



MATEKING.HU

Képletgyűjtemény

STATISZTIKA ÉS VALSZÁM ALAPOK tantárgy

Kiadás dátuma: 2026. 02. 05.

Tartalomjegyzék

Statisztika alapfogalmak.....	2
Valszám alapok, kombinatorika.....	6
Teljes valószínűség tétele és Bayes tétel.....	7
Mintavétel, binomiális és hipergeometriai eloszlás.....	8
Valószínűségi változók.....	10
Várható érték és szórás.....	13
Nevezetes eloszlások.....	14
Egy ismerv szerinti elemzés.....	17
Két ismerv szerinti elemzés.....	19
Standardizálás.....	25
Indexek.....	27

Statisztika alapfogalmak

Az ismérvek olyan vizsgálati szempontok, amelyek alapján a [sokaság](#) részekre osztható.

Vannak olyanok, amik csak két részre osztják a sokaságot, például azokra, akik megbuktak statisztikából és azokra akik nem. Olyan is van, ami mondjuk öt részre osztja a sokaságot, sőt olyan is lehet, hogy végtelen sok részre osztja.

Négy fő ismérvfajta különböztethető meg:

- területi
- időbeli
- mennyiségi
- minőségi

[Megnézem a kapcsolódó epizódot](#)

Az ismérveken belül négyféle mérési szintet tudunk megkülönböztetni.

A nominális (névleges) mérési skála a [sokaság](#) elemeit valamilyen tulajdonság szerint csoportokba sorolja, de a csoportok között nincsen semmilyen sorrendiség.

Ilyenek pl.:

- Az áldozatok halálának oka (MINŐSÉGI)
- A terroristák származási országa (TERÜLETI)
- A pilóták állampolgársága (MINŐSÉGI)

[Megnézem a kapcsolódó epizódot](#)

Az ismérveken belül négyféle mérési szintet tudunk megkülönböztetni.

Az ordinális (sorrendi) mérési skála a [sokaság](#) elemeit valamilyen tulajdonság szerint csoportokba sorolja, és a csoportok között van sorrendiség.

Ilyenek pl.:

- A [statisztika](#) vizsga osztályzata (MINŐSÉGI)
- A hotelek csillagos értékelése (MINŐSÉGI)

[Megnézem a kapcsolódó epizódot](#)

Az intervallum-skála a [sokaság](#) elemeit valamilyen mérés szerint rendezi sorba.

Az intervallum-skála jellegzetes tulajdonsága, hogy a "mennyivel több" kérdésre választ tud adni, de a "hányszor annyi" kérdésre nem.

Ilyenek pl.:

- Hőmérséklet (MENNYISÉG)
- Születési dátum (IDŐBELI)

[Megnézem a kapcsolódó epizódot](#)

Az arány-skála a [sokaság](#) elemeit szintén valamilyen mérés szerint rendezi sorba, de abban különbözik az intervallum-skálától, hogy ennek van valódi nullpontja.

Innen ered az arány-skála elnevezés is, van értelme az egymáshoz viszonyított arányoknak.

Ilyenek pl.:

- Életkor (MENNYISÉGI)
- Testmagasság (MENNYISÉGI)

[Megnézem a kapcsolódó epizódot](#)

Hogyha egy [sokaság](#) elemei egymástól jól elkülöníthető egységek, akkor a [sokaság](#) diszkrét.

Egy évfolyam hallgatói például diszkrét [sokaság](#), egy borászat által termelt éves bormennyiség viszont már nem.

[Megnézem a kapcsolódó epizódot](#)

Hogyha egy [sokaság](#) nem diszkrét, akkor az folytonos.

Egy évfolyam hallgatói például diszkrét [sokaság](#), egy borászat által termelt éves bormennyiség viszont már folytonos [sokaság](#).

[Megnézem a kapcsolódó epizódot](#)

Az időpontra vonatkozó sokaságokat álló sokaságnak nevezzük.

Például egy város lakosainak száma január elsején egy álló [sokaság](#).

[Megnézem a kapcsolódó epizódot](#)

Az időtartamra vonatkozó sokaságokat mozgó sokaságnak nevezzük.

Például egy városban január elsején született lakosok száma mozgó [sokaság](#).

[Megnézem a kapcsolódó epizódot](#)

A [viszonyszámok](#) jele V és kiszámolásának módja meglehetősen semmitmondó:

$$V = \frac{A}{B}$$

A képletben A és B bármi lehet, de a képlet mégis roppant fontos.

[Megnézem a kapcsolódó epizódot](#)

Ha több viszonzszámunk van, fölmerülhet az igény ezek átlagolására.

Az egyik ilyen lehetőség a számtani átlag.

$$\bar{V} = \frac{V_1 \cdot B_1 + V_2 \cdot B_2}{B_1 + B_2} = \frac{A_1 + A_2}{B_1 + B_2}$$

[Megnézem a kapcsolódó epizódot](#)

Ha több viszonyszámunk van, fölmerülhet az igény ezek átlagolására.

Az egyik ilyen lehetőség a harmonikus átlag.

$$\bar{V} = \frac{A_1 + A_2}{\frac{A_1}{V_1} + \frac{A_2}{V_2}} = \frac{A_1 + A_2}{B_1 + B_2}$$

[Megnézem a kapcsolódó epizódot](#)

A dinamikus [viszonyszámok idősorok](#) adataiból számított hányadosok.

[Megnézem a kapcsolódó epizódot](#)

A megoszlási viszonyszám egy [sokaság](#) valamely részének az egészhez viszonyított arányát írja le.

[Megnézem a kapcsolódó epizódot](#)

Az intenzitási viszonyszám két, egymással valamilyen kapcsolatban álló [sokaság](#) mennyiségeinek hányadosa.

[Megnézem a kapcsolódó epizódot](#)

A tartam[idősorok](#) egy vizsgált időtartamra vonatkozó megfigyeléseket tartalmaznak.

Például egy év baleseteinek a számát, egy hónapban eladott fogkrémek számát, stb. Ilyenkor az adatok összeadása értelmes eredményt ad.

[Megnézem a kapcsolódó epizódot](#)

Az állapot[idősorok](#) egy vizsgált időtartam egy adott pillanatára vonatkozó megfigyeléseket tartalmazzák, például az ország lakosságának számát egy adott év adott pillanatában, vagy a raktáron lévő fogkrémkészletet egy adott hónap adott pillanatában, stb. és ilyenkor az adatok összeadásával nem kapunk értelmezhető eredményt.

[Megnézem a kapcsolódó epizódot](#)

Egy speciális átlag, például ha négy hónap adataiból számoljuk ki az átlagot, viszont csak három hónapos időtartamra.

Az állapotidősort mindig kronologikus átlagot számolunk:

$$\bar{y}_k = \frac{\frac{y_1}{2} + y_2 + y_3 + \dots + \frac{y_n}{2}}{n-1}$$

[Megnézem a kapcsolódó epizódot](#)

A statisztikai táblákat három fő csoportba tudjuk sorolni.

A legegyszerűbb típust egyszerű táblának nevezzük.

Ezek tulajdonképpen egymás mellett szerepeltetett valamilyen adatok.

A táblában nincsen sem "Összesen" sor, sem pedig "Összesen" oszlop.

Például:

Ország	GDP/fő (USD)	Gépjárművek (db/1000 fő)
Ausztria	50 380	550
Belgium	46 237	503
Hollandia	52 646	481
Svájc	82 484	539

[Megnézem a kapcsolódó epizódot](#)

A statisztikai táblákat három fő csoportba tudjuk sorolni.

Az egyik típus az úgynevezett csoportosító tábla, aminek lényege, hogy az adatokat valamelyik [ismérv](#) szerint tudjuk összesíteni.

Például:

Ország	Népesség (millió)	Sertések száma (millió)
Ausztria	8,9	2,8
Belgium	11,5	6,2
Hollandia	17,4	11,9
Svájc	8,6	1,4
Összesen:	46,4	22,3

[Megnézem a kapcsolódó epizódot](#)

A statisztikai táblákat három fő csoportba tudjuk sorolni.

Az egyik típus az úgynevezett [kombinációs tábla](#) vagy más néven [kontingencia tábla](#), amely esetében mindegyik [ismérv](#) szerint tudjuk az adatokat összesíteni.

Például:

	Nő	Férfi	Össz,
Vezető	7	18	25
Közép-vezető	11	23	34
Beosztott	756	185	941
Össz.	774	226	1000

[Megnézem a kapcsolódó epizódot](#)

Valszám alapok, kombinatorika

Eseményeknek nevezzük a valószínűségi kísérlet során bekövetkező lehetséges kimeneteket.

Megkülönböztetünk elemi eseményeket, ilyen például, hogy egy dobókockával 1-est dobunk. Vannak azonban olyan események is amik több elemi eseményből épülnek fel, ilyen például az, hogy párosat dobunk.

Az eseményeket az ABC nagybetűivel jelöljük.

[Megnézem a kapcsolódó epizódot](#)

A valószínűség kiszámításának klasszikus modelljét akkor alkalmazhatjuk, ha egy kísérletnek véges sok kimenetele van és ezek valószínűsége egyenlő. Ekkor az [esemény](#) valószínűségét úgy kaphatjuk meg, hogy megszámloljuk hány elemi eseményből áll és ezt elosztjuk az összes [elemi esemény](#) számával.

[Megnézem a kapcsolódó epizódot](#)

Az A és B eseményt egymástól függetlennek nevezzük, ha teljesül rájuk, hogy

$$P(A \cap B) = P(A) \cdot P(B)$$

[Megnézem a kapcsolódó epizódot](#)

Az A és B eseményt kizárónak nevezünk, ha

$$A \cap B = \emptyset$$

[Megnézem a kapcsolódó epizódot](#)

Az A [esemény valószínűsége](#), ha tudjuk, hogy a B [esemény](#) biztosan bekövetkezik:

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

[Megnézem a kapcsolódó epizódot](#)

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cap B) = P(A) + P(B) - P(A \cup B)$$

$$P(A \setminus B) = P(A) - P(A \cap B)$$

$$P(\bar{A}) = 1 - P(A)$$

[Megnézem a kapcsolódó epizódot](#)

Teljes valószínűség tétele és Bayes tétel

Ha B_1, B_2 és így tovább B_n teljes eseményrendszer, valamint A tetszőleges [esemény](#), akkor

$$P(A) = P(A | B_1)P(B_1) + P(A | B_2)P(B_2) + \dots + P(A | B_n)P(B_n)$$

[Megnézem a kapcsolódó epizódot](#)

A Bayes tételt akkor használjuk, ha egy korábban bekövetkezett (B_k) [esemény](#) valószínűségét akarjuk kiszámolni egy később bekövetkezett (A) tükrében.

Ha B_1, B_2 és így tovább B_n teljes eseményrendszer, valamint A tetszőleges [esemény](#), akkor bármely B_k eseményre

$$P(B_k | A) = \frac{P(A|B_k)P(B_k)}{P(A|B_1)P(B_1)+P(A|B_2)P(B_2)+\dots+P(A|B_n)P(B_n)}$$

[Megnézem a kapcsolódó epizódot](#)

Mintavétel, binomiális és hipergeometriai eloszlás

Ezt a képletet hívjuk [binomiális](#) eloszlásnak:

$$P = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}$$

ahol n a kísérletek száma,

k a sikeres kísérletek száma,

p pedig a sikeres kísérlet valószínűsége.

[Megnézem a kapcsolódó epizódot](#)

Visszatevéses mintavételről beszélünk, ha egy p valószínűségű elem többszöri kihúzásának esélyét vizsgáljuk úgy, hogy ha kihúzunk egy ilyen elemet, akkor ezt követően azt visszarakjuk.

Például ha azt vizsgáljuk, hogy egy kosárban van 8 piros és 5 kék golyó, és mennyi a valószínűsége, hogy háromszor húzva két piros és egy kék golyót húznánk úgy, hogy a kihúzott golyókat mindig visszatesszük, akkor az egy visszatevéses [mintavétel](#).

A visszatevéses mintavételhez kapcsolódó [eloszlás](#) a [binomiális eloszlás](#).

[Megnézem a kapcsolódó epizódot](#)

A visszatevés nélküli [mintavétel](#) tipikus példája, hogy van egy doboz, benne N darab elem. Közülük K darab valamilyen tulajdonságú, az egyszerűség kedvéért hívjuk selejtesnek. Mondjuk sárga vagy szép vagy ronda. Kihúzunk n darab elemet, és ez a képlet meg fogja nekünk mondani, hogy mekkora az esélye, hogy közülük k darab a vizsgált tulajdonságú:

$$P(X = k) = \frac{\binom{K}{k} \cdot \binom{N-K}{n-k}}{\binom{N}{n}}$$

De vannak olyan esetek, amikor a visszatevés nélküli mintavételnél másik képletet kell használnunk. Ezt a másik képletet [binomiális](#) eloszlásnak nevezzük, és olyankor használjuk, amikor a selejtek száma helyett csak a selejtek arányát ismerjük.

Ez a [binomiális eloszlás](#) képlete:

$$P = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}$$

ahol n a kísérletek száma,

k a sikeres kísérletek száma,

p pedig a sikeres kísérlet valószínűsége.

És, hogy mi alapján döntjük el, hogy a két képlet közül melyiket kell használni? A dolog nagyon logikus, nézd meg a kapcsolódó epizódot és minden világos lesz.

[Megnézem a kapcsolódó epizódot](#)

A [hipergeometriai eloszlás](#) a visszatevés nélküli mintavételhez kapcsolódó [eloszlás](#), képlete pedig:

$$P(X = k) = \frac{\binom{K}{k} \cdot \binom{N-K}{n-k}}{\binom{N}{n}}$$

[Megnézem a kapcsolódó epizódot](#)

Valószínűségi változók

Folytonosnak nevezzük azokat a valószínűségi változókat, amik folytonos mennyiségeket mérnek, ilyen például az idő, a távolság. Ebben az esetben az [eloszlás](#) függvény is mindig folytonos függvény lesz.

[Megnézem a kapcsolódó epizódot](#)

Diszkrétnek nevezzük azokat a valószínűségi változókat, amik megszámlálhatóan sok értéket vesznek fel. Ez azt jelenti, hogy vagy véges sokat, vagy végtelent, de úgy, hogy fel tudjuk sorolni az értékeit.

[Megnézem a kapcsolódó epizódot](#)

Az X [valószínűségi változó](#) eloszlásfüggvénye:

$$F(x) = P(X < x)$$

Ha az X [valószínűségi változó](#) diszkrét és értékei $X = a, X = b, X = c$ meg ilyenek, akkor az [eloszlásfüggvény](#) mindig egy lépcsőzetes függvény, ami minden számnál pontosan akkorát ugrik, mint az adott szám valószínűsége, amíg el nem érjük az 1-et.

$$F(x) = \begin{cases} 0 & \text{ha } x \leq a \\ P(X = a) & \text{ha } a < x \leq b \\ P(X = a) + P(X = b) & \text{ha } b < x \leq c \\ \dots \\ 1 \end{cases}$$

Ha az X [valószínűségi változó](#) folytonos, akkor az a és b számok között bármilyen valós értéket fölvehet. Ilyenkor az [eloszlásfüggvény](#) is folytonos, ami a -ig nullát vesz föl, a és b közt növekszik és b után végig egyet vesz föl. Vagyis ahol az X [valószínűségi változó](#) működik, ott a függvény életre kel, előtte és utána pedig hibernált állapotban van.

[Megnézem a kapcsolódó epizódot](#)

A [sűrűségfüggvény](#) úgy működik, hogy a valószínűségeket a görbe alatti területek adják meg. Az [eloszlásfüggvény](#) jele $F(x)$ volt, a [sűrűségfüggvény](#) jele $f(x)$. Az $a < X < b$ valószínűség éppen a görbe alatti terület a -tól b -ig.

$$P(a < X < b) = \int_a^b f(x) dx$$

Ha az $X < a$ valószínűséget szeretnénk kiszámolni:

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

Ha a $b < X$ valószínűséget:

$$P(b < X) = \int_b^{+\infty} f(x) dx$$

Ha ezt a három területet összeadjuk, akkor éppen a teljes görbe alatti területet kapjuk, ami a 100%-ot jelenti, így hát ez a terület éppen 1.

A [sűrűségfüggvény](#) tulajdonságai:

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

nem negatív

[Megnézem a kapcsolódó epizódot](#)

$$1. \lim_{-\infty} F(x) = 0$$

$$2. \lim_{\infty} F(x) = 1$$

3. monoton nő

4. balról folytonos

[Megnézem a kapcsolódó epizódot](#)

$$1. \int_{-\infty}^{\infty} f(x) dx = 1$$

2. nem negatív

[Megnézem a kapcsolódó epizódot](#)

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

$$P(b < X) = 1 - F(b) = \int_b^{+\infty} f(x) dx$$

$$P(a < X < b) = F(b) - F(a) = \int_a^b f(x) dx$$

[Megnézem a kapcsolódó epizódot](#)

Az X [valószínűségi változó](#) $F(x)$ eloszlásfüggvényéből úgy kapjuk meg az $f(x)$ sűrűségfüggvényét, hogy az $F(x)$ eloszlásfüggvényt deriváljuk, azaz:

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

Ha az X [valószínűségi változó](#) $f(x)$ sűrűségi függvényét ismerjük, és meg akarjuk adni az $F(x)$ eloszlásfüggvényét, akkor azt pedig így tehetjük:

$$F(x) = \int_{-\infty}^x f(t) dt$$

[Megnézem a kapcsolódó epizódot](#)

Várható érték és szórás

A [várható érték](#) jele $E(X)$.

Diszkrét esetben úgy kell kiszámolni, hogy

$$E(X) = \sum X_i P(X_i)$$

[Megnézem a kapcsolódó epizódot](#)

A szórás azt mutatja meg, hogy a [várható érték](#) körül milyen nagy ingadozásra számíthatunk.

Jele: $D(X)$

Kiszámításának módja diszkrét esetben:

$$D(X) = \sqrt{E(X^2) - E^2(X)}$$

[Megnézem a kapcsolódó epizódot](#)

Folytonos [valószínűségi változók](#) esetén a [várható érték](#):

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

[Megnézem a kapcsolódó epizódot](#)

Folytonos [valószínűségi változó](#) esetén a szórást ugyanúgy kell számolni, mint diszkrét [valószínűségi változó](#) esetén:

$$D(X) = \sqrt{E(X^2) - E^2(X)}$$

[Megnézem a kapcsolódó epizódot](#)

Nevezetes eloszlások

A [hipergeometriai eloszlás](#) egy diszkrét [eloszlás](#).

Ismert, hogy mennyi az összes elem és az összes selejt, vagyis N , K és n .

$$P(X = k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$$

A [hipergeometriai eloszlás](#) várható értéke:

$$E(X) = n \frac{K}{N}$$

A [hipergeometriai eloszlás](#) szórása:

$$D(X) = \sqrt{n \frac{K}{N} \left(1 - \frac{K}{N}\right) \frac{N-n}{N-1}}$$

[Megnézem a kapcsolódó epizódot](#)

A [binomiális eloszlás](#) egy diszkrét [eloszlás](#).

Csak valami %-os izé ismert, a [várható érték](#), az átlag, az arány, a valószínűség, továbbá X korlátos diszkrét [valószínűségi változó](#).

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

A [binomiális eloszlás](#) várható értéke:

$$E(X) = np$$

A [binomiális eloszlás](#) szórása:

$$D(X) = \sqrt{np(1-p)}$$

[Megnézem a kapcsolódó epizódot](#)

A [Poisson eloszlás](#) egy diszkrét [eloszlás](#), ahol előre ismert a [várható érték](#), és a [valószínűségi változó](#) nem korlátos, vagyis tetszőlegesen bármilyen nagy érték is lehet.

Például valamilyen anyagban a hibák száma, vagy egy adott idő alatt bekövetkező események száma. A [Poisson](#) eloszlásos feladatokban általában valamilyen százalék vagy arány vagy [várható érték](#) vagy átlag vagy valószínűség van megadva. Mondjuk egy könyvben az oldalak 80%-ában nincs hiba, vagy az 20 méter hosszú ruhaszövetek harmadában nincs hiba, vagy egy üzletben óránként várhatóan 13 vevő érkezik, vagy egy bankban percenként átlag 24 tranzakció történik, vagy 0,2 a valószínűsége, hogy 10 perc alatt nem érkezik segélyhívás. Ezek mind Poisson eloszlások, ahol az X nem korlátos diszkrét [valószínűségi változó](#).

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

A [Poisson eloszlás](#) várható értéke:

$$E(X) = \lambda$$

A [Poisson eloszlás](#) szórása:

$$D(X) = \sqrt{\lambda}$$

[Megnézem a kapcsolódó epizódot](#)

Az [exponenciális eloszlás](#) egy folytonos [eloszlás](#).

Eloszlásfüggvénye:

$$F(x) = \begin{cases} 0 & \text{ha } x \leq 0 \\ 1 - e^{-\lambda x} & \text{ha } 0 < x \end{cases}$$

Sűrűségfüggvénye:

$$f(x) = \begin{cases} 0 & \text{ha } x \leq 0 \\ \lambda e^{-\lambda x} & \text{ha } 0 < x \end{cases}$$

Az [exponenciális eloszlás](#) várható értéke:

$$E(X) = \frac{1}{\lambda}$$

Az [exponenciális eloszlás](#) szórása:

$$D(X) = \frac{1}{\lambda}$$

[Megnézem a kapcsolódó epizódot](#)

Az [egyenletes eloszlás](#) egy folytonos [eloszlás](#).

Eloszlásfüggvénye:

$$F(x) = \begin{cases} 0 & \text{ha } x \leq A \\ \frac{x-A}{B-A} & \text{ha } A < x \leq B \\ 1 & \text{ha } B < x \end{cases}$$

Sűrűségfüggvénye:

$$f(x) = \begin{cases} \frac{1}{B-A} & \text{ha } A < x \leq B \\ 0 & \text{különben} \end{cases}$$

Az [egyenletes eloszlás](#) várható értéke:

$$E(X) = \frac{A+B}{2}$$

Az [egyenletes eloszlás](#) szórása:

$$D(X) = \frac{B-A}{\sqrt{12}}$$

[Megnézem a kapcsolódó epizódot](#)

A [normális eloszlás](#) egy folytonos [eloszlás](#).

Eloszlásfüggvénye:

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

Sűrűségfüggvénye:

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

A [normális eloszlás](#) várható értéke:

$$E(X) = \mu$$

A [normális eloszlás](#) szórása:

$$D(X) = \sigma$$

[Megnézem a kapcsolódó epizódot](#)

Egy ismerv szerinti elemzés

Az [alakmutatók](#) arról szólnak, hogy az [eloszlás](#) mennyire asszimmetrikus.

Az egyik legegyszerűbb és leggyakrabban használt [alakmutatók](#), az úgynevezett Pearson-féle mérőszámok:

$$P = 3 \frac{\bar{X} - Me}{\sigma} \quad A = \frac{\bar{X} - Mo}{\sigma}$$

A negatív értékek jobb oldali asszimetriát jelentenek. A pozitív értékek esetén pedig bal oldali asszimetria van.

A Pearson-féle P és A mutatók általában -1 és 1 között tartózkodnak és csak extrém esetekben vesznek föl 1-nél nagyobb vagy -1-nél kisebb értéket.

[Megnézem a kapcsolódó epizódot](#)

Az [alakmutatók](#) arról szólnak, hogy az [eloszlás](#) mennyire asszimmetrikus.

Az egyik legegyszerűbb és leggyakrabban használt [alakmutatók](#), az úgynevezett Pearson-féle mérőszámok mellett az F-mutatók:

$$F_{0,25} = \frac{(Q_3 - Me) - (Me - Q_1)}{(Q_3 - Me) + (Me - Q_1)}$$

$$F_{0,1} = \frac{(D_9 - Me) - (Me - D_1)}{(D_9 - Me) + (Me - D_1)}$$

ahol D_1 az első, D_9 pedig a kilencedik decilist jelenti.

A negatív értékek jobb oldali asszimetriát jelentenek. A pozitív értékek esetén pedig bal oldali asszimetria van.

Az F mutató csak -1 és 1 között lehet.

[Megnézem a kapcsolódó epizódot](#)

A csúcosság azt jelenti, hogy az [eloszlás](#) görbéje mennyire csúcsosodik ki.

A csúcosság mérésére a következő mutató van forgalomban:

$$\alpha_4 = \frac{M_4(\bar{X})}{\sigma^4} - 3$$

Itt $M_4(\bar{X})$ az úgynevezett negyedik momentum, és így számolható ki:

$$M_4(\bar{X}) = \frac{\sum (\bar{X} - x_i)^4}{N}$$

[Megnézem a kapcsolódó epizódot](#)

A Herfindahl-index egy eszköz a [koncentráció](#) vizsgálatára.

$$HI = \sum Z_i^2$$

A Herfindahl-index $1/N$ és 1 között vesz fel értékeket és minél közelebb van az 1-hez, annál nagyobb a [koncentráció](#).

[Megnézem a kapcsolódó epizódot](#)

A Lorenz-görbe egy eszköz a [koncentráció](#) vizsgálatára.

A Lorenz-görbe az úgynevezett koncentrációs területtel szemlélteti a [koncentráció](#) mértékét.

Minél nagyobb ez a terület, a [koncentráció](#) annál erősebb.

Olyankor pedig, amikor a [Lorenz görbe](#) egybeesik a négyzet átlójával, a [koncentráció](#) nulla.

[Megnézem a kapcsolódó epizódot](#)

A tartam[idősorok](#) egy vizsgált időtartamra vonatkozó megfigyeléseket tartalmaznak.

Például egy év baleseteinek a számát, egy hónapban eladott fogkrémek számát, stb. Ilyenkor az adatok összeadása értelmes eredményt ad.

[Megnézem a kapcsolódó epizódot](#)

Az állapot[idősorok](#) egy vizsgált időtartam egy adott pillanatára vonatkozó megfigyeléseket tartalmazzák, például az ország lakosságának számát egy adott év adott pillanatában, vagy a raktáron lévő fogkrémkészletet egy adott hónap adott pillanatában, stb. és ilyenkor az adatok összeadásával nem kapunk értelmezhető eredményt.

[Megnézem a kapcsolódó epizódot](#)

Egy speciális átlag, például ha négy hónap adataiból számoljuk ki az átlagot, viszont csak három hónapos időtartamra.

Az állapotidősornál mindig kronologikus átlagot számolunk:

$$\bar{y}_k = \frac{\frac{y_1}{2} + y_2 + y_3 + \dots + \frac{y_n}{2}}{n-1}$$

[Megnézem a kapcsolódó epizódot](#)

Két ismerv szerinti elemzés

Egy sokaságot egyszerre több [ismerv](#) szerint is vizsgálhatunk.

A következő ismérveket különböztetjük meg:

- minőségi [ismerv](#) (pl. férfi vagy nő)
- területi [ismerv](#) (pl. városban vagy faluban lakik)
- [mennyiségi ismerv](#) (pl. milyen magasak)

[Megnézem a kapcsolódó epizódot](#)

Ha mindkét [ismerv](#) minőségi (vagy területi), akkor asszociációs kapcsolatról beszélünk.

Ilyen például egy cég alkalmazottjainak megoszlása neme és beosztása szerint.

Minőségi				
Minőségi		Nő	Férfi	Össz.
	Vezető	7	18	25
	Közép-vezető	11	23	34
	Beosztott	756	185	941
	Össz.	774	226	1000

Az így létrejövő táblát kombinációs táblának nevezzük. Átlagot, szórást és egyéb mutatókat egyik [ismerv](#) szerint sem tudunk számolni.

[Megnézem a kapcsolódó epizódot](#)

Ha az egyik [ismerv](#) minőségi (vagy területi), a másik mennyiségi, akkor vegyes kapcsolatról beszélünk.

Ilyen például egy város szállodáinak megoszlása az éjszakák ára és a szállodák besorolása alapján.

Minőségi					
Mennyiségi	Árak (EUR/fő/éj)	Szálloda típusa			Össz.
		**	***	****	
	0-50	37	8	1	46
	51-100	15	40	3	58
	101-150	10	33	12	55
	151-200	4	22	15	41
	Össz.	66	103	31	200

Átlagot, szórást és egyéb mutatókat csak az egyik [ismerv](#), az árak szerint tudunk számolni.

[Megnézem a kapcsolódó epizódot](#)

Ha mindkét [ismérv](#) mennyiségi, akkor korrelációs kapcsolatról beszélünk.

Ilyen például Európa 4 országának egy főre jutó GDP-je és az ezer főre jutó gépjárművek számának megoszlása.

	Mennyiségi	Mennyiségi
Ország	GDP/fő (USD)	Gépjárművek (db/1000 fő)
Ausztria	50 380	550
Belgium	46 237	503
Hollandia	52 646	481
Svájc	82 484	539

A táblázatban mindkét [ismérv](#) szerint tudunk átlagot, szórást és egyéb mutatókat számolni.

[Megnézem a kapcsolódó epizódot](#)

Ha mindkét ismerv sorrendi, akkor rangkorrelációs kapcsolatról beszélünk.

Ilyen például ha két társadalmi csoportot kérdezzünk meg, hogy 1-től 10-ig rangsorolják az alábbi országokat, az alapján, hogy mennyire szívesen nyaralnának ott.

Ország	Egyik csoport	Másik csoport
Ausztria	10	2
Belgium	9	6
Csehország	4	7
Franciaország	3	3
Görögország	1	8
Hollandia	7	5
Lengyelország	2	9
Magyarország	6	10
Németország	5	4
Svájc	8	1

[Megnézem a kapcsolódó epizódot](#)

Két [ismérv](#) akkor független, ha minden feltételes megoszlás egyforma és megegyezik a feltétel nélküli megoszlással.

[Megnézem a kapcsolódó epizódot](#)

Két [ismérv](#) között függvényszerű kapcsolat van, ha nem minden feltételes megoszlás egyforma, de minden feltételes [eloszlás](#) szórása nulla.

Függvényszerű kapcsolatnál az egyik [ismérv](#) ismeretében a másik egyértelműen kitalálható.

A két [ismérv](#) kapcsolata akkor függvényszerű, ha nem minden feltételes megoszlás egyforma, de a feltételes megoszlások szórása nulla.

[Megnézem a kapcsolódó epizódot](#)

Ha a két [ismérv](#) közötti kapcsolat nem független és nem is függvényszerű, akkor sztochasztikus kapcsolatról beszélünk. Kicsit összefüggnek ugyan az adatok, de olyan nagyon azért nem.

A két [ismérv](#) kapcsolata akkor sztochasztikus, ha nem minden feltételes megoszlás egyforma de a feltételes megoszlások szórása nem mind nulla.

[Megnézem a kapcsolódó epizódot](#)

A Cramer-féle asszociációs együttható arra való, hogy amikor mindkét [ismérv](#) minőségi, rávilágítson a két [ismérv](#) közötti kapcsolat szorosságára.

$$C = \sqrt{\frac{\chi^2}{N \cdot \min\{(r-1); (c-1)\}}}$$

Itt N = az összes elem, r = a táblázat sorainak száma és c = a táblázat oszlopainak száma, továbbá

$$\chi^2 = \sum \frac{(f_{ij} - f_{ij}^*)^2}{f_{ij}^*}$$

A Cramer-mutató függvényszerű kapcsolat esetén 1, független esetén pedig 0.

[Megnézem a kapcsolódó epizódot](#)

A [kombinációs tábla](#) általános sémája:

	C_1	C_2	$\dots$	C_j		Össz.
R_1	f_{11}	f_{12}	$\dots$	f_{1j}		$f_{1\bullet}$
R_2	f_{21}	f_{22}	$\dots$	f_{2j}		$f_{2\bullet}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
R_i	f_{i1}	f_{i2}	$\dots$	f_{ij}		$f_{i\bullet}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
Össz.	$f_{\bullet 1}$	$f_{\bullet 2}$	$\dots$	$f_{\bullet j}$		N

Az első oszlop elemei, amint látjuk f_{11} aztán f_{21} és így tovább az általános tag f_{1i} , ami közös bennük az az, hogy a második indexe mindegyiknek 1-es.

Az oszlop alján összegezzük őket, az összeg $f_{\bullet 1}$ ami azt jelenti, hogy ez azoknak az elemeknek az összege, ahol a második index 1, az első index pedig tökmindegy, hogy mi, ezt hivatott jelezni a $\bullet$ jel.

Aztán a második oszlopban pontosan ugyanez a helyzet, az oszlopban lévő elemek f_{12} alatta f_{22} és így tovább, összegük pedig $f_{\bullet 2}$.

Ugyanez megy a sorokra is, az első sor elemei f_{11} aztán f_{12} és így tovább, itt az elemek első indexe egyezik meg, mindegyiknek 1-es, összegüket pedig úgy jelöljük, hogy $f_{1\bullet}$.

[Megnézem a kapcsolódó epizódot](#)

A Csuprov-féle mutató segítségével két [ismérv](#) közötti kapcsolatot vizsgálhatjuk.

A Csuprov-féle mutató:

$$\Gamma = \sqrt{\frac{\chi^2}{N \cdot \sqrt{r-1} \cdot \sqrt{c-1}}}$$

[Megnézem a kapcsolódó epizódot](#)

Ha azt vizsgáljuk, hogy az egyes értékek mennyire térnek el a részátlagoktól, akkor belső szórást számolunk.

Jele: σ_B

A rész-szórásokból úgy lesz belső szórás, hogy súlyozzuk őket a rész-sokaságok szórásával.

A belső szórást kiszámolhatjuk a rész-szórások nélkül is.

[Megnézem a kapcsolódó epizódot](#)

Ha a részátlagoknak nézzük a főátlagtól való eltérését, az a külső szórás.

Jele: σ_K

[Megnézem a kapcsolódó epizódot](#)

Ha az egyes értékeknek nézzük a főátlagtól való eltérést, az a teljes szórás.

A teljes szórás az egész [sokaság](#) szórását jelenti, vagyis ha nem bontjuk fel a sokaságot rész-sokaságokra.

Jele : σ

A háromféle szórásra mindig teljesül, hogy

$$\sigma^2 = \sigma_B^2 + \sigma_K^2$$

[Megnézem a kapcsolódó epizódot](#)

A belső eltérés-négyzetösszeg a belső szórás gyök alatti részének számlálója.

Jele: SSB

[Megnézem a kapcsolódó epizódot](#)

A külső eltérés-négyzetösszeg a külső szórás gyök alatti részének számlálója.

Jele: SSK

[Megnézem a kapcsolódó epizódot](#)

A teljes eltérés-négyszetösszeg a teljes szórás gyök alatti részének számlálója.

Jele: SST

$$SST = SSB + SSK$$

[Megnézem a kapcsolódó epizódot](#)

A PRE egy rövidítés, Proportional Reduction Errors, ami relatív hibacsökkenésnek fordítható.

A módszer lényege, hogy a PRE érték kiszámolásával megállapítható, az egyik [ismérv](#) ismerete hány százalékkal csökkenti a másik [ismérv](#) nagyságával kapcsolatos bizonytalanságot.

Ha $PRE = 0$, az azt jelenti, hogy ez a bizonytalanság egyáltalán nem csökken. Ebben az esetben a két [ismérv](#) egymástól független.

Ha $PRE = 1$, akkor a bizonytalanság 100%-kal csökken. Ilyenkor a két [ismérv](#) között függvényszerű kapcsolat van.

Ha pedig PRE értéke valahol nulla és egy között van, akkor a kapcsolat nem független és nem is függvényszerű, tehát sztochasztikus.

$$PRE = H^2 = \frac{\sigma^2 - \sigma_B^2}{\sigma^2} = \frac{\sigma_K^2}{\sigma^2}$$

vagy

$$PRE = H^2 = \frac{SST - SSB}{SST} = \frac{SSK}{SST}$$

[Megnézem a kapcsolódó epizódot](#)

Ha két [ismérv](#) között [korrelációs kapcsolat](#) van, akkor a két [ismérv](#) közötti kapcsolat szorosságát a lineáris [korrelációs együttható](#) írja le:

$$r = \frac{\sum dX \cdot dY}{\sqrt{\sum d^2 X \cdot \sum d^2 Y}}$$

A lineáris [korrelációs együttható](#) azt méri, hogy X és Y között milyen szoros lineáris kapcsolat van. Értéke mindig $-1 \geq r \geq 1$.

Ha $r = \pm 1$, akkor X és Y között függvényszerű lineáris kapcsolat van, ha $r = 0$, akkor nincs lineáris kapcsolat. De attól, hogy nincs lineáris kapcsolat, másfajta kapcsolat még lehet, tehát $r = 0$ esetén X és Y nem biztos, hogy független.

[Megnézem a kapcsolódó epizódot](#)

A determinációs együttható a lineáris [korrelációs együttható](#) négyzete, azaz r^2 .

A determinációs együttható pontosan úgy értelmezhető, mint a PRE mutató a vegyes kapcsolatnál.

[Megnézem a kapcsolódó epizódot](#)

Ha pl. egy verseny eredményét ketten is megtippelik, és el kell döntenünk melyikük találta el jobban a valós eredményt...

Erre való a rang[korrelációs együttható](#):

$$\rho = 1 - \frac{6 \cdot \sum (R_X - R_Y)^2}{N(N^2 - 1)}$$

Hogyha valaki éppen eltalálja a helyes sorrendet, akkor a rang[korreláció](#) értéke 1.

Ha pedig éppen a fordított sorrendet találja el, akkor -1.

És minél inkább eltalálja valaki a valós sorrendet, a rang[korreláció](#) annál nagyobb.

[Megnézem a kapcsolódó epizódot](#)

Standardizálás

Az első szint a [főátlagok](#) összehasonlítása:

$$K = \bar{V}_0 - \bar{V}_1$$

A második szint a részhatás különbség kiszámolása:

(ilyenkor az összetételhatást vesszük standardnek)

$$K' = \bar{V}'_0 - \bar{V}'_1 = \frac{\sum B_0 \cdot V_0}{\sum B_0} - \frac{\sum B_0 \cdot V_1}{\sum B_0}$$

A harmadik szint az összetételhatás különbség kiszámolása:

$$K'' = \bar{V}''_0 - \bar{V}''_1 = \frac{\sum B_0 \cdot V_1}{\sum B_0} - \frac{\sum B_1 \cdot V_1}{\sum B_1}$$

[Megnézem a kapcsolódó epizódot](#)

Területi összehasonlítás:

$$K = \bar{V}_1 - \bar{V}_0$$

$$K' = \bar{V}'_1 - \bar{V}'_0$$

$$K'' = \bar{V}''_1 - \bar{V}''_0$$

$$K = K' + K''$$

[Megnézem a kapcsolódó epizódot](#)

Időbeli összehasonlítás:

$$I = \frac{\bar{V}_1}{\bar{V}_0}$$

$$I' = \frac{\bar{V}'_1}{\bar{V}'_0}$$

$$I'' = \frac{\bar{V}''_1}{\bar{V}''_0}$$

$$I = I' \cdot I''$$

[Megnézem a kapcsolódó epizódot](#)

Az első szint a [főátlagok](#) összehasonlítása:

$$I = \frac{\bar{V}_1}{\bar{V}_0}$$

A második szint a részhatásindex kiszámolása:

(ilyenkor az összetételhatást vesszük standardnek és mindig a tárgyidőszakét)

$$I' = \frac{\bar{V}'_1}{\bar{V}'_0} = \frac{\frac{\sum B_1 \cdot V_1}{\sum B_1}}{\frac{\sum B_1 \cdot V_0}{\sum B_1}}$$

A harmadik szint az összetételhatás-index kiszámolása:

(ilyenkor a részhatást vesszük standardnek és mindig a bázisidőszakit, de ha I és I' már megvan, akkor egyszerűbb az $I = I' \cdot I''$ alapján számolni.)

$$I'' = \frac{\bar{V}''_1}{\bar{V}''_0} = \frac{\frac{\sum B_1 \cdot V_0}{\sum B_1}}{\frac{\sum B_0 \cdot V_0}{\sum B_0}}$$

[Megnézem a kapcsolódó epizódot](#)

Indexek

A [volumenindex](#) a forgalom változásának mértéke, amit súlyozhatunk a bázisidőszak vagy a tárgyidőszak áraival.

[Volumenindex](#) Laspeyres /bázis időszak szerint/:

$$I_q^0 = \frac{\sum p_0 q_1}{\sum p_0 q_0}$$

[Volumenindex](#) Paasche /tárgyidőszak szerint/:

$$I_q^1 = \frac{\sum p_1 q_1}{\sum p_1 q_0}$$

[Megnézem a kapcsolódó epizódot](#)

Az [árindex](#) a szektort érintő árváltozást méri, és súlyozhatjuk a bázisidőszak vagy a tárgyidőszak volumeneivel.

[Árindex](#) Laspeyres /bázis időszak szerint/:

$$I_p^0 = \frac{\sum p_1 q_0}{\sum p_0 q_0}$$

[Árindex](#) Paasche /tárgyidőszak szerint/:

$$I_p^1 = \frac{\sum p_1 q_1}{\sum p_0 q_1}$$

[Megnézem a kapcsolódó epizódot](#)

Az indexek átlagformái:

$$I_p^0 = \frac{\sum p_0 q_0 i_p}{\sum p_0 q_0} = \frac{\sum p_1 q_0}{\sum \frac{p_1 q_0}{i_p}} = \frac{\sum v_0 i_p}{\sum v_0}$$

$$I_p^1 = \frac{\sum p_0 q_1 i_p}{\sum p_0 q_1} = \frac{\sum p_1 q_1}{\sum \frac{p_1 q_1}{i_p}} = \frac{\sum v_1}{\sum \frac{v_1}{i_p}}$$

$$I_q^0 = \frac{\sum p_0 q_0 i_q}{\sum p_0 q_0} = \frac{\sum p_0 q_1}{\sum \frac{p_0 q_1}{i_q}} = \frac{\sum v_0 i_q}{\sum v_0}$$

$$I_q^1 = \frac{\sum p_1 q_0 i_q}{\sum p_1 q_0} = \frac{\sum p_1 q_1}{\sum \frac{p_1 q_1}{i_q}} = \frac{\sum v_1}{\sum \frac{v_1}{i_q}}$$

[Megnézem a kapcsolódó epizódot](#)

A Fischer-féle [árindex](#) és [volumenindex](#):

$$I_p^F = \sqrt{I_p^0 \cdot I_p^1} \quad I_q^F = \sqrt{I_q^0 \cdot I_q^1}$$

[Megnézem a kapcsolódó epizódot](#)

Az [értékindex](#):

$$I_v = \frac{\sum p_1 q_1}{\sum p_0 q_0} = I_p^1 \cdot I_q^0 = I_q^1 \cdot I_p^0 = I_p^F \cdot I_q^F$$

[Megnézem a kapcsolódó epizódot](#)

A vásárlóerő mérésére van forgalomban az úgynevezett vásárlóerő-paritás. Angol megfelelője Purchasing Power Parity alapján rövidítése PPP.

A vásárlóerő-paritás az egyedi vásárlóerő-parításoknak a fogyasztással súlyozott átlaga.

[Megnézem a kapcsolódó epizódot](#)
