



MATEKING.HU

Képletgyűjtemény

MATEMATIKA 3 EDUTUS tantárgy

Kiadás dátuma: 2026. 09. 15.

Tartalomjegyzék

Kamatos kamat és pénzügyi számítások.....	2
Kombinatorika.....	3
Valszám alapok, klasszikus valszám.....	5
Teljes valószínűség tétele, Bayes tétel.....	6
Geometriai valószínűség, Binomiális tétel.....	7
Eloszlás, eloszlásfüggvény, sűrűségfüggvény.....	8
A binomiális eloszlás és a hipergeometriai eloszlás.....	11
Várható érték és szórás.....	13
Markov és Csebisev egyenlőtlenségek.....	14
Nevezetes diszkrét és folytonos eloszlások.....	15
Kétváltozós eloszlások.....	18
Statisztikai alapfogalmak.....	20
Becslések.....	22
Hipotézisvizsgálat.....	28
Regressziószámítás.....	33

Kamatos kamat és pénzügyi számítások

A kamatos kamat lényege, hogy beteszünk a bankba egy összeget, amit tőkének neveznek és T_0 -al jelölünk. Erre egy bizonyos időszak alatt $p\%$ -os kamatot kapunk. Eddig ezt úgy hívjuk, hogy egyszerű kamat. Attól lesz belőle kamatos kamat, hogy a kamattal megnövelt összeget újra kamatoztatjuk, és így elindul a kamatos kamat folyamata. Magának a kamatos kamatnak a képlete nagyon egyszerű, csupán néhány dologra kell figyelni, amiket részletesen be is mutatunk a kamatos kamat feladatok megoldása közben.

A T_0 összegből n darab kamatperiódus után a következő T_n összeg lesz, ha minden periódusban p -os a kamat:

$$T_n = T_0 \cdot \left(1 + \frac{p}{100}\right)^n$$

A képletben p jelenti a kamatot és n pedig a kamatperiódusok számát. Ezek azok a periódusok, amiknek a végén jóváírják a kamatot. Hogyha a kamatot havonta írják jóvá, akkor a periódusok hónapok, és n a hónapok száma. Hogyha féléves jóváírás van, akkor n a félévek száma, hogyha évente írják jóvá a kamatot akkor n az évek száma és így tovább. Vannak mindenféle bonyolult képletek, amik ezt megpróbálják kezelni, de kár foglalkozni velük, sokkal egyszerűbb megérteni a dolog lényegét és a sima kamatos kamat képletet használni. Mindig azt a kérdést tegyük föl magunknak, hogy milyen gyakorisággal írják jóvá a kamatot, hány jóváírás van és egy periódusra mekkora kamat jut.

Ha például az éves kamat 6% és évenkénti a kamatozás, akkor $p = 6$. Ha a jóváírás félévente történik, akkor a 6% -os éves kamatot is felezzük, tehát $p = 3$, vagy éppen havi jóváírás esetén az éves kamatot 12-vel kell osztani és így $p = 0,5$. Ha ezt a gondolatmenetet megértjük, a kamatos kamat képlete örülden egyszerű.

[Megnézem a kapcsolódó epizódot](#)

T összeget n darab perióduson keresztül azonos méretű pénzösszegekben törlesztünk. Minden periódusban $p\%$ -os a kamat. Az egy periódusra eső törlesztőrészlet:

$$a = T \cdot \frac{q^n \cdot (q-1)}{q^n - 1} \quad q = 1 + \frac{p}{100}$$

[Megnézem a kapcsolódó epizódot](#)

n darab perióduson keresztül azonos méretű a pénzösszegeket fizetünk be, periódusonkénti $p\%$ -os kamat mellett. Az n periódus végén összegyűlt pénzmennyiség:

$$S_n = a \cdot q \cdot \frac{q^n - 1}{q - 1} \quad q = 1 + \frac{p}{100}$$

[Megnézem a kapcsolódó epizódot](#)

Kombinatorika

Egy adott n elemű halmaz elemeinek egy ismétlés nélküli permutációján az n különböző elem egy sorba rendezését értjük.

n darab különböző elem permutációinak száma:

$$1 \cdot 2 \cdot 3 \cdot \dots \cdot n = n!$$

[Megnézem a kapcsolódó epizódot](#)

n faktoriálisán az n -nél kisebb vagy egyenlő pozitív egész számok szorzatát értjük.

$$n! = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot 3 \cdot 2 \cdot 1$$

pl.:

$$4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$$

$$5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$$

$$1! = 1$$

Továbbá definíció szerint $0! = 1$.

[Megnézem a kapcsolódó epizódot](#)

Ha n db. egymástól különböző elem közül kiválasztunk k ($k \leq n$) db.-ot úgy, hogy a kiválasztott elemek sorrendje is számít, akkor az n elem k -ad osztályú ismétlés nélküli variációját kapjuk.

n darab különböző elemből kiválasztott k darab elem variációinak száma:

$$n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-k+1) = \frac{n!}{(n-k)!}$$

[Megnézem a kapcsolódó epizódot](#)

Ha n különböző elem közül kiválasztunk k ($k \leq n$) db.-ot úgy, hogy a kiválasztott elemek sorrendjére nem vagyunk tekintettel, akkor n elem k -ad osztályú ismétlés nélküli kombinációját kapjuk.

n darab különböző elem közül kiválasztott k darab elem kombinációinak száma:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

[Megnézem a kapcsolódó epizódot](#)

Ha n elem között van $k_1, k_2, \dots, k_r$ egymással megegyező, akkor az elemek egy sorba rendezését ismétléses permutációnak nevezzük.

n elem közötti $k_1, k_2, \dots, k_r$ egymással megegyező ismétléses permutációinak száma:

$$\frac{n!}{k_1! \cdot k_2! \cdot \dots \cdot k_r!}$$

[Megnézem a kapcsolódó epizódot](#)

Ha n db. egymástól különböző elem közül kiválasztunk k db.-ot úgy, hogy a kiválasztott elemek sorrendje is számít és ugyanazt az elemet többször is választhatjuk, akkor az n elem k -ad osztályú ismétléses variációját kapjuk.

Az n elem k -ad osztályú ismétléses variációk száma: n^k .

[Megnézem a kapcsolódó epizódot](#)

Ha kör alakban helyezünk el n különböző elemet és azok sorrendjét vizsgáljuk, akkor ciklikus permutációról beszélünk.

n darab különböző elem ciklikus permutációinak száma $\frac{n!}{n} = (n - 1)!$

[Megnézem a kapcsolódó epizódot](#)

Valszám alapok, klasszikus valszám

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, 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](#)

Geometriai valószínűség, Binomiális tétel

Ha egy [esemény](#) előfordulását geometriai alakzat (vonal, síkidom, test) mértékével jellemezzük, akkor geometriai valószínűségről beszélünk.

Ilyenkor a szokásos $P = \frac{\text{kedvező}}{\text{összes}}$ lehet mondjuk $P = \frac{T_{\text{kedvező}}}{T_{\text{összes}}}$

[Megnézem a kapcsolódó epizódot](#)

$$(a + b)^n = \sum_{k=0}^n \binom{n}{k} a^{n-k} b^k$$

[Megnézem a kapcsolódó epizódot](#)

[Binomiális](#) tétel:

$$(a + b)^n = \sum_{k=0}^n \binom{n}{k} a^{n-k} b^k = \binom{n}{0} a^n + \binom{n}{1} a^{n-1} b + \binom{n}{2} a^{n-2} b^2 + \dots + \binom{n}{n} b^n$$

[Megnézem a kapcsolódó epizódot](#)

Eloszlás, eloszlásfüggvény, sűrűségfüggvény

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.

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

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](#)

A binomiális eloszlás és a 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 kihúzzunk 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. Kihúzzunk 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](#)

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](#)

Markov és Csebisev egyenlőtlenségek

A Markov-egyenlőtlenség egy nagyon egyszerű dolgot állít. Az, hogy az X [valószínűségi változó](#) sokkal nagyobb legyen a várható értéknél nem túl valószínű:

$$P(X \geq t \cdot E(X)) \leq \frac{1}{t}$$

[Megnézem a kapcsolódó epizódot](#)

A [Csebisev egyenlőtlenség](#) arról szól, hogy a várható értéktől való eltérés nem lehet túl nagy.

Ha ez az eltérés nagyobb, mint a szórás t -szerese, akkor ennek a valószínűsége kicsi:

$$P(|X - E(X)| \geq t \cdot D(X)) \leq \frac{1}{t^2}$$

Ha az eltérés kisebb, mint a szórás t -szerese, akkor ennek valószínűsége nagy:

$$P(|X - E(X)| < t \cdot D(X)) > 1 - \frac{1}{t^2}$$

[Megnézem a kapcsolódó epizódot](#)

Ha egy [esemény](#) bekövetkezésének elméleti valószínűsége p , akkor minél többször végezzük el a kísérletet, a relatív gyakoriság és az elméleti valószínűség eltérése annál kisebb lesz.

$$P\left(\left|\frac{X}{n} - p\right| < \epsilon\right) \geq 1 - \frac{p(1-p)}{n\epsilon^2} \quad P\left(\left|\frac{X}{n} - p\right| > \epsilon\right) < \frac{p(1-p)}{n\epsilon^2}$$

[Megnézem a kapcsolódó epizódot](#)

Nevezetes diszkrét és folytonos 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](#)

Kétváltozós eloszlások

X és Y együttes eloszlása egy táblázat, amelyben szerepel X és Y összes lehetséges értéke és a hozzájuk tartozó valószínűségek.

Ha a táblázat sorait összeadjuk, akkor Y peremeloszlását kapjuk. Ha az oszlopokat adjuk össze, akkor X peremeloszlását kapjuk.

[Megnézem a kapcsolódó epizódot](#)

A [korreláció](#) X és Y valószínűségi változók közötti kapcsolatot írja le.

$$COV(X, Y) = E(X \cdot Y) - E(X)E(Y)$$

$$R(X, Y) = \frac{COV(X, Y)}{D(X)D(Y)}$$

[Megnézem a kapcsolódó epizódot](#)

X és Y kétváltozós [eloszlás](#) esetén...

X peremeloszlásfüggvénye:

$$F_X(x) = P(X < x) = \lim_{y \rightarrow \infty} F(x, y)$$

Y peremeloszlásfüggvénye:

$$F_Y(y) = P(Y < y) = \lim_{x \rightarrow \infty} F(x, y)$$

[Megnézem a kapcsolódó epizódot](#)

X és Y valószínűségi változók együttes eloszlásfüggvénye:

$$F(x, y) = P(X < x, Y < y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) dv du$$

[Megnézem a kapcsolódó epizódot](#)

Az együttes [eloszlás](#) sűrűségfüggvénye:

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

ahol

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) dv du$$

Az együttes [eloszlásfüggvény](#).

[Megnézem a kapcsolódó epizódot](#)

X perem-sűrűségfüggvénye:

$$f_X(x) = F'_X(x)$$

ahol

$$F_X(x) = \int_{-\infty}^{\infty} f(u) du$$

az X perem-eloszlásfüggvénye.

Y perem-sűrűségfüggvénye:

$$f_Y(y) = F'_Y(y)$$

ahol

$$F_Y(y) = \int_{-\infty}^{\infty} f(v) dv$$

az Y perem-eloszlásfüggvénye.

[Megnézem a kapcsolódó epizódot](#)

X és Y kétváltozós [eloszlás](#) esetén...

X peremeloszlásfüggvénye:

$$F_X(x) = P(X < x) = \lim_{y \rightarrow \infty} F(x, y)$$

Y peremeloszlásfüggvénye:

$$F_Y(y) = P(Y < y) = \lim_{x \rightarrow \infty} F(x, y)$$

[Megnézem a kapcsolódó epizódot](#)

Statisztikai alapfogalmak

Az adatsor leggyakoribb értéke a [módusz](#). Hogyha például Bob matekjegyei ezek:

2, 3, 1, 4, 1, 2, 2, 3, 5, 2, 3, 2, 3, 2, 4, 3, 2, 4, 2, 4

Akkor egyszerűen meg kell számolni, hogy melyikből van a legtöbb, és az a matekjegy lesz a [módusz](#). Most 2-esből van a legtöbb, így Bob matekjegyeinek a módusza 2. A [módusz](#) jele M_o és így most $M_o=2$.

Léteznek olyan eloszlások is, amelyeknek több módusza van. Hogyha például Bob jegyei:

1, 2, 2, 3, 5, 3, 3, 4, 2

Itt 2-esből és 3-asból ugyanannyi van, mindkettőből 3 darab. Ez egy kétmódusú [eloszlás](#).

[Megnézem a kapcsolódó epizódot](#)

A [medián](#) a növekvő sorba rendezett adatsor középső értéke. Ha az adatsorban páros sok elem van, akkor nincs középső elem, ilyenkor a két középső elem átlagát vesszük.

Hogyha például Bob matekjegyei ezek:

2, 3, 1, 4, 1, 2, 2, 3, 5, 2, 3, 2, 3, 2, 4, 3, 2, 4, 2, 4

Akkor egyszerűen növekvő sorba kell rakni..

1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 4, 4, 4, 4, 5

És aztán meg kell keresni melyik a középső. Most nincsen középső, mert páros sok elem van, így ilyenkor a két középen lévőt átlagoljuk:

1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 4, 4, 4, 4, 5

Ezeknek az átlaga 2,5 vagyis a [medián](#) most 2,5. A [medián](#) jele M_e , így $M_e=2,5$

[Megnézem a kapcsolódó epizódot](#)

Az átlagot úgy kapjuk meg, hogy az összes elemet összeadjuk, és aztán elosztjuk az elemek számával.

Jele: $\bar{x}$

[Megnézem a kapcsolódó epizódot](#)

Az átlagtól való átlagos eltérés egyik legjobb mérőszáma a szórás. Hátránya, hogy egy kicsit ronda a szórás képlete. A szórást egy szigma nevű görög betűvel jelöljük.

$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}}$$

[Megnézem a kapcsolódó epizódot](#)

Az adatsor első felének a felezőpontja az alsó kvartilis.

Az alsó kvartilis jele: Q_1

[Megnézem a kapcsolódó epizódot](#)

Az adatsor második felének a felezőpontja a felső kvartilis.

A felső kvartilis jele: Q_3

[Megnézem a kapcsolódó epizódot](#)

A kvartilisek és a [medián](#) azt szemlélteti, hogyan oszlanak el az adatsorban szereplő adatok. Ezek segítségével készíthető el a doboz-ábra, vagy másnéven dobozdiagram. Szokás még sodrófa diagramnak is nevezni, és az angol elnevezést is gyakran használják, ami a box plot.

Egy sobarendezett adatsorban öt darab speciális negyedelőpontot fogunk használni. Az első az adatsor legkisebb értéke, ez a Q_0 . Aztán a következő negyedelő az alsó kvartilis, ami Q_1 utána jön a felezőpont vagyis a [medián](#), ezt Me -vel és Q_2 -vel is jelöljük, végül a felső kvartilis, ami a Q_3 . Az adatsor legnagyobb értéke pedig Q_4 . A legnagyobb és a legkisebb érték különbsége a terjedelem, míg a két kvartilis különbségét félterjedelemnek vagy más néven interkvartilisnek hívjuk. Ezekből épül föl a doboz-ábra vagy másként dobozdiagram.

Előfordulhat, hogy az adatsorban kiugró értékek is szerepelnek. A kiugró érték az, ami az alsó kvartilisnél legalább a félterjedelem másfélszeresénél kisebb, vagy pedig a felső kvartilisnél legalább a félterjedelem másfélszeresénél nagyobb. Huh, ez elég bonyolult hangzik. De valójában nagyon egyszerű, csak nézd meg kapcsolódó epizódot és kiderül.

[Megnézem a kapcsolódó epizódot](#)

A relatív szórás azt mondja meg, hogy a szórás az átlagnak hány százaléka:

$$V = \frac{\sigma}{\bar{x}}$$

[Megnézem a kapcsolódó epizódot](#)

Becslések

Olyan esetekben, amikor valamiért nem tudjuk vagy nem akarjuk a teljes sokaságot megvizsgálni, hogy meghatározzuk a fontosabb statisztikai mutatóit, becslést alkalmazunk. A becslés lényege, hogy egy minta alapján próbálunk ezekre a mutatókra következtetni.

[Megnézem a kapcsolódó epizódot](#)

A megbízhatósági szintet konfidencia szintnek nevezzük. A konfidencia szint szokásos jelölése $1 - \alpha$.

[Megnézem a kapcsolódó epizódot](#)

Az $1 - \alpha$ megbízhatósági szinthez, vagy másként konfidencia szinthez tartozó konfidencia intervallumok azok az intervallumok, amik a sokasági átlagot $1 - \alpha$ valószínűséggel tartalmazzák.

A konfidencia intervallum végpontjai:

$$\bar{x} \pm Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \text{ ahol}$$

$\bar{x}$ = a minta átlaga

n = a minta elemszáma

σ = a teljes [sokaság](#) szórása

$Z_{1-\frac{\alpha}{2}}$ pedig a [standard normális eloszlás](#) $1 - \frac{\alpha}{2}$ -höz tartozó Z értéke.

[Megnézem a kapcsolódó epizódot](#)

$$\bar{x} \pm Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \text{ ahol}$$

$\bar{x}$ = a minta átlaga

n = a minta elemszáma

σ = a teljes [sokaság](#) szórása

$Z_{1-\frac{\alpha}{2}}$ pedig a [standard normális eloszlás](#) $1 - \frac{\alpha}{2}$ -höz tartozó Z értéke.

[Megnézem a kapcsolódó epizódot](#)

A FAE minta azt jelenti, hogy a [mintavétel](#) során bármely mintaelemet azonos eséllyel választunk ki. Ilyen a visszatevéses [mintavétel](#), vagy pedig abban az esetben ha az alap [sokaság](#) elemszáma nagyon nagy, akkor a visszatevés nélküli [mintavétel](#) is.

[Megnézem a kapcsolódó epizódot](#)

$$\bar{x} \pm t_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \text{ ahol}$$

$1 - \alpha =$ konfidencia szint

$\bar{x} =$ a minta átlaga

$n =$ a minta elemszáma

$s =$ a teljes [sokaság](#) szórása, a sokasági szórás nem ismert

$t_{1-\frac{\alpha}{2}}$ pedig a [t-eloszlás](#) $1 - \frac{\alpha}{2}$ -höz tartozó értéke.

[Megnézem a kapcsolódó epizódot](#)

$$\bar{p} \pm Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \text{ ahol}$$

$1 - \alpha =$ konfidencia szint

$\bar{p} =$ a minta alapján kapott valószínűség

$n =$ a minta elemszáma

$Z_{1-\frac{\alpha}{2}}$ pedig a [standard normális eloszlás](#) $1 - \frac{\alpha}{2}$ -höz tartozó Z értéke.

[Megnézem a kapcsolódó epizódot](#)

$$\frac{(n-1) \cdot s^2}{\chi^2_{1-\frac{\alpha}{2}}(v)} < \sigma^2 < \frac{(n-1) \cdot s^2}{\chi^2_{\frac{\alpha}{2}}(v)} \text{ ahol}$$

$1 - \alpha =$ konfidencia szint

$\bar{p} =$ a minta alapján kapott valószínűség

$n =$ a minta elemszáma

$s =$ a minta szórása, a sokasági szórás nem ismert

$\chi^2(v)$ pedig a khi-négyzet [eloszlás](#) megfelelő értéke

[Megnézem a kapcsolódó epizódot](#)

Az EV-minta abban különbözik a FAE-mintától, hogy a kiválasztott mintaelemek nem függetlenek egymástól.

Ez olyankor fordulhat elő, ha a teljes [sokaság](#) mérete viszonylag kicsi a minta elemszámához képest. EV-minták esetén tehát a minta fontos jellemzőjévé válik, hogy mekkora a teljes [sokaság](#), amelynek elemszámát N jelöli.

[Megnézem a kapcsolódó epizódot](#)

$$\bar{x} \pm t_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \cdot \sqrt{1 - \frac{n}{N}} \text{ ahol}$$

$1 - \alpha =$ konfidencia szint

$\bar{x} =$ a minta átlaga

$n =$ a minta elemszáma

$N =$ a teljes [sokaság](#) elemszáma

$s =$ a minta szórása

$t_{1-\frac{\alpha}{2}}$ pedig a [t-eloszlás](#) $1 - \frac{\alpha}{2}$ -höz tartozó értéke.

[Megnézem a kapcsolódó epizódot](#)

$$\bar{p} \pm Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \cdot \sqrt{1 - \frac{n}{N}} \text{ ahol}$$

$1 - \alpha =$ konfidencia szint

$\bar{p} =$ a minta alapján kapott valószínűség

$n =$ a minta elemszáma

$N =$ a teljes [sokaság](#) elemszáma

$Z_{1-\frac{\alpha}{2}}$ pedig a [standard normális eloszlás](#) $1 - \frac{\alpha}{2}$ -höz tartozó Z értéke.

[Megnézem a kapcsolódó epizódot](#)

Ha a teljes sokaságot felosztjuk viszonylag homogén rétegekre, és a mintát is ezen a rétegek szerint vizsgáljuk, a variancia csökkenthető.

$$\hat{\bar{x}}_R \pm Z_{1-\frac{\alpha}{2}} \cdot s_{\hat{\bar{X}}_R}$$

$1 - \alpha$ = konfidencia szint

$\bar{x}$ = a minta átlaga

n = a minta elemszáma

n_j = a minta j-edik rétegének elemszáma

N = a teljes [sokaság](#) elemszáma

N_j = a teljes [sokaság](#) j-edik rétegének elemszáma

W_j = a teljes [sokaság](#) j-edik rétegének a teljes sokasághoz viszonyított aránya

s_j = a minta j-edik rétegének szórása

$$\hat{\bar{X}}_R = \sum_{j=1}^M W_j \bar{x}_j$$

$$s_{\hat{\bar{X}}_R}^2 = \sum_{j=1}^M W_j^2 \frac{s_j^2}{n_j} \left(1 - \frac{n_j}{N_j}\right)$$

[Megnézem a kapcsolódó epizódot](#)

A kétmintás becslésekre akkor van szükség, amikor két [sokaság](#) valamilyen paraméterét, leginkább az átlagát szeretnénk összehasonlítani.

A kétmintás [becslések](#) lehetnek független mintás [becslések](#) vagy páros mintás [becslések](#).

[Megnézem a kapcsolódó epizódot](#)

Ha mindkét [sokaság](#) közel normális eloszlású, akkor az [átlagok](#) különbségének becslésére ez a formula van forgalomban.

$$d \pm t_{1-\frac{\alpha}{2}} \cdot s_d \text{ ahol } d = \bar{x} - \bar{y}$$

$$s_d = s_c \cdot \sqrt{\frac{1}{n_Y} + \frac{1}{n_X}} \text{ itt } s_c^2 = \frac{(n_X-1)s_X^2 + (n_Y-1)s_Y^2}{n_X+n_Y-2}$$

$1 - \alpha$ = konfidencia szint

$\bar{x}$ = az egyik minta átlaga

$\bar{Y}$ = a másik minta átlaga

n_X = az egyik minta elemszáma

n_Y = a másik minta elemszáma

A szabadságfok $v = n_X + n_Y - 2$

[Megnézem a kapcsolódó epizódot](#)

Egy becslést torzítatlannak nevezünk, ha az egyes mintákból kapott [becslések](#) várható értéke megegyezik a becslni kívánt mennyiséggel.

Ez a tulajdonság azt jelenti, hogy a becslés során kapott értékek a becslni kívánt érték körül ingadoznak, és ez az ingadozás szimmetrikus. A torzítatlan becsléseket mindig előnyben részesítjük a torzítottakkal szemben.

[Megnézem a kapcsolódó epizódot](#)

A kérdés az, hogy ha egy sokasági jellemzőre több becslés jöhet szóba, hogyan válasszunk közülük, vagyis mikor tekintünk egy becslést jónak, kettő közül melyiket tekintjük jobbnak és kijelenthetjük-e valamelyikről, hogy a legjobb?

Két alapvető szempont alapján szoktuk a becsléseket versenyeztetni. Az egyik, a már jól ismert torzítatlanság, vagyis a becslésnek az a tulajdonsága, hogy az összes lehetséges mintán vett [becslések](#) átlaga megegyezik a becslni kívánt sokasági jellemzővel. A másik az úgynevezett minimális variancia kritérium.

A minimális variancia kritérium azt jelenti, hogy ha van két torzítatlan becslésünk, akkor a kettő közül azt tekintjük jobbnak, aminek az összes mintán vett értékeinek varianciája kisebb.

[Megnézem a kapcsolódó epizódot](#)

$$MSE(\hat{\theta}) = var(\hat{\theta}) + (E(\theta) - \theta)^2$$

Az első tag a varianciát, a második tag a várható értéktől való eltérést, vagyis a torzítottságot méri. Ha a becslés torzítatlan, $E(\hat{\theta}) = \theta$ így ez a második tag nulla. Két becslés közül azt részesítjük előnyben, amelyre MSE kisebb.

Az $E(\hat{\theta}) - \theta$ különbségre, vagyis a torzítás mértékére az angol bias szó alapján a $Bs\hat{\theta}$ jelölés van forgalomban. Használatos tehát az

$$MSE(\hat{\theta}) = var(\hat{\theta}) + Bs^2(\hat{\theta})$$

Képlet is.

[Megnézem a kapcsolódó epizódot](#)

Eddigi vizsgálódásaink egyik legfontosabb eredménye a mintaátlagot eloszlásának jellemzése. Ha a teljes [sokaság](#) átlaga μ és szórása pedig σ , akkor az ebből vett n elemű minták átlagai olyan eloszlással helyezkednek el, aminek átlaga szintén μ , a szórása pedig $\frac{\sigma}{\sqrt{n}}$.

Ezt az utóbbit a minta standard hibájának szokás nevezni. A standard hiba tehát azt mondja meg, hogy a minta [átlagok](#) mekkora szórással ingadoznak a tényleges sokasági átlag körül.

[Megnézem a kapcsolódó epizódot](#)

Mintavételi hibának azokat a hibákat nevezzük, amik kimondottan azért fordulnak elő, mert nem tudjuk, vagy nem akarjuk a teljes sokaságot vizsgálni. A mintavételi hiba tehát a [sokaság](#) eloszlásán és a mintavételi eljárásen kívül főleg a minta elemszáma határozza meg. Mivel pedig ezeket általában már a mintavételt megelőzően ismerjük, a mintavételi hibának megvan az a kellemes tulajdonsága, hogy legtöbbször előre megállapítható. Vagyis még el sem végeztük a mintavételt, de már tudjuk, hogy mekkora lesz a [mintavétel](#) során elkövetett hiba. Ez a kellemes tulajdonság lesz a kiindulópont a [becslések](#) és később a hipotézisvizsgálatok elméletének kiépítésében.

[Megnézem a kapcsolódó epizódot](#)

Hipotézisvizsgálat

Az [elfogadási tartomány](#) az a tartomány, ahová ha a próba értéke kerül, akkor a nullhipotézist elfogadjuk.

[Megnézem a kapcsolódó epizódot](#)

A [kritikus tartomány](#) az a tartomány, ahová ha a próba értéke kerül, akkor a nullhipotézist elvetjük.

[Megnézem a kapcsolódó epizódot](#)

A [szignifikanciaszint](#) a hibás döntés valószínűsége.

[Megnézem a kapcsolódó epizódot](#)

ELSŐ LÉPÉS: A HIPOTÉZIS MEGFOGALMAZÁSA

Minden [hipotézisvizsgálat](#) két egymásnak ellentmondó felvetés felírásával kezdődik. Az egyiket nullhipotézisnek nevezzük és H_0 -al jelöljük, a másikat pedig ellenhipotézisnek és jele H_1 .

MÁSODIK LÉPÉS: A PRÓBAFÜGGVÉNY KIVÁLASZTÁSA

A próbafüggvények kiválasztása magától a hipotézistől, illetve a [mintavétel](#) módjától is függ.

HARMADIK LÉPÉS: [SZIGNIFIKANCIASZINT](#) ÉS [KRITIKUS TARTOMÁNY](#)

Ha a próbafüggvény értéke az elfogadási tartományba fog esni, akkor ezt a tényt a nullhipotézist igazoló jelnek fogjuk tekinteni. Hogyha pedig a kritikus tartományba, akkor a nullhipotézist elvetjük.

NEGYEDIK LÉPÉS: [MINTAVÉTEL](#) ÉS DÖNTÉS

Ha a mintavétellel kapott eredményünk szerint a próbafüggvény az elfogadási tartományba esik, akkor a H_0 nullhipotézist tekintjük igaznak, a H_1 ellenhipotézist pedig elvetjük.

Ha viszont a próbafüggvény a minta alapján a kritikus tartományba esik, akkor a H_0 nullhipotézist vetjük el és a H_1 ellenhipotézist tekintjük igaznak.

[Megnézem a kapcsolódó epizódot](#)

A [sokaság](#) normális eloszlású, szórása σ , H_0 a [sokaság](#) átlagára vonatkozik, a minta elemszáma n .

$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

[Megnézem a kapcsolódó epizódot](#)

A [sokaság](#) normális eloszlású, szórása nem ismert, H_0 a [sokaság](#) átlagára vonatkozik, a minta elemszáma n .

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

[Megnézem a kapcsolódó epizódot](#)

A [sokaság](#) tetszőleges eloszlású, szórása nem ismert, H_0 a [sokaság](#) átlagára vonatkozik, a minta n elemű, elemszáma nagy.

$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

[Megnézem a kapcsolódó epizódot](#)

A [sokaság](#) tetszőleges eloszlású, H_0 a sokasági arányra vonatkozik, a minta n elemű, elemszáma nagy.

$$Z = \frac{P - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}}$$

[Megnézem a kapcsolódó epizódot](#)

A [sokaság](#) normális eloszlású, H_0 a sokasági szórásra vonatkozik, a minta n elemű.

$$\chi^2 = \frac{(n-1) \cdot s^2}{\sigma_0^2}$$

[Megnézem a kapcsolódó epizódot](#)

A [sokaság](#) eloszlására irányuló vizsgálat.

H_0 : mindegyik osztályköz valószínűsége egy adott eloszlásnak megfelelő érték, vagyis minden i -re az i -edik osztályköz valószínűsége a P_i érték.

Az ellenhipotézis pedig, H_1 : van olyan osztályköz, ami nem az adott eloszlásnak megfelelő P_i érték. A próbát $\chi^2_{1-\alpha}(v)$ jobb oldali kritikus értékkel végezzük el, a nullhipotézist az ennél kisebb, az ellenhipotézist az ennél nagyobb értékek igazolják. A minta elemszáma n .

$$\chi^2(v) = \sum_{i=1}^k \frac{(f_i - nP_i)^2}{nP_i}$$

ahol a v szabadságfok: $v = k - b - 1$.

Itt k = az osztályközök száma és b = az adott [eloszlás](#) azon paramétereinek száma, amit a mintából becsléssel határozunk meg.

[Megnézem a kapcsolódó epizódot](#)

A sokaságon belül két [ismérv](#) függetlenségére irányuló vizsgálat. H_0 : a két [ismérv](#) független, az ellenhipotézis pedig, H_1 : a két [ismérv](#) közti kapcsolat sztochasztikus vagy függvényszerű.

A próbát $\chi^2_{1-\alpha}(v)$ jobb oldali kritikus értékkel végezzük el, a nullhipotézist az ennél kisebb, az ellenhipotézist az ennél nagyobb értékek igazolják. A minta elemszáma n , a minta alapján készített [kontingencia tábla](#) sorainak száma r , oszlopainak száma c .

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

ahol a v szabadságfok $v = (r - 1)(c - 1)$.

[Megnézem a kapcsolódó epizódot](#)

Két sokaságban valamely változó eloszlásának egyezőségére irányuló vizsgálat. H_0 : a két sokaságban az [eloszlás](#) egyező, az ellenhipotézis pedig, H_1 : a két [eloszlás](#) nem egyező.

A próbát $\chi^2_{1-\alpha}(v)$ jobb oldali kritikus értékkel végezzük el, a nullhipotézist az ennél kisebb, az ellenhipotézist az ennél nagyobb értékek igazolják. Mintát ezúttal mindkét sokaságból veszünk, az X sokaságból vett minta elemszáma n_X az Y sokaságból vett mintáé n_Y mindkét mintában az osztályközök száma k .

$$\chi^2(v) = n_X \cdot n_Y \cdot \sum_{i=1}^k \left(\frac{1}{n_X + n_Y} \cdot \left(\frac{n_{Xi}}{n_X} - \frac{n_{Yi}}{n_Y} \right)^2 \right)$$

ahol a v szabadságfok $v = k - 1$.

[Megnézem a kapcsolódó epizódot](#)

Mindkét [sokaság](#) normális eloszlású, szórásaik σ_X és σ_Y .

$$Z = \frac{(\bar{y} - \bar{x}) - \delta_0}{\sqrt{\frac{\sigma_Y^2}{n_Y} + \frac{\sigma_X^2}{n_X}}}$$

A nullhipotézis: $H_0: \mu_X - \mu_Y = \delta_0$, ahol δ tetszőleges, de előre megadott érték. A minták elemszáma n_X és n_Y .

[Megnézem a kapcsolódó epizódot](#)

A két [sokaság](#) normális eloszlású és szórásaik egyformák.

$$t(v) = \frac{(\bar{y} - \bar{x}) - \delta_0}{s \cdot \sqrt{\frac{1}{n_Y} + \frac{1}{n_X}}}$$

$$\text{itt } s^2 = \frac{(n_X - 1)s_X^2 + (n_Y - 1)s_Y^2}{n_X + n_Y - 2}$$

A nullhipotézis $H_0: \mu_X - \mu_Y = \delta_0$, ahol δ tetszőleges, de előre megadott érték.

A minták elemszáma n_X és n_Y , szórása s_X és s_Y , a szabadságfok $v = n_Y + n_X - 2$

[Megnézem a kapcsolódó epizódot](#)

A két [sokaság](#) eloszlása és szórása nem ismert, mindkettő szórása véges, és mindkét minta elemszáma elég nagy.

$$Z = \frac{(\bar{y} - \bar{x}) - \delta_0}{\sqrt{\frac{s_Y^2}{n_Y} + \frac{s_X^2}{n_X}}}$$

A nullhipotézis $H_0: \mu_X - \mu_Y = \delta_0$, ahol δ tetszőleges, de előre megadott érték.

A minták elemszáma n_X és n_Y , szórása s_X és s_Y .

[Megnézem a kapcsolódó epizódot](#)

Két [sokaság](#) szórásának összehasonlítására irányuló próba, ha mindkét [sokaság](#) normális eloszlású. A nullhipotézis

$$H_0: \sigma_1^2 = \sigma_2^2$$

$$F = \frac{s_1^2}{s_2^2} \quad F_{1-p}(v_1; v_2) = \frac{1}{F_p(v_2; v_1)}$$

Az F-[eloszlás](#) két szabadságfoka

$$v_1 = n_1 - 1 \text{ és } v_2 = n_2 - 1$$

$$\text{Bal oldali kritikus érték: } \frac{1}{F_{1-\alpha}(v_2; v_1)}$$

$$\text{Jobb oldali kritikus érték: } F_{1-\alpha}(v_1; v_2)$$

Kétoldali kritikus érték:

$$\frac{1}{F_{1-\frac{\alpha}{2}}(v_2; v_1)} \text{ és } F_{1-\frac{\alpha}{2}}(v_1; v_2)$$

[Megnézem a kapcsolódó epizódot](#)

Több [sokaság](#) várható értékének összehasonlítására vonatkozó próba, ha mindegyik [sokaság](#) normális eloszlású és azonos szórású.

A H_0 nullhipotézis: $\mu_1 = \mu_2 = \mu_3 = \dots = \mu_M = \mu$, vagyis az, hogy a várható értékek az összes sokaságra (M db) megegyeznek, míg az ellenhipotézis az, hogy van olyan μ_j amire $\mu_j \neq \mu$.

[Megnézem a kapcsolódó epizódot](#)

A Bartlett-próba több [sokaság](#) szórásának összehasonlítására vonatkozó próba, ha mindegyik [sokaság](#) normális eloszlású.

A H_0 nullhipotézis: $\sigma_1 = \sigma_2 = \sigma_3 = \dots = \sigma_M = \sigma$, vagyis az, hogy az összes [sokaság](#) (M db.) szórása megegyezik, míg az ellenhipotézis az, hogy van olyan σ_j , amire $\sigma_j \neq \sigma$.

$$SSB = \sum_{j=1}^M (n_j - 1) s_j^2 \quad s_b = \frac{SSB}{n-M}$$

A próbafüggvény

$$B^2 = \frac{1}{c} \left(v \cdot \ln s_b^2 - \sum_{j=1}^M v_j \ln s_j^2 \right)$$

$$c = 1 + \frac{1}{3(M-1)} \left(\sum_{j=1}^M \frac{1}{v_j} - \frac{1}{v} \right)$$

Jobb oldali kritikus érték: $\chi_{1-\alpha}^2(M-1)$

[Megnézem a kapcsolódó epizódot](#)

Regressziószámítás

A [regresszió](#) egyenes egyenlete:

$$y = b_0 + b_1 \cdot x$$

$$\text{Ahol } b_1 = \frac{\sum dx \cdot dy}{\sum d^2 x} \text{ és } b_0 = \bar{y} - b_1 \cdot \bar{x}$$

A regressziós egyenes egyenletében szereplő regressziós paraméterek közül b_1 az egyenes meredeksége. A b_0 érték kevésbé jelentős, ez azt adja meg, hogy a magyarázó változó nulla értékéhez milyen y érték tartozik.

[Megnézem a kapcsolódó epizódot](#)

A regressziós egyenes egyenlete $\hat{y} = \hat{b}_0 + \hat{b}_1 \cdot x$

Ez egy [lineáris függvény](#), ami mindegyik x -hez hozzárendel valamilyen y -t. Ezek általában eltérnek a valódi y -októl. Ezeket az eltéréseket reziduumoknak nevezzük.

[Megnézem a kapcsolódó epizódot](#)

A reziduumokból képzett mutató az úgynevezett SSE, jelentése sum of squares of the errors vagyis eltérés-négyzetösszeg.

$$SSE = \sum e_i^2 = \sum (y_i - \hat{y}_i)^2$$

Ha a [regresszió](#) tökéletesen illeszkedik, akkor az $e_i = y_i - \hat{y}_i$ különbségek mindegyike 0, így $SSE=0$. Ha az illeszkedés nem tökéletes, akkor SSE egy pozitív érték, ami az illeszkedés pontatlanságát méri.

[Megnézem a kapcsolódó epizódot](#)

Ha az SSE értékeit elosztjuk a megfigyelt pontok számával és a kapott eredménynek vesszük a gyökét, akkor kapjuk a reziduális szórást:

$$s_e^* = \sqrt{\frac{SSE}{n}} = \sqrt{\frac{\sum e_i^2}{n}} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n}}$$

[Megnézem a kapcsolódó epizódot](#)

Az illeszkedés egy mérőszáma a lineáris [korrelációs együttható](#):

$$r = \frac{\sum dx \cdot dy}{\sqrt{\sum d^2 x \cdot \sum d^2 y}}$$

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

[Megnézem a kapcsolódó epizódot](#)

A magyarázóerőt méri az úgynevezett determinációs együttható, melynek jele R^2 . Ez a kétváltozós lineáris modell esetében megegyezik r^2 -tel.

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

Itt SSE az eltérés-négyzetösszeg, míg SSR az úgynevezett regressziós, vagy magyarázó négyzetösszeg, SST pedig a teljes négyzetösszeg, a köztük lévő kapcsolat pedig...

$$SST = \sum d^2 y \quad SSR = \sum (\hat{y}_i - \bar{\hat{y}})^2 = b_1^2 \sum d^2 x \quad SSE = \sum (y_i - \hat{y}_i)^2 = \sum e_i^2$$

[Megnézem a kapcsolódó epizódot](#)

A [regresszió](#) egyenes egyenlete:

$$\hat{y} = \hat{b}_0 + \hat{b}_1 x$$

Amiből

$$\lg \hat{y} = \lg \hat{b}_0 + \hat{b}_1 \cdot \lg x$$

$$\text{Ahol } \hat{b}_1 = \frac{\sum d \lg x \cdot d \lg y}{\sum d^2 \lg x} \text{ és } \lg \hat{b}_0 = \overline{\lg y} - \overline{\lg x} \cdot \hat{b}_1$$

[Megnézem a kapcsolódó epizódot](#)

A [regresszió](#) egyenes egyenlete:

$$\hat{y} = \hat{b}_0 + \hat{b}_1 x$$

Amiből

$$\lg \hat{y} = \lg \hat{b}_0 + x \cdot \hat{b}_1$$

$$\text{Ahol } \lg \hat{b}_1 = \frac{\sum dx \cdot d \lg y}{\sum d^2 x} \text{ és } \lg \hat{b}_0 = \overline{\lg y} - \bar{x} \cdot \lg \hat{b}_1$$

[Megnézem a kapcsolódó epizódot](#)

Az elaszticitás két összefüggő jelenség közti kapcsolat.

$$\text{Lineáris modellben: } El(\hat{y}, x) = \frac{\hat{b}_1 x}{\hat{y}} = \frac{\hat{b}_1 x}{\hat{b}_0 + \hat{b}_1 x}$$

$$\text{Hatványkitevős modellben: } El(\hat{y}, x) = \hat{b}_1$$

$$\text{Exponenciális modellben: } El(\hat{y}, x) = x \cdot \ln \hat{b}_1$$

[Megnézem a kapcsolódó epizódot](#)

- I. A magyarázó változók nem valószínűségi változók.
- II. A magyarázó változók lineárisan független rendszert alkotnak.
- III. Az eredményváltozó közel lineáris függvénye a magyarázó változónak.
- IV. Az ϵ hibatag feltételes eloszlása normális, várható értéke nulla.
- V. Az ϵ hibatag különböző x -ekhez tartozó értékei korrelálatlanok.

[Megnézem a kapcsolódó epizódot](#)

A paraméterek becslése:

$$\hat{b}_i \pm t_{1-\frac{\alpha}{2}} \cdot (n - k - 1) \cdot s_{\hat{b}_i}$$

A [regresszió](#) becslése:

$$\hat{y}_* \pm t_{1-\frac{\alpha}{2}} \cdot (n - k - 1) \cdot s_{\hat{y}_*}$$

[Megnézem a kapcsolódó epizódot](#)

A többváltozós regressziós modelleket olyankor alkalmazzuk, amikor az eredményváltozó alakulását több magyarázó változó tükrében vizsgáljuk.

A többváltozós [lineáris regresszió](#) egyenlete:

$$y = \hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2 + \dots + \hat{b}_k x_k + \epsilon$$

Az y eredményváltozó itt k darab magyarázó változótól és a hibatagtól függ.

A képletben a $\hat{b}_0$ paraméter a tengelymetszet, a többi $\hat{b}_i$ paraméter pedig azt jelenti, hogy az i -edik magyarázó változó egy egységgel történő változása, mennyivel változtatja az $\hat{y}$ értéket, ha a többi magyarázó változót rögzítjük.

[Megnézem a kapcsolódó epizódot](#)

A kétváltozós esethez hasonlóan a [korreláció](#) itt is a változók közti kapcsolat szorosságát írja le, csak hogy itt egy fokkal rosszabb a helyzet, ugyanis most bármely két változó korrelációját vizsgálhatjuk. Ezt tartalmazza a korrelációmátrix.

$$R = \begin{pmatrix} 1 & r_{y1} & r_{y2} & \dots & r_{yk} \\ r_{1y} & 1 & r_{12} & \dots & r_{1k} \\ r_{2y} & r_{21} & 1 & \dots & r_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ r_{ky} & r_{k1} & r_{k2} & \dots & 1 \end{pmatrix}$$

Itt r_{ij} az x_i és az x_j magyarázó változók közti korrelációt írja le, tehát például r_{12} az x_1 és az x_2 közti korrelációt jelenti.

r_{iy} pedig az x_i magyarázó változó és az y eredményváltozó közti kapcsolatot jelenti.

Mivel $r_{ij} = r_{ji}$ a [korreláció-mátrix](#) szimmetrikus. Az áttekinthetőbb felírás kedvéért a felső háromszöget el is szokták hagyni.

[Megnézem a kapcsolódó epizódot](#)

A [lineáris regresszió](#) egyenlete: $\hat{y} = \hat{b}_0 + \hat{b}_1x_1 + \hat{b}_2x_2 + \dots + \hat{b}_kx_k$

A tesztelés úgy zajlik, hogy nullhipotézisnek tekintjük a $H_0 : b_i = 0$ feltevést, ellenhipotézisnek pedig azt, hogy $H_1 : b_i \neq 0$.

A nullhipotézis azt állítja, hogy a modellben a b_i paraméter szignifikánsan nulla, vagyis az i -edik magyarázó változó felesleges, annak hatása az eredményváltozóra nulla. Az ellenhipotézis ezzel szemben az, hogy $b_i \neq 0$ vagyis az i -edik magyarázó változónak a regresszióban nem nulla hatása van.

[Megnézem a kapcsolódó epizódot](#)

Szóródás oka	Négyzetösszeg	Szabadságfok	Átlagos négyzetösszeg	F
Regresszió	SSR	k	$MSR = \frac{SSR}{k}$	$F = \frac{MSR}{MSE}$
Hiba	SSE	$n - k - 1$	$MSE = \frac{SSE}{n - k - 1}$	
Teljes	SST	$n - 1$		

[Megnézem a kapcsolódó epizódot](#)

A multikollinearitás röviden összefoglalva azt jelenti, hogy két vagy több magyarázó változó között túl szoros [korrelációs kapcsolat](#) van, és ez zavarja a becslést.

A multikollinearitás mérésére az úgynevezett VIF (variance inflator factor) variancia növelő faktor van forgalomban.

$$VIF_j = \frac{1}{1 - R_j^2}$$

A képletben szereplő R_j^2 a j -edik magyarázó változó és az összes többi magyarázó változó közti determinációs együttható.

[Megnézem a kapcsolódó epizódot](#)

Az auto[korreláció](#) a [regresszió](#) maradéktagjának a saját későbbi értékeivel való korrelációját jelenti, vagyis egyfajta szabályszerűséget a maradékváltozóban. Ideális esetben a maradéktagnak véletlenszerűnek kell lennie, bármiféle szabályszerűségért a magyarázó változók felelnek a regresszióban.

Az autó[korreláció](#) tesztelésére a Durbin-Wattson-tesztet használjuk.

[Megnézem a kapcsolódó epizódot](#)

A Durbin-Wattson-teszt lényegében egy hipotézisvizsgálat, aminek részletezésére nem térünk ki, mindössze a használatát nézzük meg.

Maga a próbafüggvény

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=2}^n e_t^2}$$

A [szignifikanciaszint](#) α , a próba elvégzése pedig az alábbi módon történik:

d_L és d_U értékeket kikeressük a táblázatból,

n = a megfigyelések száma,

k = a magyarázó változók száma,

végül megnézzük, hogy a próbafüggvény melyik tartományba esik.

[Megnézem a kapcsolódó epizódot](#)