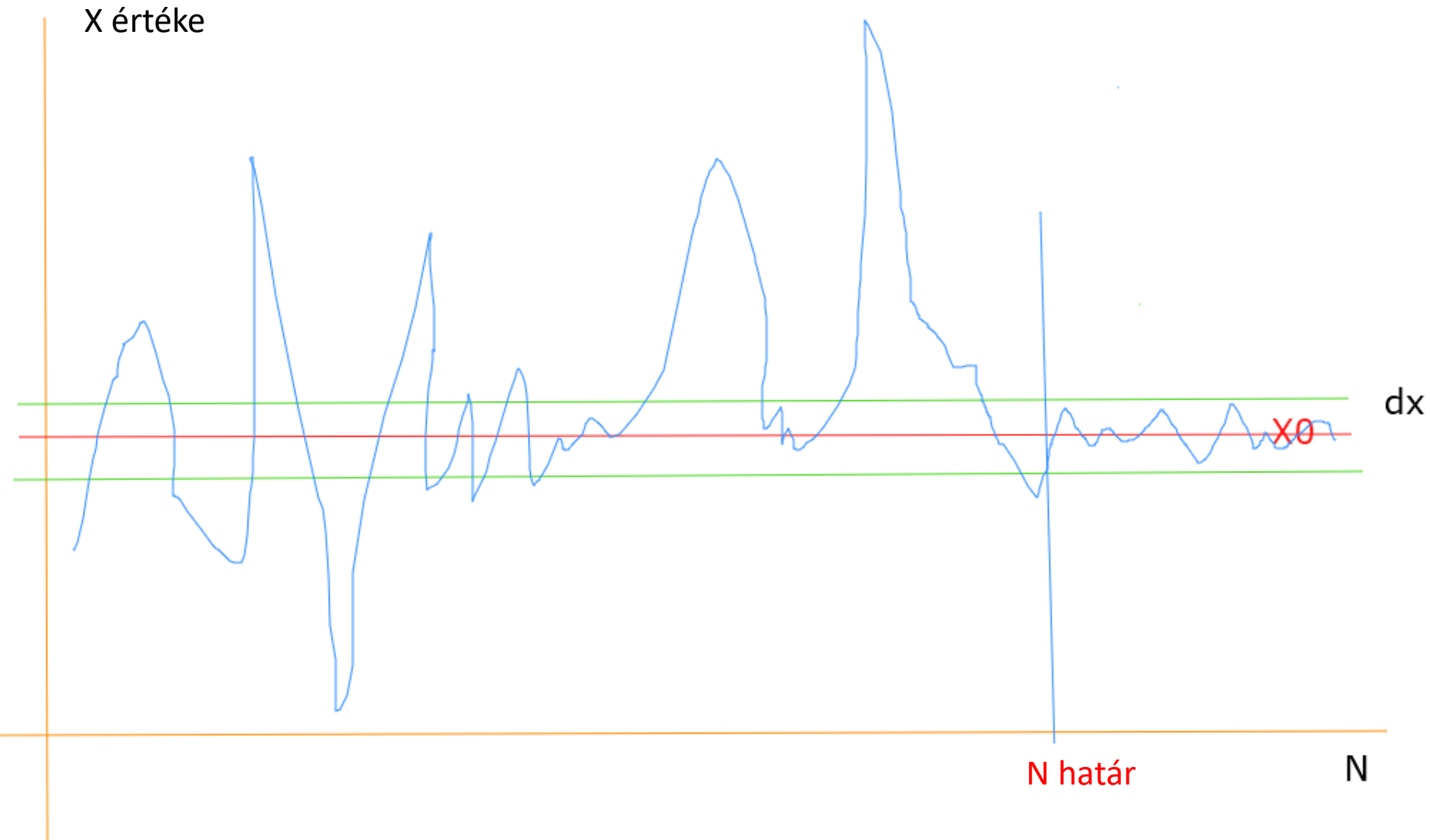
\*: ez a *frekventista* definíció

Létezik a Bayes-féle definíció is.

Fej vagy írás kísérletekből származó relatív gyakoriságok

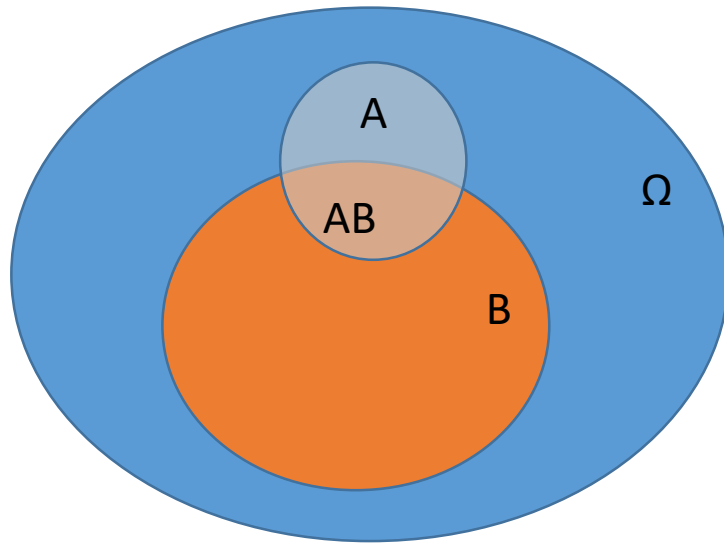


**Kiegészítés:** a határértéket úgy mondjuk, hogy az elvben létező  $X_0$  határértékhez, és  $dx$  maximális eltéréshez mindig megadható egy olyan  $N$  határnagyság, hogy ha ennél több azonos kísérletet végzünk el, akkor a relatív gyakoriság  $X_0$ -tól nem fog jobban eltérni mint  $dx$ .



## Események egymáshoz képesti viszonya

Az hogy mit tekintünk eseménynek tőlünk függ! (az ami minket éppen érdekel)



Lehetnek összetett események

- páros számot dobunk
- kapunk eltérést a vértékben

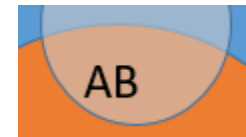
logikai összefüggések események között:



vagy



és



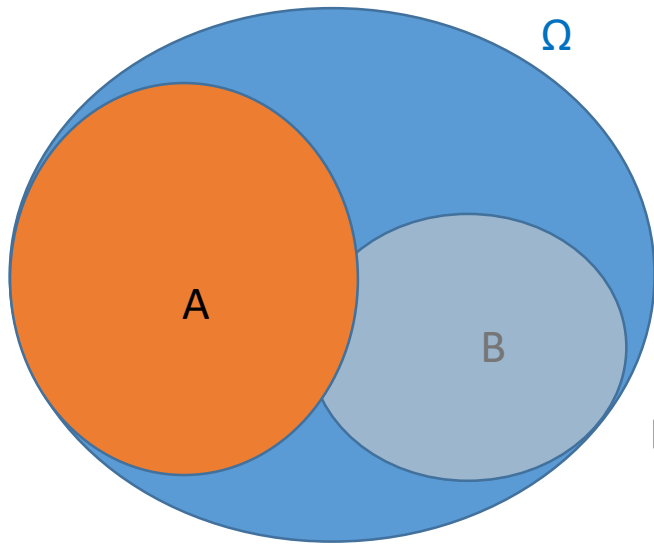
Az egyik valósul meg

Egyszerre valósulnak meg

**Egymást kizáró események:** CSAK az egyik valósulhat meg egyszerre, egy adott kísérletben  
pl. párosat vagy páratlant dobtunk? Egyszerre csak az egyik lehet, együtt nem fordulnak elő.

**Ellentett esemény:** Nem-A : azaz nem az A esemény következik be.

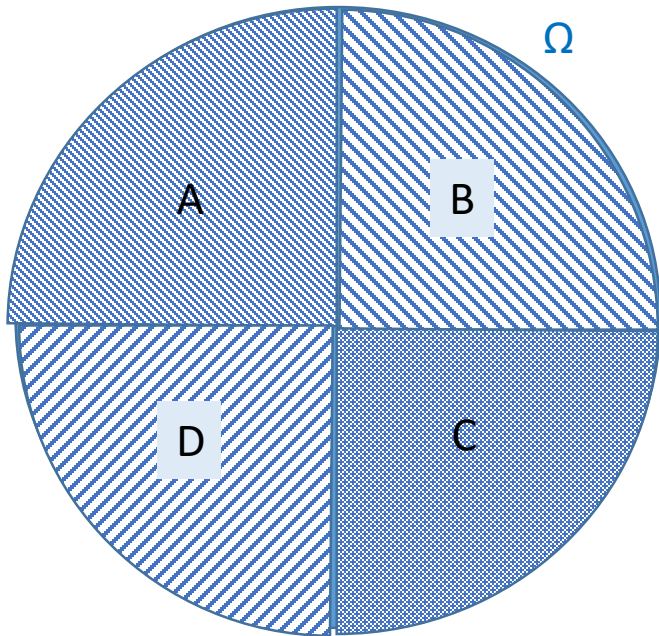
pl. „nem lázas a beteg”



## Egymást kizáró események

vagy-vagy

Egymást kizáró eseményekre igaz hogy  $P(A \text{ vagy } B) = P(A) + P(B)$



## Elemi események:

Olyan események melyeket nem lehet, vagy nem szükséges további darabokra bontani az adott kérdés vizsgálatakor.

Fontos, hogy az egész  $\Omega$  eseményteret lefedjék az elemi eseményeink!



## Szokásos jelölések:

és  $P(A \text{ és } B) = P(A * B) = P(A \cap B) = P(A \cdot B) = P(AB)$

vagy  $P(A \text{ vagy } B) = P(A + B) = P(A \cup B) = P(A \vee B)$

Nem  $P(\text{nem } A) = P(!A) = P(\sim A) = P(\bar{A})$

### Kolmogorov-axiómák:

$$0 \leq P(A) \leq 1$$

$$P(A \cdot \bar{A}) = P(\emptyset) = 0 \quad \text{lehetetlen esemény}$$

$$P(A + \bar{A}) = P(\Omega) = 1 \quad \text{biztosan bekövetkező esemény}$$

$\emptyset$  : üres halmaz, azaz a „semmi”

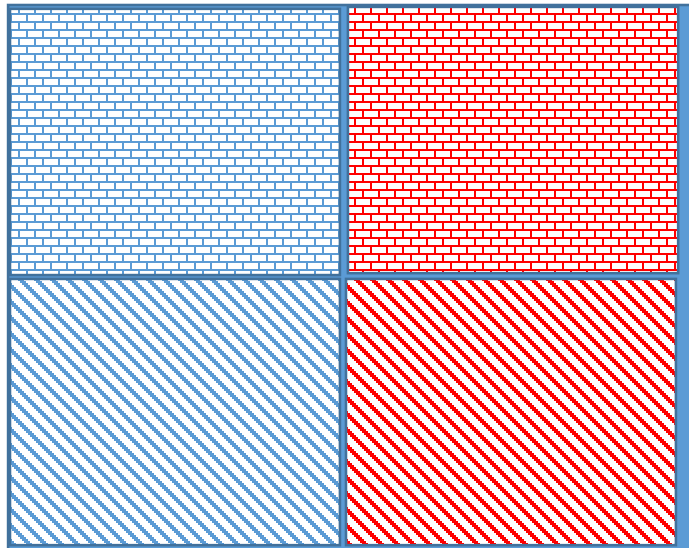
$\Omega$  : a teljes eseménytér azaz „bármilyen”

## stochasztikus függetlenség

Két esemény független, ha egymás gyakoriságát nem befolyásolják, tehát ha az egyik esemény bekövetkezése nincs hatással a másik esemény bekövetkeztére.

pl. az egyik páciens köhög, a következőnek pedig bőrgyógyászati problémája van.

Vagy a kockával először 2-est dobunk, utána pedig 5-öst. Mivel a kocka „nem emlékszik” így minden dobás teljesen új, és minden oldal egyformán esélyes



$$P(\text{rácsos}) = 2/4 = 0.5$$

$$P(\text{csíkos}) = 2/4 = 0.5$$

$$P(\text{kék}) = 2/4 = 0.5$$

$$P(\text{piros}) = 2/4 = 0.5$$

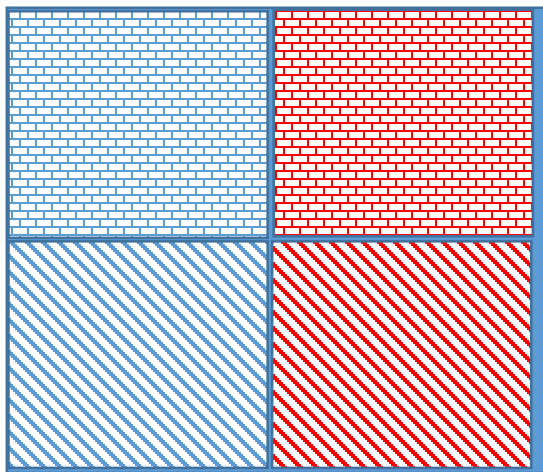
$$P(\text{csíkos ÉS piros}) = 1/4 = 1/2 * 1/2 = P(\text{piros}) * P(\text{csíkos})$$

**Általában: két esemény (A és B) akkor és csak akkor független,  
ha mindig teljesül rájuk hogy  $P(A \text{ és } B) = P(A) * P(B)$**

## Feltételes valószínűség

$P(A \mid B)$  = annak a valószínűsége, hogy A esemény bekövetkezik, ***feltéve hogy*** B bekövetkezett.

pl. :            lázas-e a páciens, *feltéve hogy* tudjuk hogy COVID-19 fertőzött.  
5-öst kapok statisztikából *feltéve hogy* eddig minden vizsgám jeles volt.



$\Omega$  egy része érdekel minket csupán, a kérdésünk az hogy azon belül milyen az előfordulási valószínűség.

$P(\text{kék} \mid \text{csíkos})$  = kékek valószínűsége a csíkosak között. Itt két csíkos van, abból egy a kék, tehát  $P(\text{kék} \mid \text{csíkos}) = \frac{1}{2}$

Megjegyzések:

1. független eseményekre  $P(A \mid B) = P(A)$
2. Két eseményre  $P(AB) = P(A \mid B) \cdot P(B)$  ez a „szorzási szabály” vagy Bayes-tétel

-> bővebben a Bayes-statisztika részben!

példák:

- Részhalmaz esete: Annak a valószínűsége hogy a páciensünk cukorbeteg (B) pl. 15%. Az hogy II. típusú legyen (A) pl. 10%.

Akkor annak a valószínűsége, hogy II. típusú cukorbetegségben szenved, HA tudjuk hogy cukorbeteg, tehát

$$P(A|B) = P(AB)/P(B) = P(A)/P(B) = 10\%/15\% = 66.7\%$$

Megjegyzés: ha A részhalmaza B-nek akkor A esetén B automatikusan teljesül, azaz  $P(AB)=P(A)$ .

- Független események: Mekkora valószínűséggel lesz a következő páciensünk szemüveges (A), HA az előző nő volt (B)?

$$\text{Tehát } P(A|B) = P(AB)/P(B) = P(A)*P(B) / P(B) = P(A)$$

Megjegyzés: itt felhasználtuk a függetlenséget, tehát hogy A-t nem befolyásolja B. Ekkor józan ésszel azt kell várjuk hogy B meg sem jelenhet az eredményben, és ez így is van 😊

## Kiegészítés:

A P értékből lehet további mérőszámot kiszámolni:

**Esély (Odds):**  $O = \frac{P}{1-P}$

Ez lényegében azt mondja meg, hogy **hányszor jobban hajlandó egy esemény bekövetkezni, mint nem-bekövetkezni.** Akkor nagyon hasznos, ha az esemény-terünk csak két részre bontott, pl.: COVID-19 pozitív **lesz vagy sem.**

**Rizikó, kockázat:** itt két – vélelmezhetően nem független - eseményt vizsgálunk egyszerre: Az egyik Odds-át nézzük meg akkor ha a másik, a **rizikófaktor** jelen van, vagy nincs jelen.

B      R

$P(\text{Betegség} | \text{Rizikó}) = \text{„a betegség kockázata ha van rizikófaktor jelen”} = P(B^*R)/P(R)$

$P(\text{Betegség} | \text{nincs rizikó}) = P(B^*\bar{R})/P(\bar{R})$

**Kockázati hányados = Rizikó hányados** = 
$$\frac{P(B | R) = P(B^*R)/P(R)}{P(B | \bar{R}) = P(B^*\bar{R})/P(\bar{R})}$$

Azaz: **hányszorosára növeli a meglévő rizikófaktor a betegség valószínűségét (kockázatát) ahhoz képest mintha nem lenne rizikófaktor jelen.**

**Esélyhányados:** Az esélyek hányadosa rizikófaktor mellett és anélkül.

(Odds Ratio)

$$OR = \frac{O_{B|R}}{O_{B|\bar{R}}}$$

Ez valójában 4 db feltételes valószínűség ismeretét jelenti a gyakorlatban

## Nevezetes eloszlások

Sok olyan eloszlás van aminek a matematikai alakja megadható, azaz az  $P(x)$  függvény ismert. Folytonos változó esetén gyakran az integrált, vagy kumulatív függvényt adják meg, tehát a  $P(\xi < x)$ -et (ahol most a görög  $\xi$  jelöli a valószínűségi változót)

Ezeket nevezetes eloszlásoknak hívjuk, sokszor alakítjuk át később az elsődleges valószínűségi változót úgy, hogy az átszámított érték valamely ismert, nevezetes eloszlást kövessen.

Példák (lesz ami később előkerül még): **Nem kell tudni felsorolni!**

- normál- vagy Gauss-eloszlás

- t-eloszlás

- egyenletes eloszlás

- exponenciális eloszlás

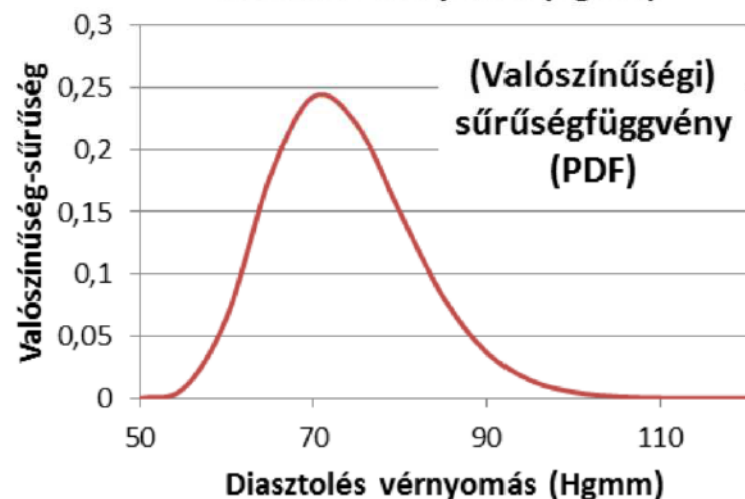
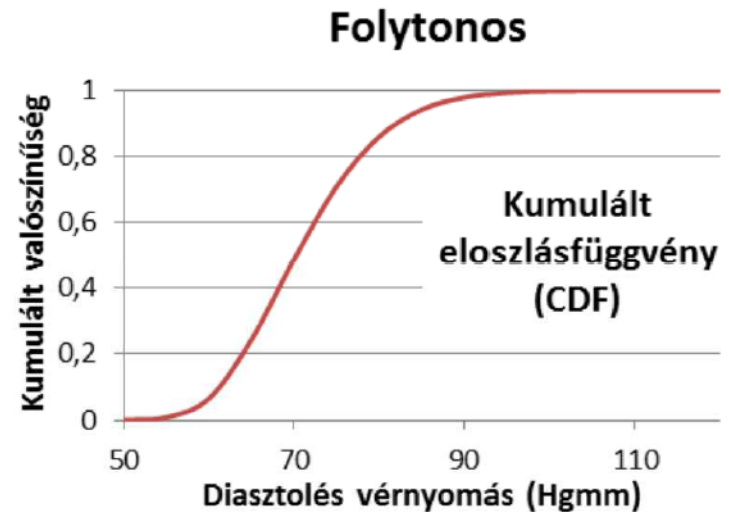
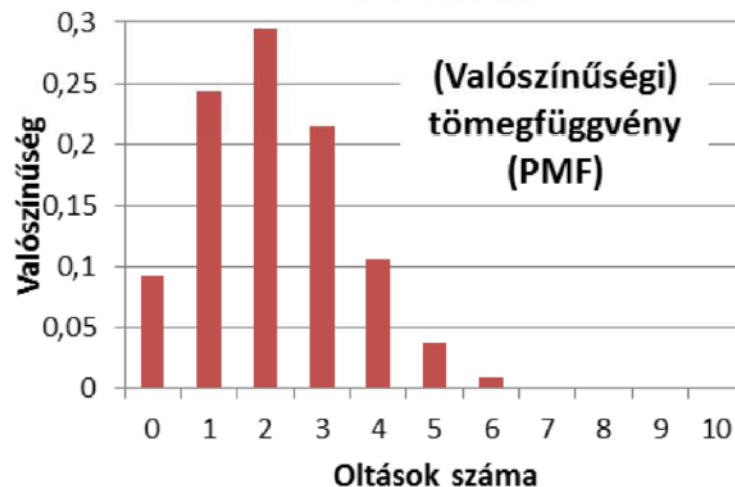
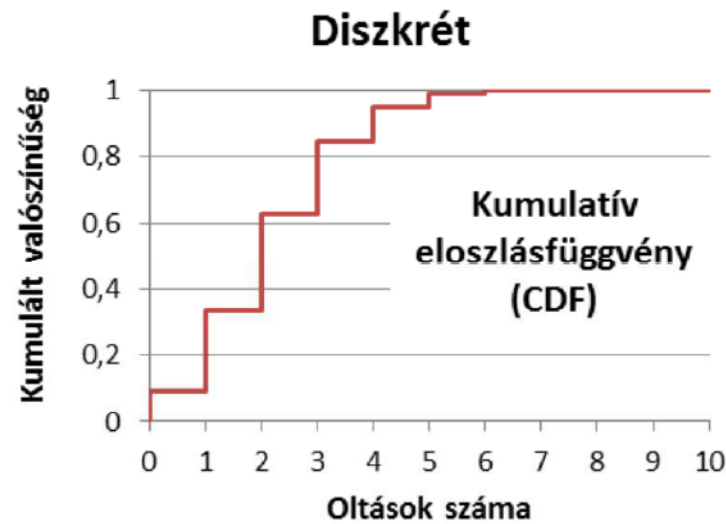
- binomiális eloszlás

- $\chi^2$ -eloszlás

- geometrikus eloszlás

- log-normál eloszlás

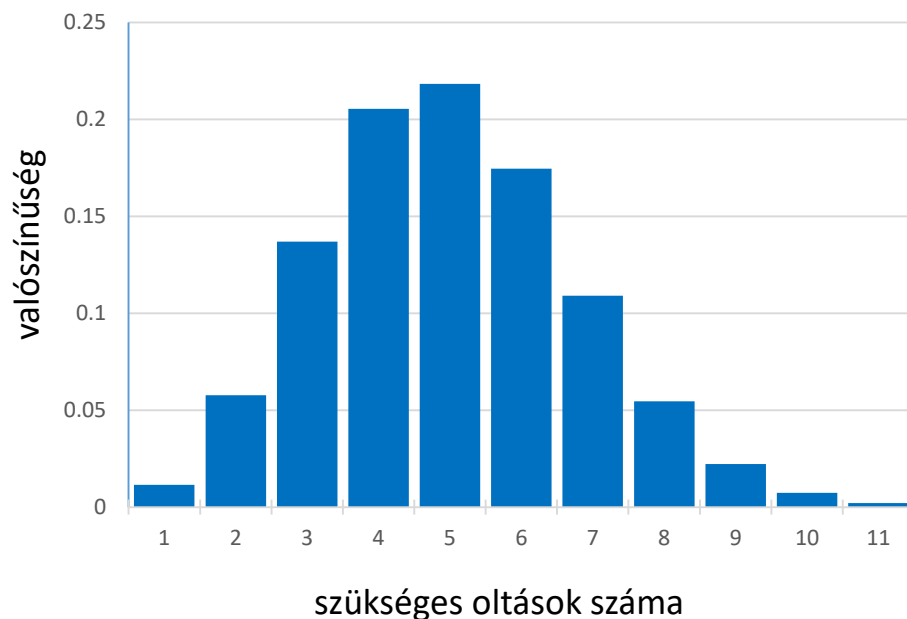
## Többféleképpen lehet eloszlásokat ábrázolni



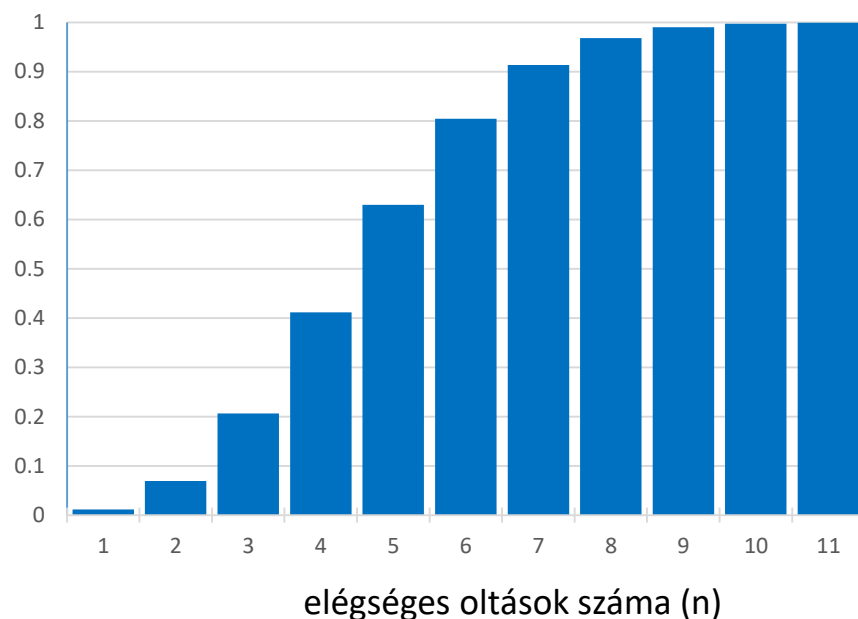
Attól függően szoktunk választani, hogy kumulált eloszlást akarunk-e, és diszkrét vagy folytonos a megfigyelt valószínűségi változónk.

## Binomiális (Bernoulli) eloszlás

Megmutatja hogy N kísérletből mekkora valószínűséggel következik be A esemény k-szor. Azaz  $P(A,k,N)$ -t kaphatjuk meg minden  $k \leq N$  esetre, feltéve hogy  $P(A)$  ismert.



$P(x)$  függvény  
(tömegfüggvény)

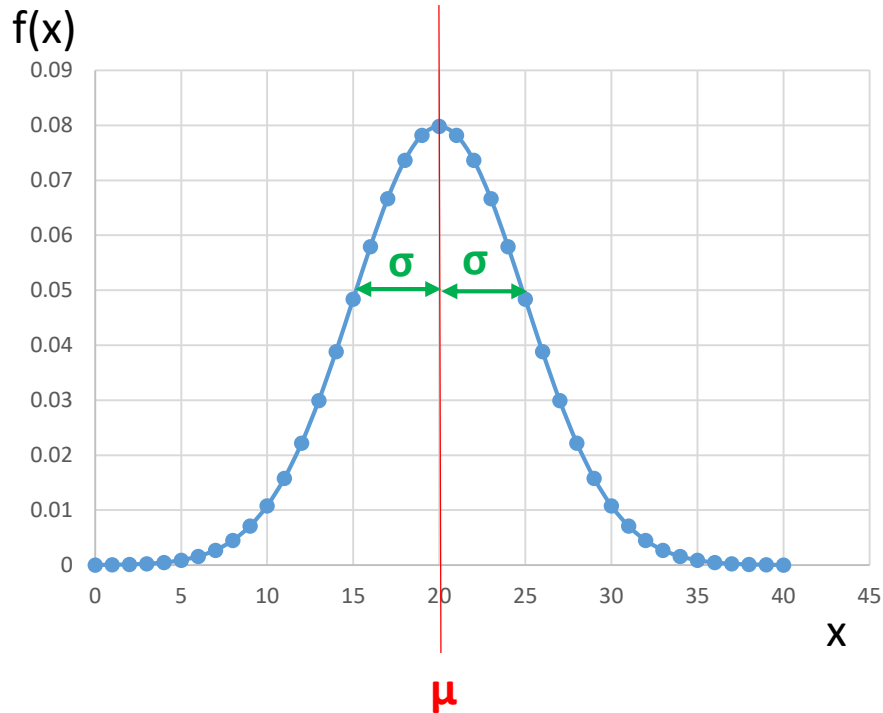


$P(x < n)$  függvény  
(kumulatív függvény)

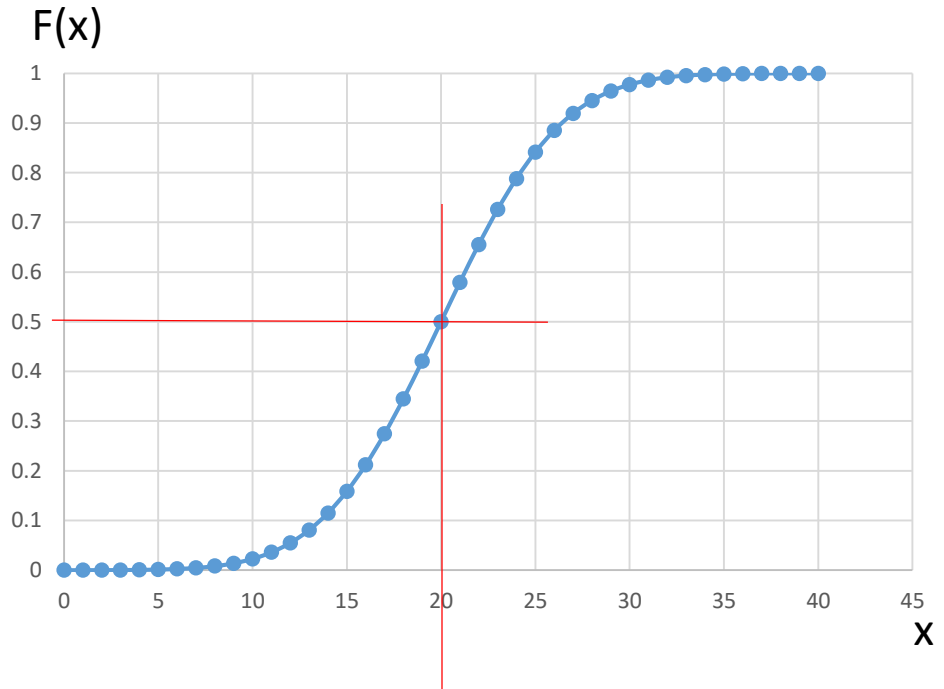
*példa:* Ha a múlt hónapban 3000 esetből volt 12 akut operációt igénylő, és ma en vagyok ügyeletben, várhatóan 20 esettel, akkor mire számítsak?



## Normál (Gauss) eloszlás



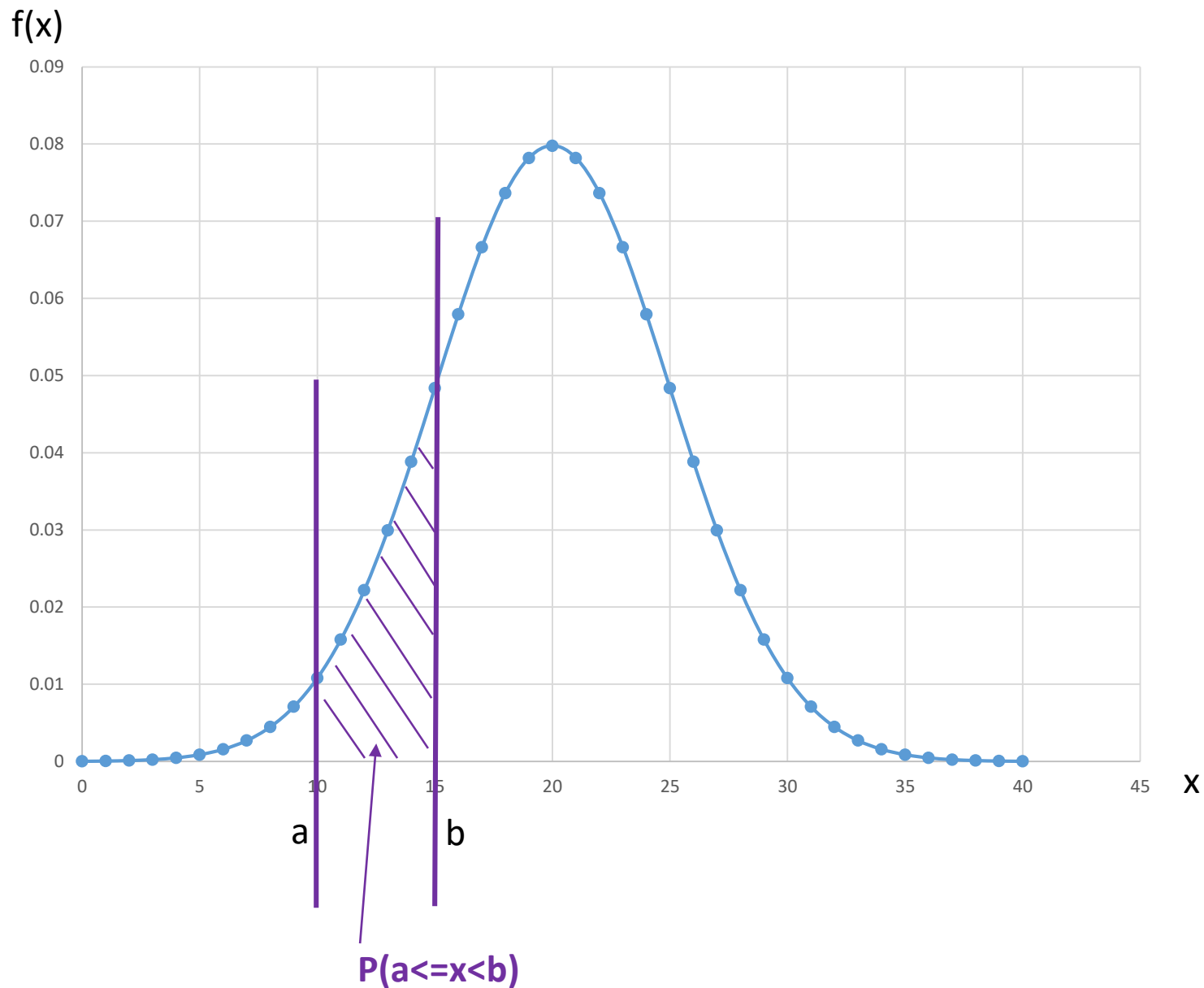
sűrűségfüggvény



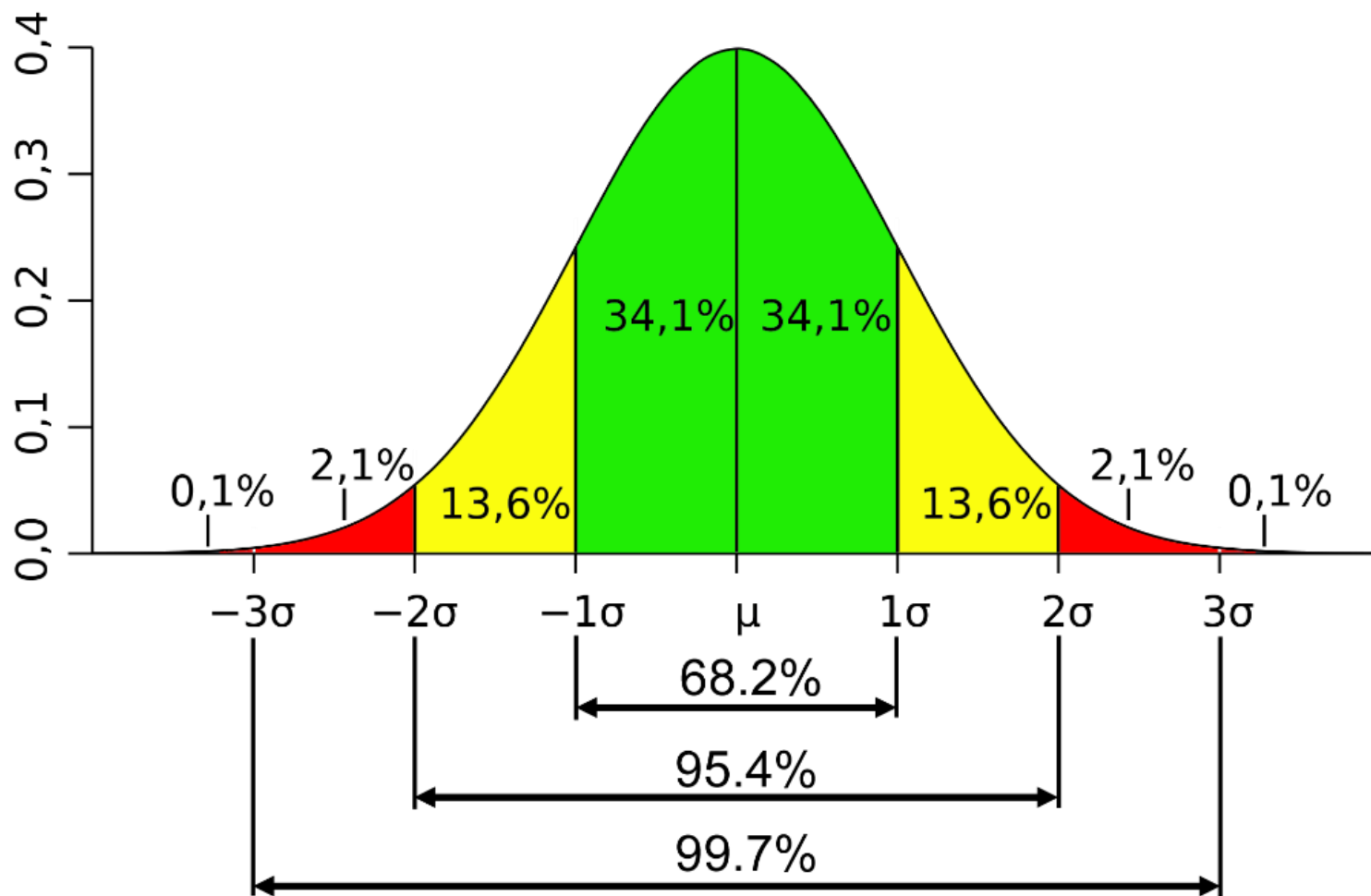
(kumulatív) eloszlásfüggvény

Folytonos változó esetén jöhet szóba, két jellemző paramétere van, a közepet megadó várható érték  $\mu$ , és a szélességet megadó szórás  $\sigma$ .

**Vigyázat! A sűrűségfüggvény értéke maga nem valószínűség, csak a görbe egy darabja alatti terület az!**

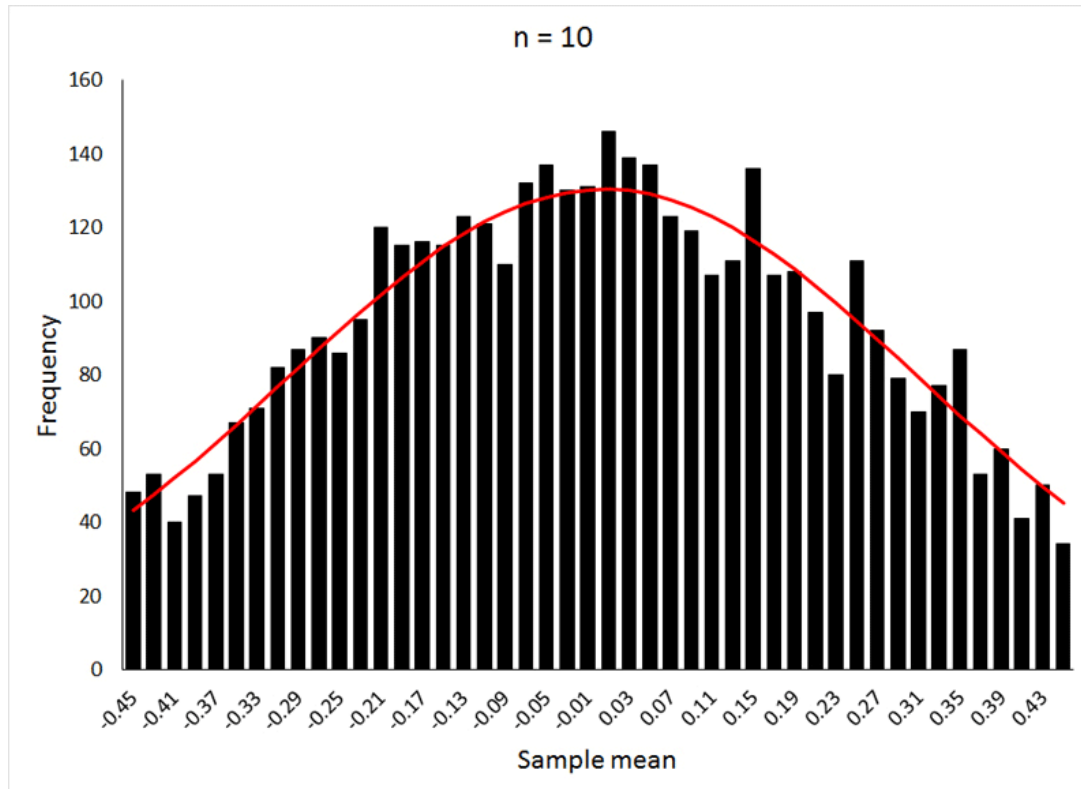


## Nevezetes területek a Gauss görbe alatt



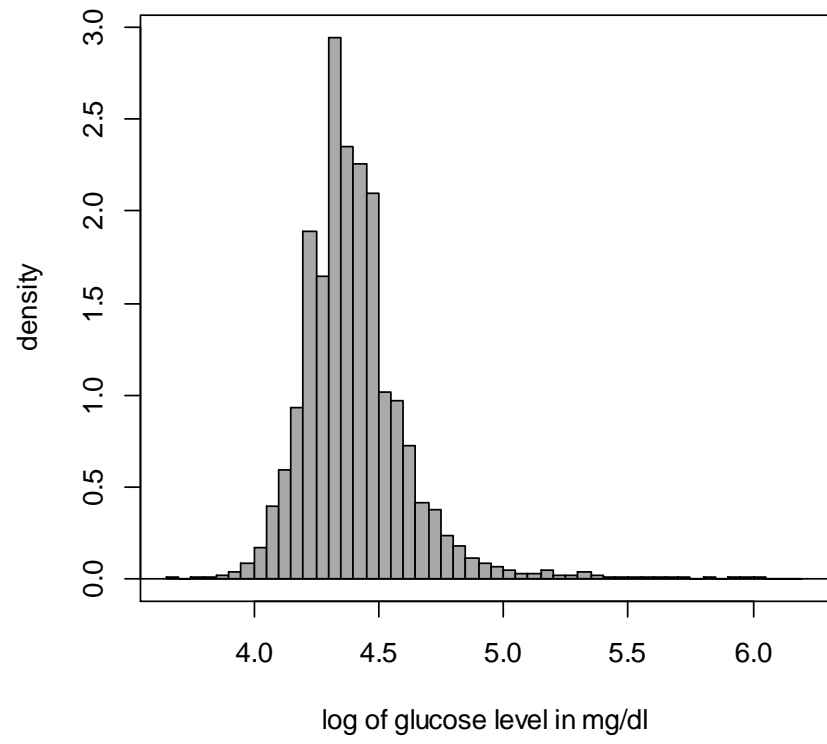
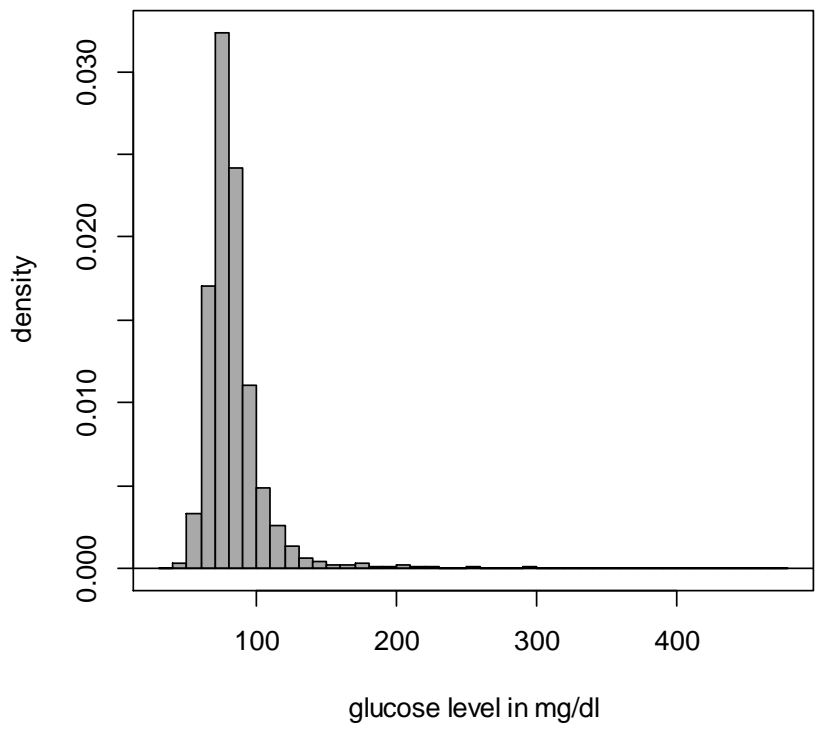
Nagyon sokszor találkozunk jó közelítéssel normál eloszlású változóval

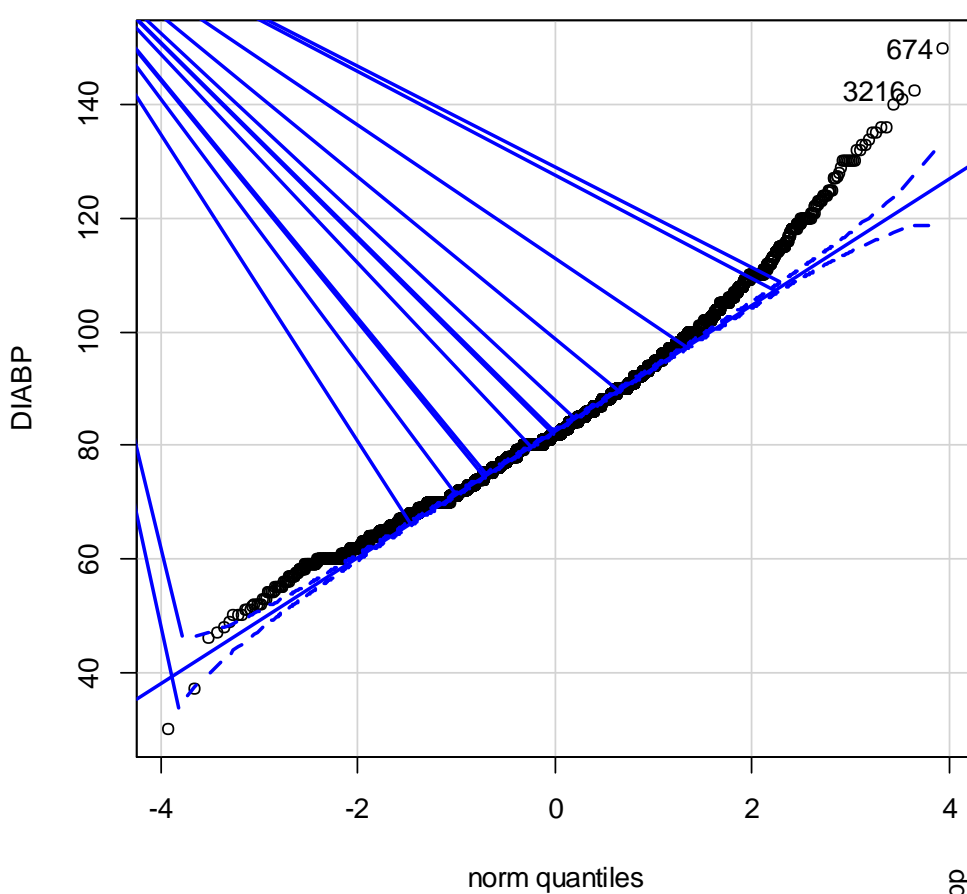
A becslések egyik célja az, hogy a valószínűségi változóhoz tartozó paramétereket ( $\mu, \sigma$ ) megpróbáljuk megmondani a minta alapján.



-> Id következő óra!

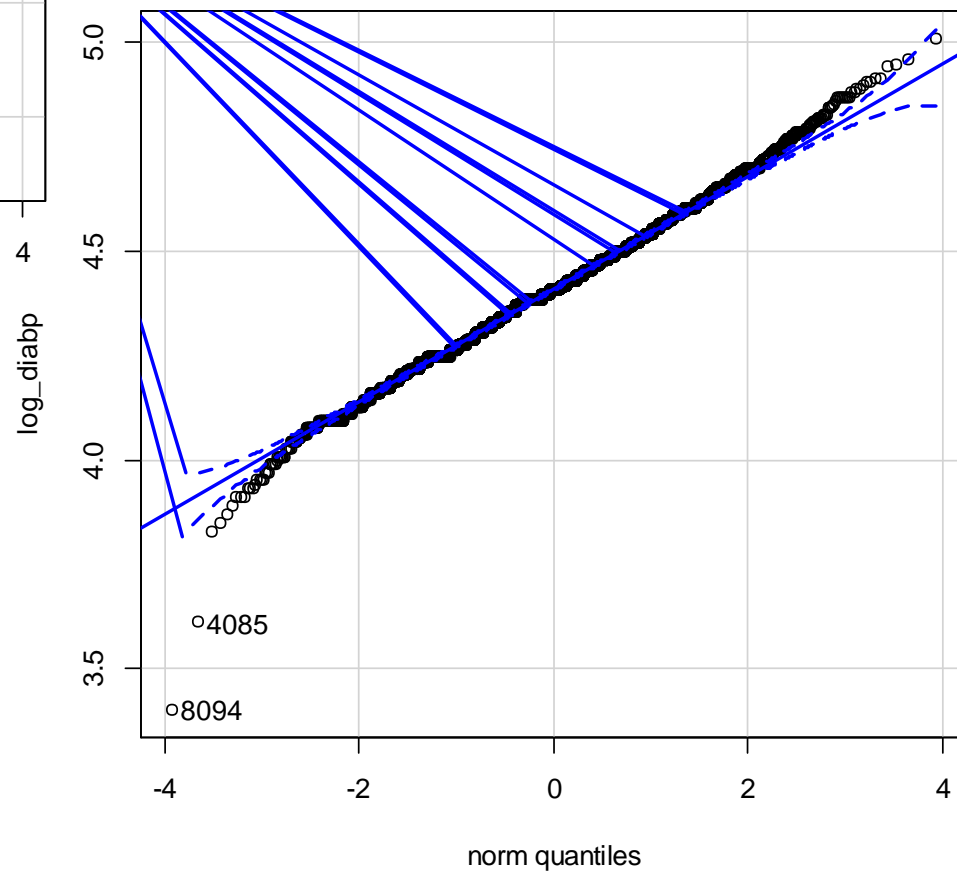
sokszor előfordul, hogy a mért értékek logaritmusa mutat inkább normál eloszlást. Ez a lognormál eloszlás. (sok vér-paraméter ilyen)





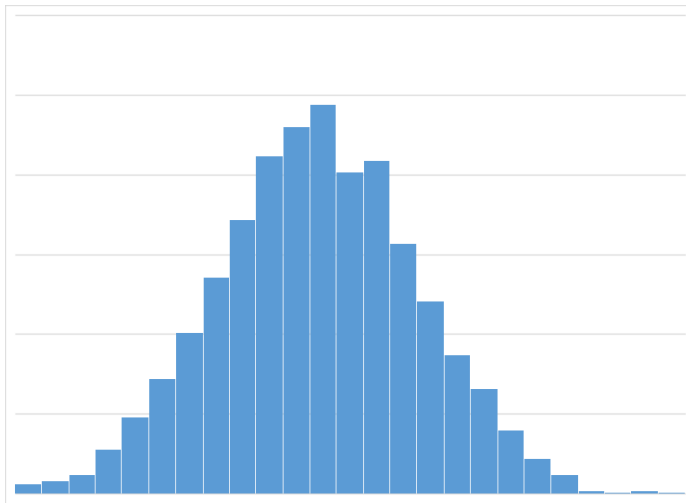
QQ plot  
(quantile-quantile plot)

a valós kvantilisokat összevetjük egy ismert elméleti eloszlásból valókkal, így látható hogy az adatok követik-e az adott eloszlást.



## Centrális határ-eloszlás tétele

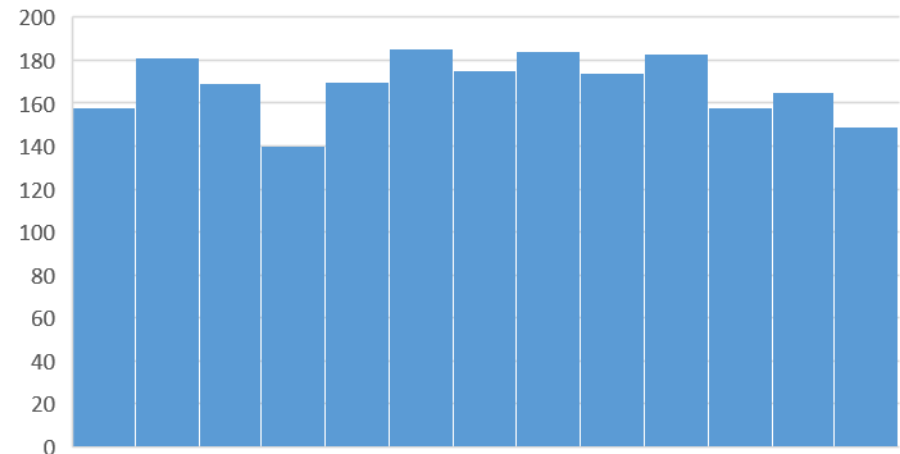
Ha van egy valószínűségi változónk amit **nagyon sok** (lehetőleg azonos eloszlású) egymástól független véletlen változó egyszerre határoz meg, akkor ez a „közös” változó normál eloszlást fog követni. Annál jobban, minél több véletlen határozza meg.



Az átlag eloszlása

30 db elemi változó átlaga

Egyes elemi változók eloszlása



A nevezetes, jól ismert eloszlások sokmindenre használhatóak 😊

-> hipotézis-vizsgálatok, döntések