1. óra

Leíró statisztika (1)

1. <u>Példa:</u>

Matematika A2 év végi eredmények a hallgatók között.

jegy	gyakoriság	jel	jegy*db	rel. gyak
5	7	v_5	35	7/55=0,12
4	10	v_4	40	10/55=0,18
3	23	v_3	69	23/55=0,44
2	15	v_2	30	15/55=0,27
summ	55		174	174/55=3,16

A jegyek: **diszkrét változó**, értéke csak 2-3-4-5 lehet, pl. 3.2 nem.

gyakoriság: a megfigyelt esemény előfordulásának száma (pl. ötös 7 db.; négyes 10 db.)

relatív gyakoriság: gyakoriság/összes események száma (pl. ötös esetén 7/55; négyes esetén 10/55)

A megfigyelés-sorozat legegyszerűbb jellemzője az **<u>Átlag:</u>**

ξ_1 ; ξ_2 ; ξ_3; …; ξ_{55} (megfigyelt jegyek sorozat 55 hallgató esetében)

$$\bar{\xi} = átlag = \frac{1}{55}\sum_{i=1}^{55} \xi_i$$

Általában:

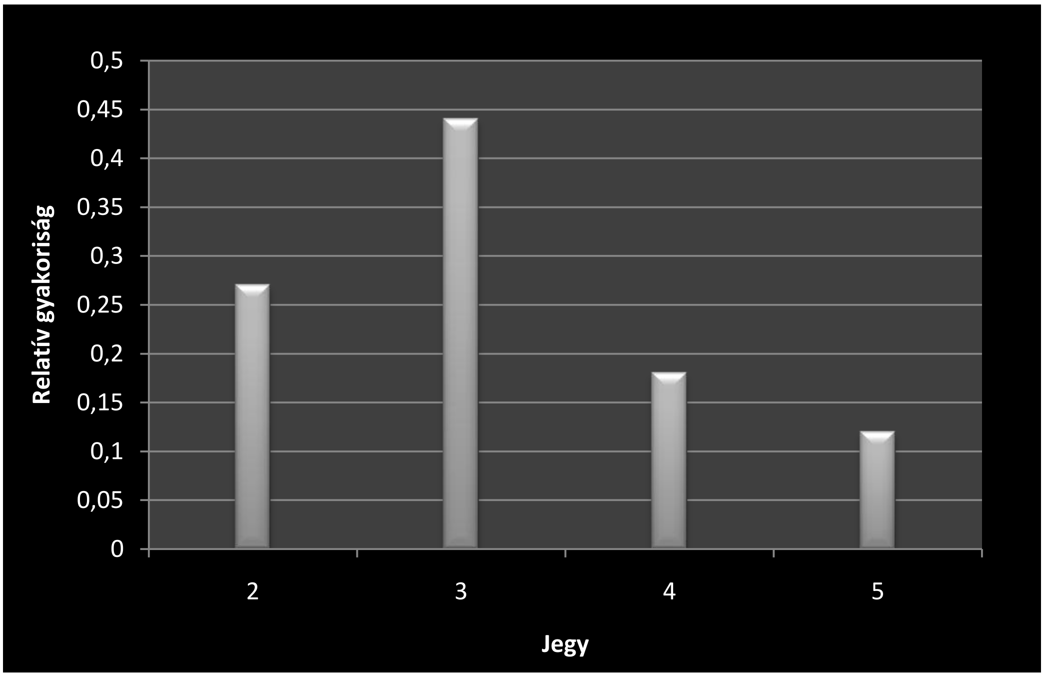
$$\bar{\xi} = \frac{1}{n}\sum_{i=1}^{n} \xi_i =$$

Egyszerűbb számolni, ha csoportokba foglalom:

$$\bar{\xi} = \frac{1}{n}(v_5 * \xi_5{}^* + \ldots + v_2 * \xi_2^*)$$

$$\overline{\xi} = \frac{1}{n}\sum_{j=2}^{5} v_j\xi^* = \sum_{j=2}^{5} \frac{v_j}{n}\xi^* = 3.16$$

A jegyek eloszlását mutatja a **<u>tapasztalati sűrűségfüggvény</u>, v. <u>hisztogram</u>**. A változó értékeihez felmérem a relatív gyakoriságot.



hisztogram (- tapasztalati sűrűségfüggvény) diszkrét változó esetén

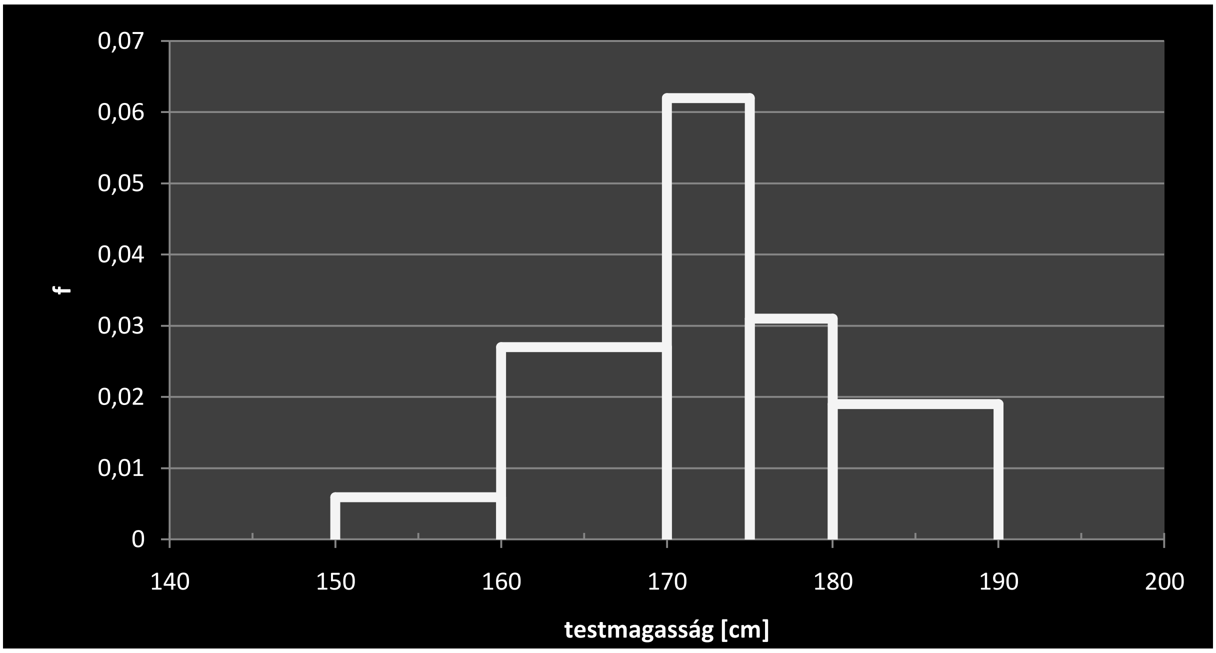
2. Példa:

Folytonos változó: pl. a hallgatók magassága

magasság cm-ben	gyakoriság	relatív gyakoriság	f
150-160	3	0,059	0,0059
160-170	14	0,27	0,027
170-175	16	0,31	0,062
175-180	8	0,155	0,031
180-190	10	0,19	0,019
190-	0	0	0

Intervallumok fölé téglalapokat rajzolunk, amik területe (T_j) az intervallumba esés relatív gyakorisága. Ha az intervallum szélessége Δx_j, akkor a téglalap magassága (f_j):

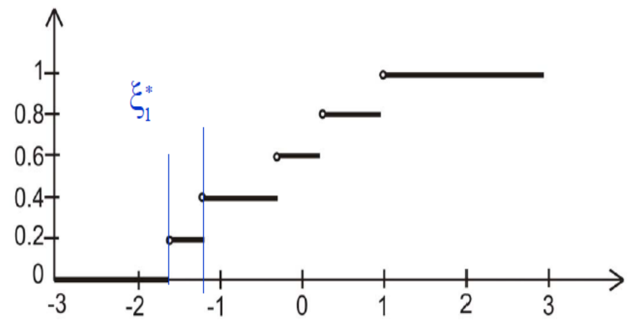
$$f_j = \frac{T_j}{\Delta x_j} = \frac{\nu_j / n}{\Delta x_j}$$



Hisztogram folytonos változó esetén

A felmérés csalóka, mert a fiúk és a lányok nem lettek külön felmérve, pedig az átlag magasságuk eltér.

<u>Tapasztalati eloszlásfüggvény:</u> (lépcsős függvény)



$$F(x) = \frac{v_1+v_2+\cdots+v_n}{n}$$

Kapcsolat a tapasztalati sűrűségfüggvény, és a tapasztalati eloszlásfüggvény között:

$$\frac{v_1}{n} + \frac{v_2}{n} + \cdots + \frac{v_n}{n} = \sum f_j \Delta x_j$$

Ha van n db. megfigyelésünk, hogyan választjuk az intervallumok számát és a határait? Gyakorlati tapasztalat, hogy a részintervallumok száma:

J= intervallumok száma $\cong \sqrt{n}$ (ha n≤ 100)
J= 1+$\log_2 n$ (ha n > 100)

<u>median</u>: megfigyelés sorozat azon eleme, amely alatt ugyanannyi elem van, mint felette. Legyen n a megfigyelések száma, ha n páratlan, akkor az m-edik elem a medián, ahol m=(n+1)/2

ha n páros, akkor a medián a „két középső” elem átlaga

$$\tilde{\xi} = \frac{1}{2}\left(\xi_{\frac{n}{2}} + \xi_{\frac{n}{2}+1}\right)$$

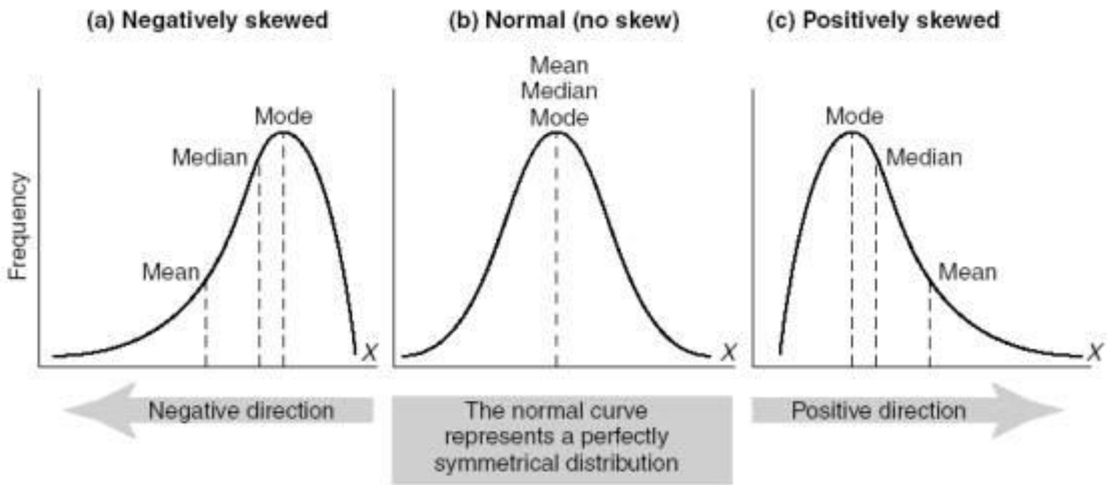
median és átlag kapcsolata:

ha az eloszlás szimmetrikus, akkor $\tilde{\xi} \cong \bar{\xi}$

ha az eloszlás „ferde”,

ha „jobbra”dől (Negatively skewed), akkor a medián > átlag

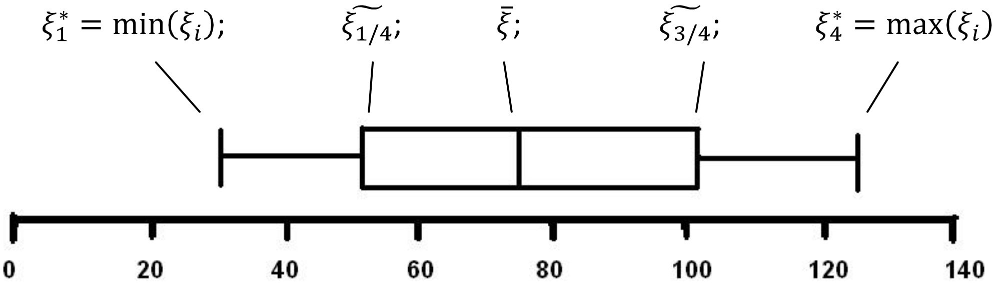
ha „balra” dől (Positively skewed), akkor a medián < átlag



Boxplot:

megfigyelési sorozat ábrázolása. Jellemző vonalai:

- **medián (**$\bar{\xi}$), ez két részre vágja a megfigyelési sorozatot.
- Az alsó mintafél mediánja az alsó negyedelő vagy **alsó kvartilis (** $\widetilde{\xi_{1/4}}$),
- a felső mintafél mediánja a felső negyedelő vagy **felső kvartilis (**$\xi_{3/4}$).
- Legkisebb mintaelem ξ_1^*
- a legnagyobb mintaelem: ξ_4^*



Fogalmak:

Gyakoriság:

Az f_i gyakoriság (abszolút gyakoriság) azt mutatja, hogy a sokaságnak hány egysége tartozik az i-edik osztályba.

n - kísérletek száma

$k=f_i$ - bennünket érdeklő események darabszáma

(például n=45 hallgató vizsgázott, és k=3 kapott 5-öst)

Relatív gyakoriság:

A gyakoriságból g_i=fi/N vagy k/n módon nyerhető relatív gyakoriság nem más, mint az fi abszolút gyakoriságból számított hányados. Gyakran %-os formában fejezzük ki.

Diszkrét változó:

Ezekben az esetekben a változó csak megadott értékeket vehet fel, például az osztályzatok csak 1, 2, 3, 4, vagy 5 értéket vehetnek fel

Folytonos változó:

A változó egy intervallumon belül bármely értéket felvehet (A diszkrét változóval ellentétben köztes értékek is felvehetőek. Például a magasság, mely, fél értékeket is felvehet. 170-180 centis tartományon belül lehet valaki 173, de 179,5 centis is.)

Tapasztalati eloszlásfüggvény $F_n(x)$:

Tetszőleges x értéket felvéve, megszámoljuk, hogy hány megfigyelt érték kisebb, mint az x. Ebből relatív gyakoriságot számolunk, és ezt a relatív gyakoriságot mérjük fel x-hez. Lépcsősfüggvény, amely minden megfigyelési értéknél (legalább) 1/n-et lép felfelé.

Tapasztalati sűrűségfüggvény (hisztogram) $f_n(x)$:

Diszkrét változó esetén a változó értékéhez felmérjük az értékhez tartozó relatív gyakoriságot.

Folytonos változó esetén intervallumokra bontjuk azt a tartományt, ahova a megfigyelt értékek esnek. Minden intervallum fölé olyan téglalapot rajzolunk, amelynek területe egyenlő az intervallumba esés relatív gyakoriságával.

Medián:

A nagysága szerint sorba rendezett értékek közül a középső érték. Vagyis az az érték, amelynél kisebb, illetve nagyobb értékek gyakorisága megegyező.

Forrás: Kerékgyártó Györgyné, L. Balogh Irén, Sugár András, Szarvas Beatrix: Statisztikai módszerek és alkalmazásuk a gazdasági és társadalmi elemzésekben