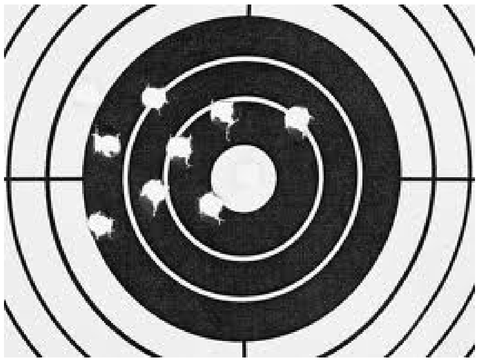
Statisztikus törvényszerűség

Mindig vannak számba nem vehető körülmények is (céltábla)



Modelljeit a valószínűségszámítás szolgáltatja

Alapfogalmak:

Jelenség: minden, ami *lényegében azonos feltételek mellett* megismétlődhet, amivel kapcsolatban megfigyeléseket lehet végezni, lehet vele „kísérletezni”.

Megfigyelés, „kísérlet”: megadjuk, hogy a jelenséggel kapcsolatban mire vagyunk kíváncsiak, illetve, hogy azt hogyan érzékeljük vagy hogyan mérjük.

Esemény: egy állítás, ami vagy bekövetkezik, vagy nem.

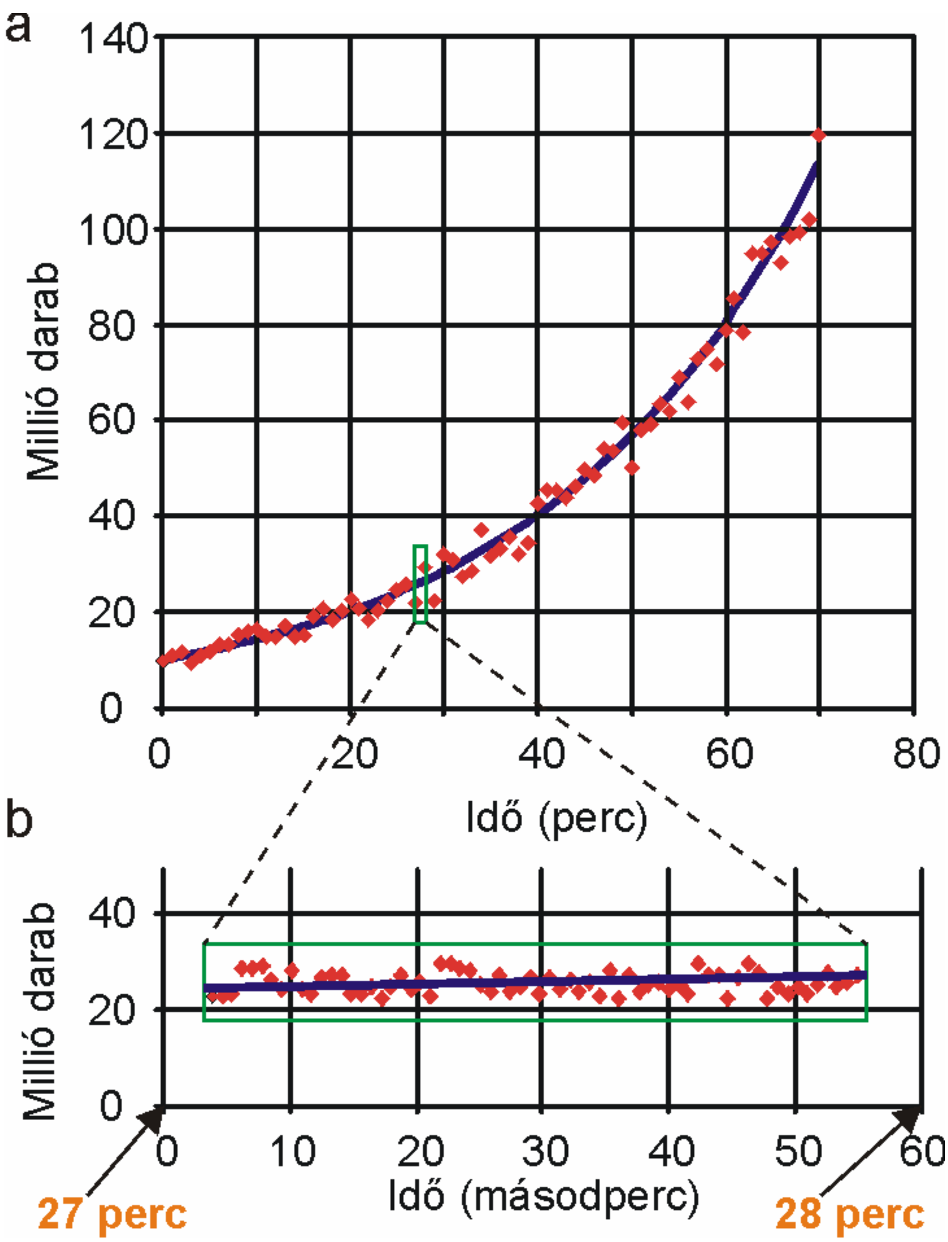
	példák			
Jelenség	orvosi vizsgálat	pénzfeldobás (1)	várakozás a buszra	pénzfeldobás (2)
Megfigyelés	a beteg bőrének színe	az érme repülési ideje	a várakozók száma	az érme melyik oldalára esik
Esemény	sárga	0,5 és 1,5 s között van	tíz	fej

Ha gyakrabban következik be, akkor valószínűbb.

A baktérium kolónia szaporodása elméletben, a megfelelő determinisztikus matematikai modell szerint (kék görbe) és gyakorlatban, a mérések alapján (piros szimbólumok).

Elmélet (modell) **Gyakorlat** (meg kell mérni)

$$N(t) = N_0 2^{\frac{t}{T}}$$



A változások determinisztikus és statisztikus része mindig **egyszerre** fordul elő.

Van-e mód a szétválasztásukra?

Első közelítésben a statisztikai tevékenységeket négy csoportba sorolhatjuk, de ezek között nincs éles határ:

1. adatgyűjtés,
2. az adatok áttekinthetővé tétele,

ehhez nincs szükség a valószínűség fogalmára:

leíró statisztika

3. az adatok elemzése,
4. következtetések

a valószínűségszámítás alapjai nélkül nem nagyon érthető:

induktív statisztika

1. adatgyűjtés (később visszatérünk erre is)

az adatgyűjtés valamilyen **cél** eléréséhez szükséges

az adatok
egy része ismert, csak meg kell kérdezni valakitől,
másik részét csak meg kell figyelni, egy
harmadikat meg kell mérni valahogy (orvosi vizsgálat)

2. az adatok áttekinthetővé tétele

A mindennapi életben is gyakran előfordul, hogy egy probléma kapcsán viszonylag sok adat áll rendelkezésünkre.
Ilyen esetekben szükséges, hogy az adatokról valamilyen áttekintésünk legyen.

2/a táblázat

Kórokozó	*Betegség*	**abszolút gyakoriság**		**relatív gyakoriság**		**feltételes relatív gyakoriság**	
	Salmonellosis (szalmonella fertőzés)	94		0,280		0,452	
baktérium	Scarlatina (skarlát)	102	208	0,303	0,619	0,490	1,000
	egyéb bakteriális eredetű	12		0,036		0,058	
	Hepatitis infectiosa (fertőző májgyulladás)	22		0,065		0,175	
vírus	Mononucleosis infectiosa	22	126	0,065	0,375	0,175	1,000
	Lyssa (veszettség)	74		0,220		0,587	
	egyéb vírusos eredetű	8		0,025		0,063	
egyéb	egyéb fertőző betegségek	2	2	0,006	0,006	1,000	1,000
összesen		336	336	1,000	1,000		

abszolút gyakoriságok

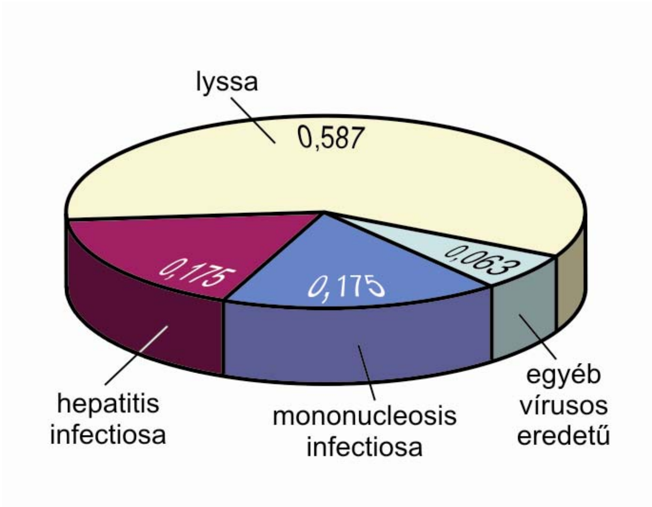
relatív gyakoriságok

hányados, ezért nemcsak azt kell tisztázni, hogy minek a relatív gyakoriságáról beszélünk, hanem azt is, hogy mihez viszonyítunk

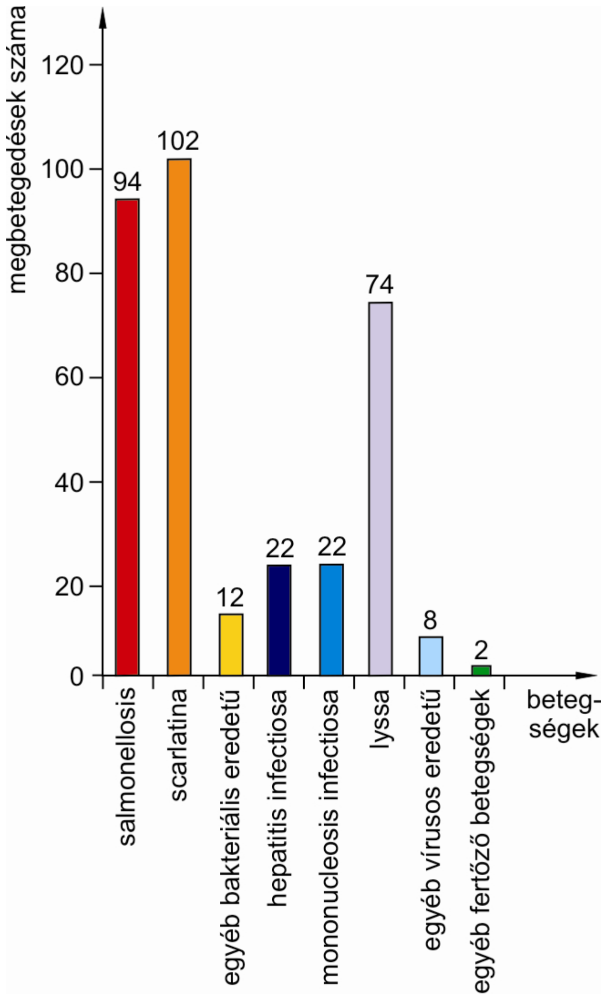
feltételes relatív gyakoriságok

a különbség csupán annyi, hogy itt szűkebb összességhez viszonyított relatív gyakoriságról van szó

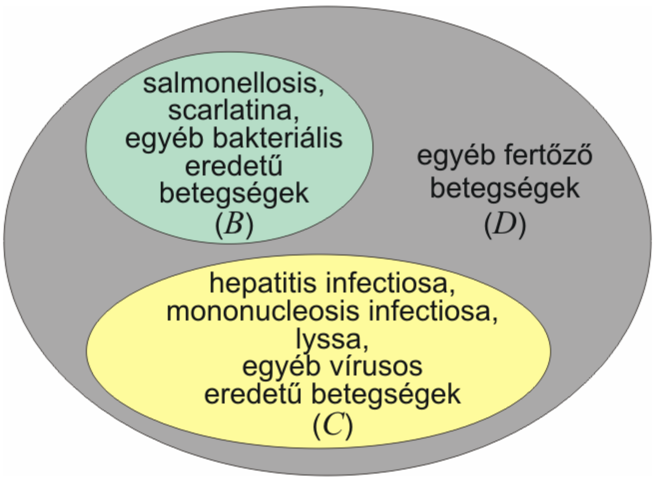
2/b grafikon



torta diagram



oszlop diagram



A táblázatban szereplő fertőző betegségek, mint **részhalmazok**

$$B \cup C \cup D = A$$

$$B \cap C \cap D = \emptyset$$

A halmazok és az események megfeleltethetők egymásnak

A vírusos eredetű megbetegedés, mint esemény

	példa
Jelenség	orvosi vizsgálat
Megfigyelés	a fertőző betegség eredete
Esemény (C)	vírusos eredetű

Összegzési (I) és szorzási (II) szabályok

Feladat:

Egyetemünkön a tavalyi vizsgákon az elégtelen osztályzatok relatív gyakorisága 0,15, a sikeres vizsgák között a jelesek relatív gyakorisága 0,2 volt. Az összes vizsgajegy között mennyi volt a jeles relatív gyakorisága?

$$1.\quad \overset{0{,}15}{\frac{\text{elégtelenek száma}}{\text{összes hallgató száma}}} + \overset{0{,}85}{\frac{\text{sikeres vizsgák száma}}{\text{összes hallgató száma}}} = 1$$

$$\downarrow$$

$$2.\quad \overset{0{,}2}{\frac{\text{jelesek száma}}{\text{sikeres } \cancel{\text{vizsgák száma}}}} \cdot \overset{0{,}85}{\frac{\text{sikeres } \cancel{\text{vizsgák}} \text{ száma}}{\text{összes hallgató száma}}} = \overset{0{,}17}{\frac{\text{jelesek száma}}{\text{összes hallgató száma}}}$$

(I)

az abszolút gyakoriságok mindig összeadhatók

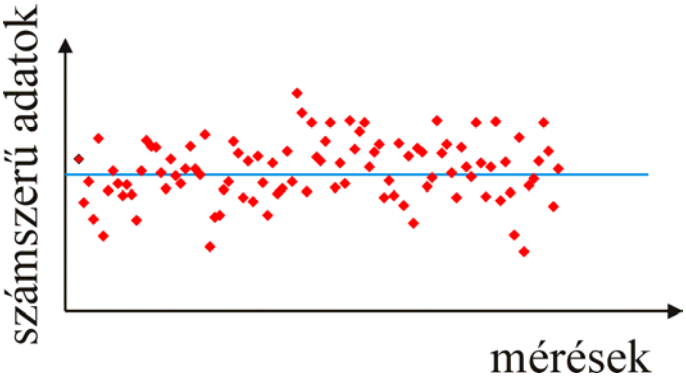
a relatív gyakoriságok is összeadhatók abban az esetben, ha ugyanahhoz az összességhez viszonyítjuk őket

(II)

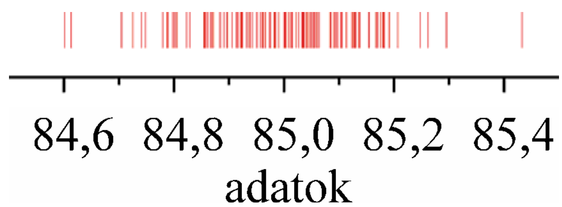
a feltételes relatív gyakoriság és a (feltétel nélküli) relatív gyakoriság a fentiek szerint összeszorozható

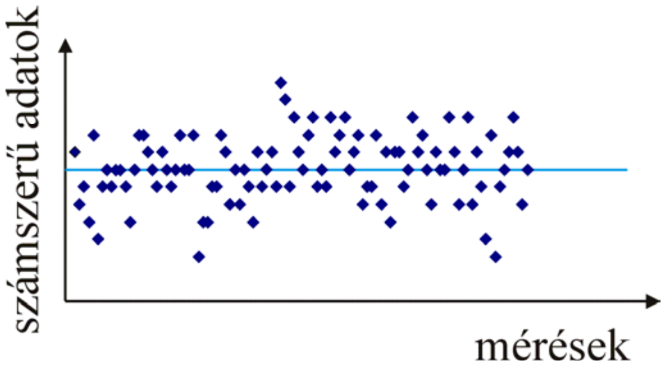
Számszerű adatok jellemzői

Kiindulásképpen tegyük fel, hogy egyelőre csak olyan adatokat vizsgálunk, ahol nincs determinisztikus változás. (100 adat)

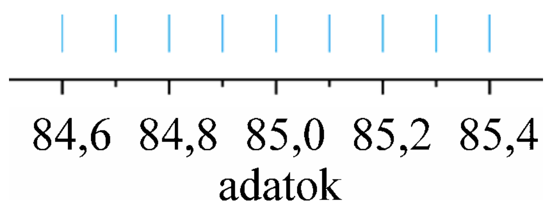


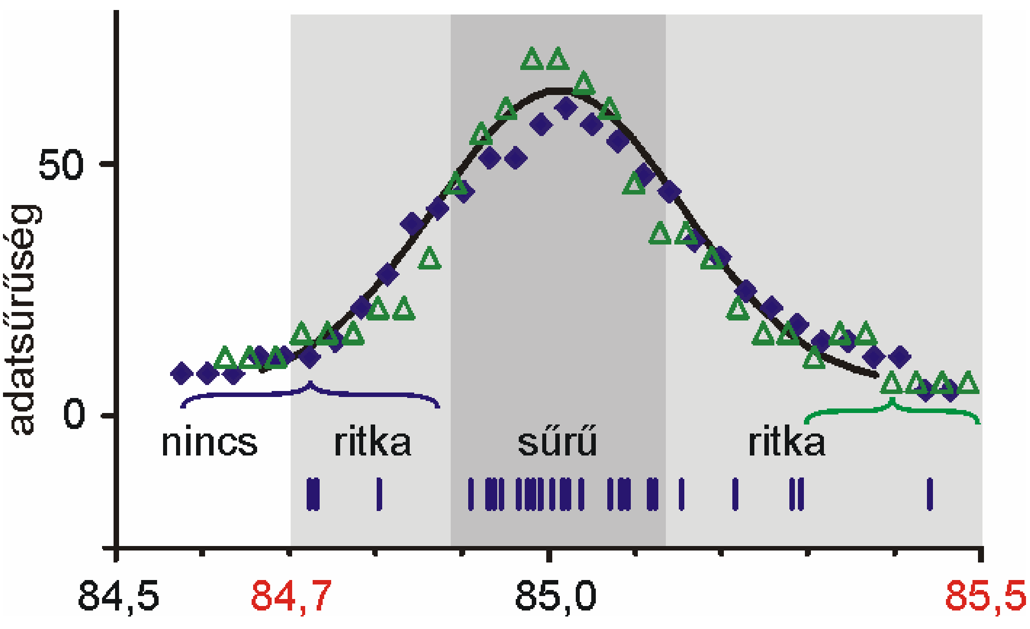
„folytonos" eset
„soha sincs"
két azonos adat
(a sorrend nem számít)



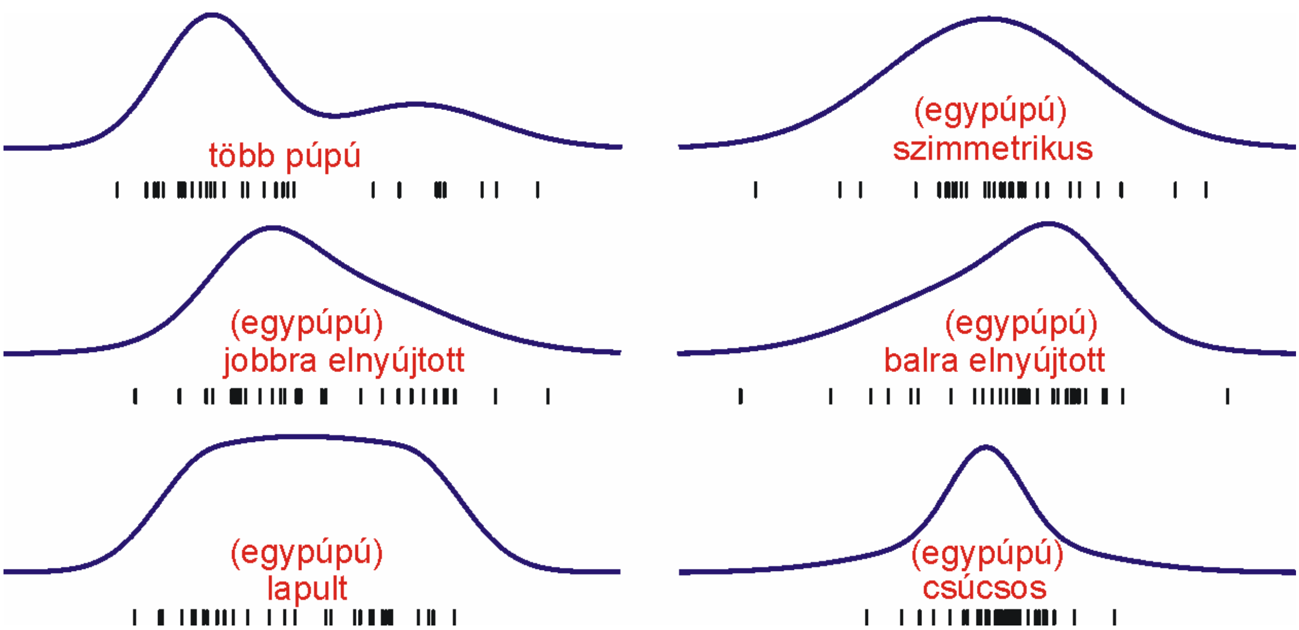


diszkrét eset
meg kell adnunk a
gyakoriságokat is



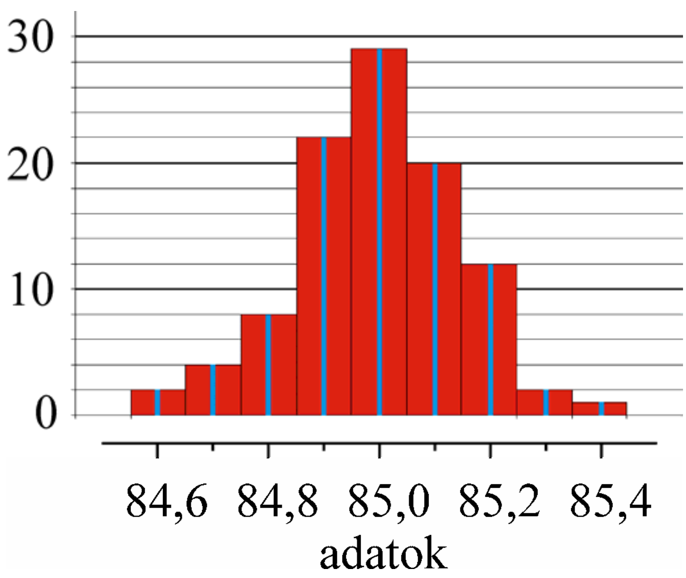


Adatsűrűség jellemzése,
jellegzetes típusok



Gyakorisági eloszlás

A diszkrét esetben egyértelmű.
A folytonos esetben alakja függ az intervallumok, más néven az **osztályok** megválasztásától (de általában nem nagyon).



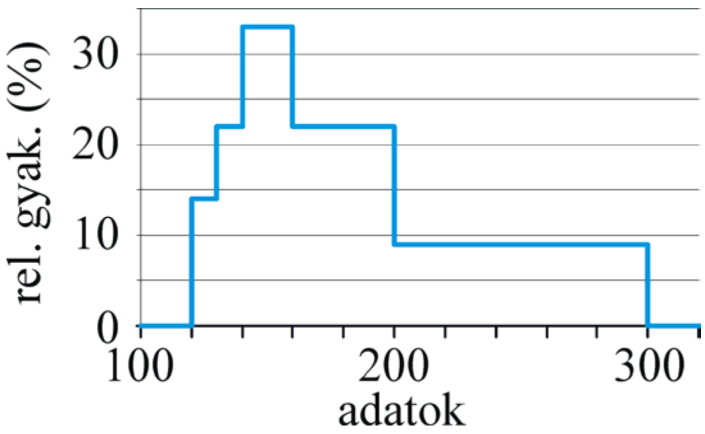
A jobb összehasonlíthatóság kedvéért sokszor a relatív gyakoriságokat adjuk meg.

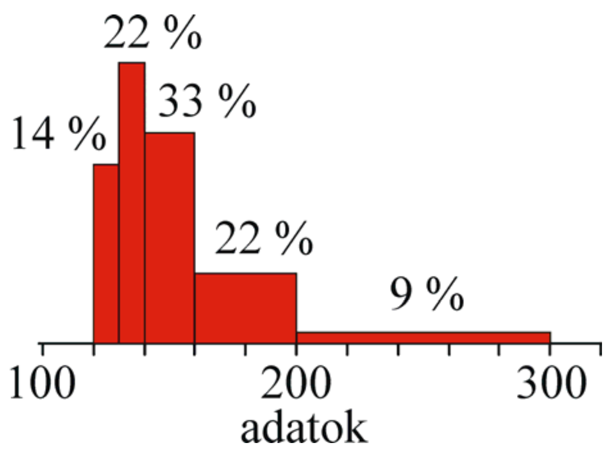
Hogyan járjunk el akkor, amikor nem ismerünk minden adatot?

Pl. Fizetési kimutatás (Ft):

120 és < 130 ezer között	124	14%
130 és < 140 ezer között	195	22%
140 és < 160 ezer között	293	33%
160 és < 200 ezer között	195	22%
200 és 300 ezer között	80	9%
összesen	887	100%

Hogyan szemléltessük?





Hisztogram

az oszlopok területe adja meg a relatív gyakoriságokat.
A teljes terület 100 %, vagyis 1.

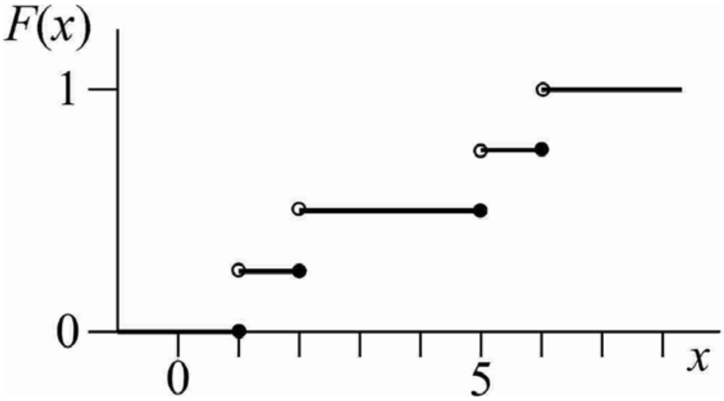
Amennyiben az oszlopok egyforma szélesek,
– tehát egyenközű osztályszélességek esetén –
a két ábrázolás azonos.

Eloszlásfüggvény

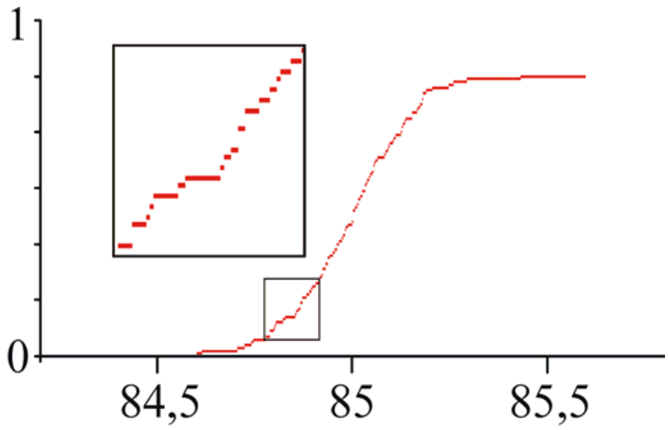
Az $x_1, x_2, x_3, \ldots, x_n$ adatrendszer F eloszlásfüggvénye:

$$F(x) = \frac{\text{az } x - \text{nél kisebb adatok darabszáma}}{n}$$

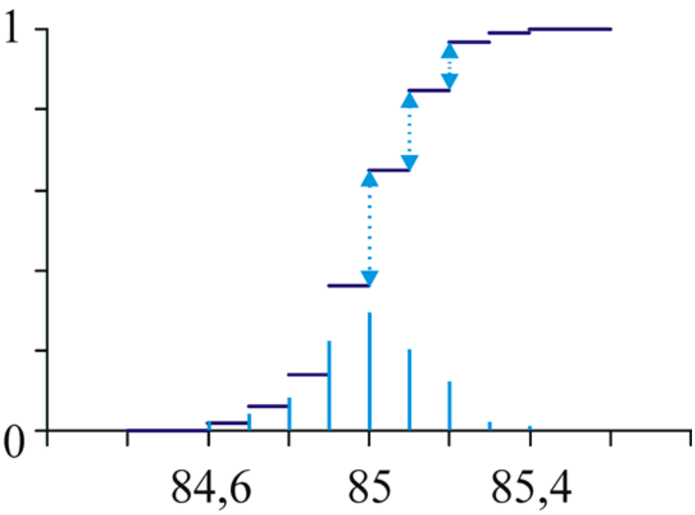
Pl. (1) [1, 2, 5, 6]



Pl. (2) A korábbi 100 adat folytonos esetben.



Pl. (3) A korábbi 100 adat diszkrét esetben.



Az oszlopok a relatív gyakoriságokat mutatják. Ezeket egymás után összeadva megkapjuk az eloszlásfüggvény megfelelő értékeit. Az állítás megfordítása is igaz. Az eloszlásfüggvény különbségei a relatív gyakoriságokat adják meg.