„TÁMOP-4.1.2/A/1-11/1-2011-0015 Egészségügyi Ügyvitelszervező Szakirány: Tartalomfejlesztés és Elektronikus Tananyagfejlesztés a BSc képzés keretében”

Biostatisztika

e-Book

Dr. Dinya Elek

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL

A projektek az Európai Unió támogatásával valósulnak meg.

Tartalomjegyzék

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

1. Bevezetés a mátrixok világába

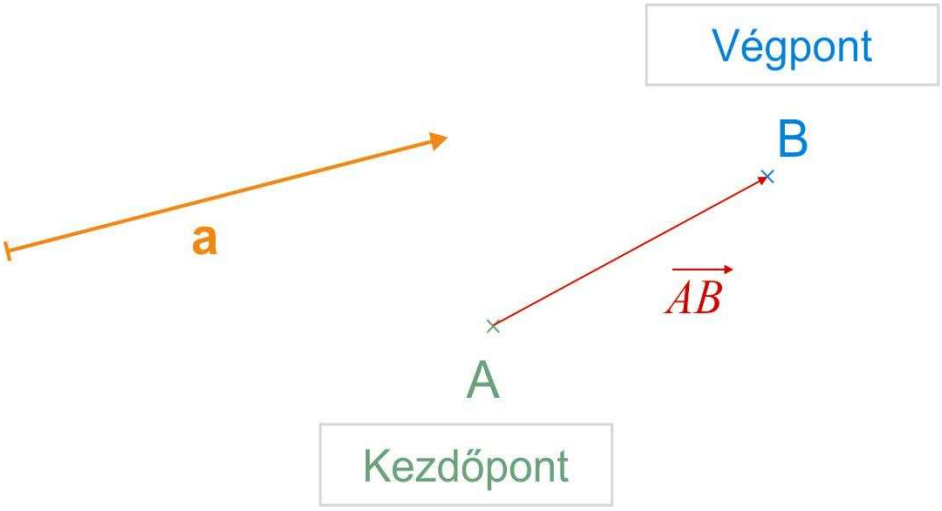
1.1. Vektorok

Alapfogalmak:

- Skalármennyiség: konkrét számérték (terület, térfogat stb.).
- Vektormennyiség: irányított érték (erő, sebesség stb.).
- Szabadvektor: önmagával párhozamosan eltolható.
- Fixvektor: fix kezdőpont.
- Csúsztatható vektor: saját irányegyenesük mentén mozgatható.

Definíció: a tér irányított szakaszait nevezzük *vektoroknak*, amelyeknek adott a nagysága és iránya. Másképp fogalmazva a vektor egy irányított szakasz, vagy azzal jellemezhető mennyiség.

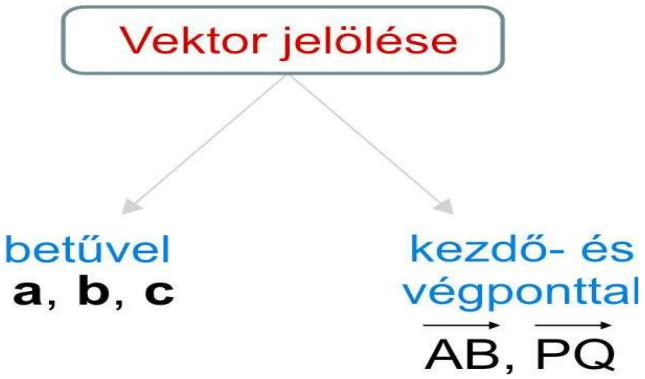
Példák vektorokra:



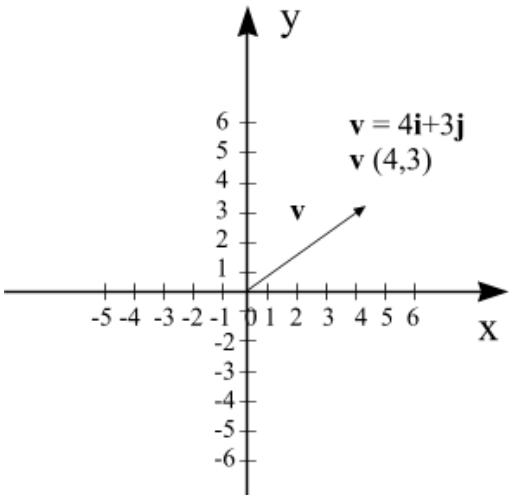
Jelölésük: a vektort megadhatjuk a kezdő és végpontja segítségével ($\overrightarrow{AB}$) vagy jelölhetjük kisbetűvel kétféle módon: **a** vagy $\underline{a}$

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.



Koordináta rendszerben origó kezdőpontú vektort rendezett számpár jellemzi a síkban, térben pedig rendezett számhármas



Definíció: két vektor azonos (egyenlő), ha hosszuk (nagyságuk) is és irányuk is megegyezik.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

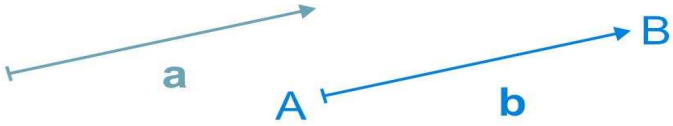
Példa:

Az ábra jelöléseivel:

$\mathbf{a} = \mathbf{b}$

$\overrightarrow{AB} = \mathbf{a}$

$\mathbf{b} \equiv \overrightarrow{AB}$



Definíció: vektorok egyenlősége ekvivalenciarelációt jelent:

- reflexív: $\forall \mathbf{a}$: $\mathbf{a} = \mathbf{a}$
- szimmetrikus: $\forall \mathbf{a}, \mathbf{b}$: ha $\mathbf{a} = \mathbf{b} \Rightarrow \mathbf{b} = \mathbf{a}$
- tranzitív: $\forall \mathbf{a}, \mathbf{b}, \mathbf{c}$: ha $\mathbf{a} = \mathbf{b}$ és $\mathbf{b} = \mathbf{c} \Rightarrow \mathbf{a} = \mathbf{c}$.

Definíció: a vektor hosszát a vektor abszolút értékének is nevezzük (nem negatív valós szám).

A fenti vektor hossza:

$$|\mathbf{v}| = \sqrt{a^2 + b^2} = \sqrt{4^2 + 3^2} = \sqrt{25} = 5$$

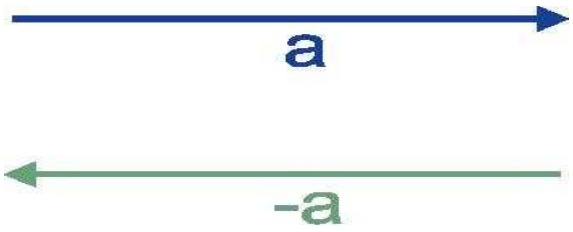
Definíció: az olyan vektort (**0**), amelynek megegyezik a kezdőpontja és a végpontja és abszolút értéke *0*, nullvektornak nevezzük. Iránya tetszőleges, minden vektorral párhuzamos és minden vektorra merőleges. Ilyen vektorból csak egy létezik.

*Definíció: h*a egy vektor abszolút értéke *1*, akor egységvektornak nevezzük. Ilyen vektorból végtelen sok létezik.

Definíció: az ***a*** vektor **ellentettje:** az ***a*** vektort, amelyik vele egyenlő abszolútértékű, egyező állású, de vele ellentétes irányú. Jelölése: – **a** .

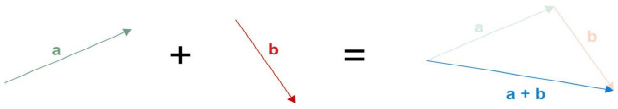
Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

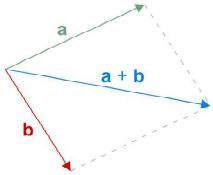


*Definíció: k*ét vektor *összegén* egy harmadik vektort értünk, amelyet meghatározhatunk paralelogramma-módszer, vagy összefűzés (háromszög-módszer, sokszög-módszer) segítségével.

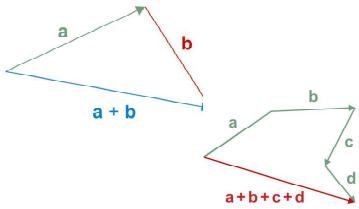
Vektorműveletek



Paralelogramma módszer



Háromszög módszer



A vektorösszeadás kommutatív és asszociatív:

$\forall \mathbf{a}, \mathbf{b}$ esetén: $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$

$\forall \mathbf{a}, \mathbf{b}, \mathbf{c}$ esetén: $(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c})$.

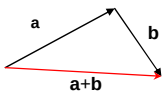
Definíció: az ***a*** és ***b*** vektorok ***a*** **−** ***b*** különbségén azt a ***c*** vektort értjük, melyre ***b*** + ***c*** = ***a.***

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

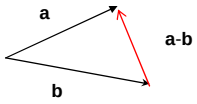
Koordinátákkal kifejezve: **a** (a1,a2) **b** (b1,b2)

Összeadás



$\mathbf{a}+\mathbf{b}$ (a_1+b_1, a_2+b_2)

Kivonás



$\mathbf{a}-\mathbf{b}$ (a_1-b_1, a_2-b_2)

Megjegyzés:

• Két vektor különbségét megkapjuk úgy, hogy közös kezdőpontba toljuk őket, mert ekkor a különbségvektor a végpontjaikat összekötő vektor lesz, a kisebbítendő felé irányítva.

• A vektorok összeadása, illetve kivonása során az eredmény esetleg a **0** is lehet.

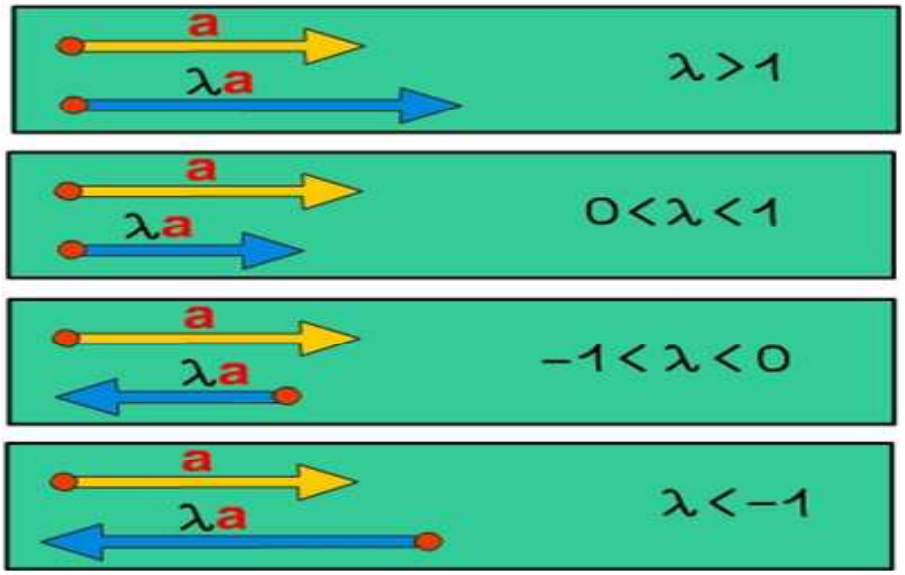
• Bármely **a** vektor esetén $\mathbf{a} + \mathbf{0} = \mathbf{a}$ és $\mathbf{a} - \mathbf{0} = \mathbf{a}$.

Definíció: Egy $\boldsymbol{a}$ vektor és egy λ szám szorzata egy vektor, amelynek hossza $|\lambda\mathbf{a}| = |\lambda|\cdot|\mathbf{a}|$, párhuzamos $\boldsymbol{a}$-val és $\lambda > 0$ esetén egyirányú, $\lambda < 0$ esetén ellentétes irányú $\boldsymbol{a}$-val.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Vektor szorzása λ számmal (skalárral)



Vektorok számmal való szorzására érvényesek a következő műveleti szabályok:

$\forall \lambda, \mu$ skalár és $\forall \boldsymbol{a}$ esetén: $\lambda(\mu \boldsymbol{a}) = (\lambda\mu)\mathbf{a}$ (asszociativitás)

$\forall \lambda$ és $\forall \mathbf{a}, \mathbf{b}$ vektor esetén: $\lambda(\boldsymbol{a} + \boldsymbol{b}) = \lambda \boldsymbol{a} + \lambda \boldsymbol{b}$ (disztributívitás)

$\forall \lambda$, μ és $\forall \mathbf{a}$ esetén: $(\lambda + \mu)\boldsymbol{a} = \lambda \boldsymbol{a} + \mu \boldsymbol{a}$ (disztributívitás)

Definíció: legyenek $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}n$ tetszőleges vektorok a térben, $c_1, c_2, \ldots, cn$ pedig valós számok. Az $c_1\boldsymbol{a}_1 + c_2\boldsymbol{a}_2 + \cdots + cn\boldsymbol{a}n$ kifejezést az $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}n$ vektorok *lineáris kombinációjának* nevezzük.

Példa: ha $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ vektorok, akkor $3\boldsymbol{a} - 4\boldsymbol{b} + 6\boldsymbol{c}$ egy lineáris kombinációjuk. Ha megadunk néhány vektort, akkor ezeknek végtelen sok lineáris kombinációja létezik, hiszen az együtthatók tetszőleges valós számok lehetnek.

Állítás: legyenek $\boldsymbol{a}, \boldsymbol{b}$ és $\boldsymbol{c}$ a tér vektorai. Ha $\boldsymbol{a}$, $\mathbf{b}$ és $\boldsymbol{c}$ nincsenek egy síkban, akkor a tér minden $\boldsymbol{v}$ vektora egyértelműen előállítható $\boldsymbol{a}$, $\boldsymbol{b}$ és $\boldsymbol{c}$ lineáris kombinációjaként.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Definíció: Az $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}n$ vektorok triviális lineáris kombinációján a $0\cdot\mathbf{a}1 + 0\cdot\mathbf{a}2 + \ldots + 0\cdot\mathbf{a}n$ kifejezést értjük.

Megjegyzés: akkor beszélünk triviális lineáris kombinációról, ha minden együttható *0*. Természetesen az eredmény csak a **0** vektor lehet.

Definíció: Az $\boldsymbol{a}1, \boldsymbol{a}2, \ldots, \boldsymbol{a}n$ vektorokat lineárisan függetlennek nevezzük, ha csak a triviális lineáris kombinációjuk **0**. Minden más esetben a vektorokat lineárisan összefüggőnek hívjuk.

*Állítás: k*ét vektor lineárisan összefüggő, ha párhuzamosak egymással.

Állítás: A tér három vektora akkor lineárisan összefüggő, ha egy síkban vannak. A tér pl. négy vektora mindenképpen lineárisan összefüggő.

Definíció: A térbeli vektorok egy lineárisan független vektorhármasát bázisnak nevezzük.

Definíció: Ha $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ a tér egy bázisa és $\boldsymbol{v} = \alpha_1\boldsymbol{e}_1 + \alpha_2\boldsymbol{e}_2 + \alpha_3\boldsymbol{e}_3$, akkor az α_1, α_2, α_3 számokat a $\boldsymbol{v}$ vektor ($\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$ bázisra vonatkozó) koordinátáinak nevezzük.

Megjegyzés: a bázisvektorok általánosan használt jelölés rendszere **i** (*abszcissza)*, **j** (ordinata**)**, **k** (kóta**).** Tulajdonságaik:

– egységnyi hosszúságúak ($|\mathbf{i}| = |\mathbf{j}| = |\mathbf{k}| = 1$),

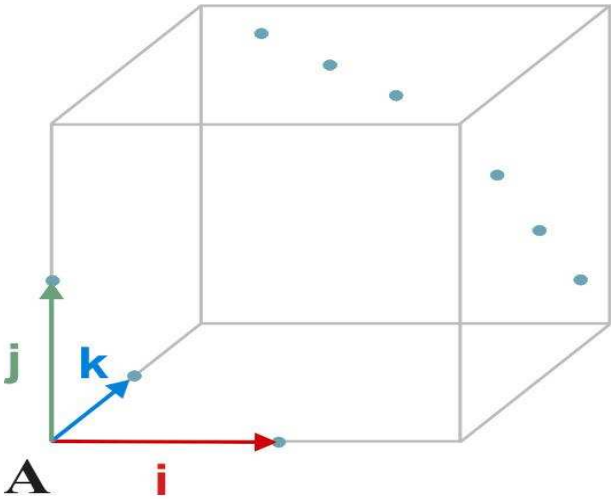
– páronként ortogonálisak egymásra,

– i, **j**, **k** sorrendben ún. *jobbrendszert* alkotnak. (ha **k** végpontja felől nézünk a másik két bázisvektor síkjára, akkor **i**-t a **j**-be pozitív irányú, óramutató járásával ellentétes,180 foknál kisebb szögű forgás viszi át.

A tér egységvektorai:

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.



*Definíció: e*gy *Q* pont helyvektorán az $\overrightarrow{OQ}$ vektort értjük, ahol *O* az origó. Az így definiált vector ún. kötöttvektor, mivel kezdőpontja rögzített.

Definíció: Egy *Q* pont koordinátáin a helyvektorának a koordinátáit értjük.

Definíció: Két vektor összegének koordinátái az eredeti vektorok megfelelő koordinátáinak összegével egyenlő.

Definíció: Két vektor különbségének koordinátái az eredeti vektorok megfelelő koordinátáinak különbségével egyenlő

.

Definíció: Ha egy vektort egy ***c*** számmal szorzunk, akkor az így kapott vektor minden koordinátája a eredeti vektor megfelelő koordinátájának ***c***-szerese lesz.

Definíció: Az a(a_1, a_2, a_3) vektor hossza

$$|\mathbf{a}| = \sqrt{a_1^2 + a_2^2 + a_3^2}$$

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Definíciók:

n koordinátával jellemzett vektorok

2 féle megadási mód:

oszlopvektor: $\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ . \\ . \\ . \\ a_n \end{bmatrix}$ sorvektor: $\mathbf{a}^* = [a_1, a_2, \ldots a_n]$

Vektorokkal való műveletek

Adott két vektor. Számítsuk ki a következőket: a+b; ab; a vektor hosszát valamint a 3a-t, a*b!

$$\mathbf{a} = \begin{bmatrix} -2 \\ 1 \\ 0 \\ 2 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} -2 \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad \mathbf{a}+\mathbf{b} = \begin{bmatrix} -4 \\ 1 \\ 0 \\ 3 \end{bmatrix} \quad |\mathbf{a}| = \sqrt{(-2)^2 + 1^2 + 0^2 + 2^2} = 3$$

$$3\mathbf{a} = \begin{bmatrix} -6 \\ 3 \\ 0 \\ 6 \end{bmatrix} \quad \mathbf{a}^*\mathbf{b} = [-2,1,0,2] \begin{bmatrix} -2 \\ 0 \\ 0 \\ 1 \end{bmatrix} = 4+0+0+2 = 6$$

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Definíció: két vektor skaláris szorzatán az alábbi szorzatot értjük:

- Két tetszőleges **a** = [a_1, a_2, ... , a_n] és **b** = [b_1, b_2, ... , b_n] vektor skaláris szorzata alatt a következőt értjük:

$$\mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^{n} a_i b_i = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n$$

- ahol Σ az összegzést és *n* a vektortér dimenzióját jelöli.

Skaláris szorzat tulajdonságai

1. Kommutatív: a·b = b·a
2. A skaláris szorzás egy *c* skaláris tényezővel

   asszociatív: *c*(a·b)=(*c*a)b
3. Disztributív: a·(b+c)= a·b + a·c

Definíció: az ***a*** és ***b*** vektorok skaláris szorzatán az

ab = |**a**|·|**b**|·cos φ

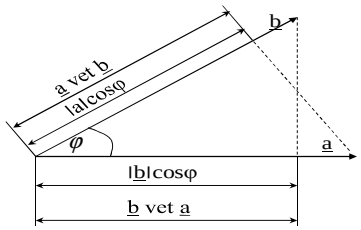
Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

számot értjük, ahol φ az $\boldsymbol{a}$ és $\boldsymbol{b}$ vektorok hajlásszöge.

Állítás: két vektor skaláris szorzata akkor és csak akkor *0*, ha a két vektor merőleges egymásra.

Két vektor skaláris szorzatának kommutatvitása



A kommutativitás következik a skaláris szorzat definíciójából vagy az ábrán látott két háromszög hasonlósága alapján, mivel $|\underline{b}|\cdot\cos\varphi = |\,\underline{b}$ vet $\underline{a}|$, ahol a

$\underline{b}$ vet $\underline{a}$ a $\underline{b}$ vektor vetülete az $\underline{a}$ vektorra, és

$|\underline{a}|\cdot\cos\varphi = |\underline{a}$ vet $\underline{b}|$, úgyhogy

$\underline{a}\cdot\underline{b} = \underline{b}\cdot\underline{a}$

Definíció: Az **a** és **b** vektorok vektoriális szorzatán azt az **a×b**-vel jelölt vektort értjük.

A vektoriális szorzatra vonatkozóan teljesülnek:

- hossza $|\mathbf{a} \times \mathbf{b}| = |\mathbf{a}|\cdot|\mathbf{b}|\cdot\sin\varphi$, ahol φ az **a** és **b** vektorok hajlásszöge,
- iránya merőleges az **a** és **b** vektorokra,

• **a**, **b** és **a × b** ebben a sorrendben jobbrendszert alkot.

Állítás: két vektor akkor és csak akkor párhuzamos, ha $\mathbf{a} \times \mathbf{b} = \mathbf{0}$

Állítás: tetszőleges $\boldsymbol{a}$ és $\boldsymbol{b}$ vektorok és λ valós számesetén igaz az alábbi egyenlőség:

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

$$\lambda\,(\mathbf{a} \times \mathbf{b}) = \lambda\mathbf{a} \times \mathbf{b} = \mathbf{a} \times \lambda\mathbf{b}$$

Állítás: az $\mathbf{a}$ (a_1, a_2, a_3) és $\mathbf{b}$ (b_1, b_2, b_3) vektorok vektoriális szorzata determináns alakban

$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} i & j & k \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}$$

Állítás: az $\mathbf{a}$ és $\mathbf{b}$ vektorok által kifeszített paralelogramma területe a két vektor vektoriális szorzatának abszolút értékével egyenlő

$$T = |\mathbf{a} \times \mathbf{b}|$$

Állítás: az $\mathbf{a}$ és $\mathbf{b}$ vektorok által kifeszített háromszög területe

$$T = |\mathbf{a} \times \mathbf{b}|/2$$

Definíció: az $\mathbf{a}$, $\mathbf{b}$ és $\mathbf{c}$ vektorokból képzett $(\mathbf{a} \times \mathbf{b})\,\mathbf{c}$ kifejezést az $\mathbf{a}$, $\mathbf{b}$ és $\mathbf{c}$ vektorok vegyesszorzatának nevezzük.

Megjegyzések:
- Az elnevezés arra utal, hogy hogy a kifejezésen belül kétfajta szorzás is szerepel.
- A vegyesszorzat eredménye skalár.

Állítás: ha $\mathbf{a}$, $\mathbf{b}$ és $\mathbf{c}$ nem esnek egy síkba, akkor vegyesszorzatuk abszolút értéke megegyezik az általuk kifeszített paralellepipedon térfogatával:

$$V = |(\mathbf{a} \times \mathbf{b})\,\mathbf{c}|$$

Állítás: az $\mathbf{a}$, $\mathbf{b}$ és $\mathbf{c}$ vektorok akkor és csak akkor esnek egy síkba, ha $(\mathbf{a} \times \mathbf{b})\,\mathbf{c}$ vegyesszorzatuk $\mathbf{0}$.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Felcserélési tétel (a vegyesszorzat eredménye nem változik): tetszőleges **a**, **b** és **c** vektorok esetén

$$(\mathbf{a} \times \mathbf{b})\,\mathbf{c} = \mathbf{a}\,(\mathbf{b} \times \mathbf{c})$$

Állítás: az **a**($a1$, $a2$, $a3$), **b**($b1$, $b2$, $b3$) és **c**($c1$, $c2$, $c3$) vektorok vegyesszorzata

$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}$$

Állítás: az **a**, **b** és **c** vektorok által kifeszített tetraéder térfogata egyenlő a vegyesszorzatuk abszolút értékének hatodrészével

$$\mathbf{V} = |\mathbf{abc}|/6$$

Cauchy–Bunyakowski–Schwarz egyenlőtlenség:

$$|\langle x, y\rangle| \leq \|x\| \cdot \|y\|$$

ahol

$$\|x\| = \sqrt{\langle x, x\rangle}.$$

vagy másképpen kifejezve az egyenlőtlenséget:

$$\left|\sum_{i=1}^{n} x_i \bar{y}_i\right|^2 \leq \sum_{j=1}^{n} |x_j|^2 \sum_{k=1}^{n} |y_k|^2.$$

azaz

$$(a_1b_1 + a_2b_2 + \ldots + a_nb_n) \leq (a_1^2 + a_2^2 + \ldots + a_n^2)(b_1^2 + b_2^2 + \ldots + b_n^2)$$

Minkowsky (háromszög egyenlőtlenség):

$$\|x + y\| \leq \|x\| + \|y\|.$$

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

1.2. Mátrixok

Általánosságban *mátrixnak* nevezünk egy téglalap elrendezésű, $m \cdot n$ számú, a_{ij} valós számot (általában, de lehet komplex szám is) tartalmazó táblázatot. A mátrixokat nagy betűvel jelöljük és szögletes zárójelben adjuk meg:

$$\underset{(\mathbf{m},\mathbf{n})}{\mathbf{A}} = \begin{bmatrix} \mathbf{a}_{11} & \mathbf{a}_{12} & \cdots & \mathbf{a}_{1n} \\ \mathbf{a}_{21} & \mathbf{a}_{22} & \cdots & \mathbf{a}_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \mathbf{a}_{m1} & \mathbf{a}_{m2} & \cdots & \mathbf{a}_{mn} \end{bmatrix}$$

Az adott mátrix *m·n típusú: m* sorból és *n* oszlopból áll, az a_{ij} a mátrix *i*-edik sorában és *j*-edik oszlopában lévő eleme. Ha *m=n*, akkor a mátrixot *négyzetes mátrixnak* (vagy kvadratikus) nevezzük és a sorok száma a mátrix *rendjét* is meghatározza. Ha egy *A* mátrix sorait és oszlopait felcseréljük, akkor kapjuk az *A** transzponált mátrixt.

$$\underset{(\mathbf{n},\mathbf{m})}{\mathbf{A}^{*}} = \begin{bmatrix} \mathbf{a}_{11} & \mathbf{a}_{21} & \cdots & \mathbf{a}_{m1} \\ \mathbf{a}_{12} & \mathbf{a}_{22} & \cdots & \mathbf{a}_{m2} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \mathbf{a}_{1n} & \mathbf{a}_{2n} & \cdots & \mathbf{a}_{mn} \end{bmatrix}$$

A transzponálás során a kvadratikus mátrix *n* rendje nem változik és a transzponált mátrix transzponáltja az eredeti mátrixt adja eredményül.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

1.2.1. Alapműveletek

a) Mátrixok egyenlősége

Két mátrix csak akkor egyenlő egymással, ha soraik és oszlopaik száma egyenlő (azonos típusúak) és az azonos helyen álló elemeik megegyeznek.

b) Összeadás, kivonás

A két művelet csak azonos típusú mátrixokra értelmezett. Az eredmény mátrix (összeg vagy különbség mátrix) a két mátrix típusával azonos, és elemei a két mátrix azonos helyén lévő elemeinek az összege vagy különbsége. A két művelet tetszőleges számú mátrixokra is elvégezhető.

$$\mathbf{C = A + B} = \begin{bmatrix} 1 & 2 & 3 \\ -5 & 3 & -1 \\ 4 & 2 & -1 \end{bmatrix} + \begin{bmatrix} 0 & 2 & 1 \\ 5 & 1 & 2 \\ 3 & -1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 4 & 4 \\ 0 & 4 & 1 \\ 7 & 1 & -1 \end{bmatrix}$$

$$\mathbf{C = A - B} = \begin{bmatrix} 1 & 2 & 3 \\ -5 & 3 & -1 \\ 4 & 2 & -1 \end{bmatrix} - \begin{bmatrix} 0 & 2 & 1 \\ 5 & 1 & 2 \\ 3 & -1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 2 \\ -10 & 2 & -3 \\ 1 & 3 & -1 \end{bmatrix}$$

Egy mátrix spurja (mátrix nyoma) a főátlóban lévő elemeknek az összege. Pl. az A mátrix spurja 5. Jelölésben *Sp(A)=3.*

c) Konstanssal való szorzás

A mátrix minden elemét megszorozzuk az adott számmal

$$\mathbf{C = k \cdot A = 2 \cdot A = 2 \cdot} \begin{bmatrix} 1 & 2 & 3 \\ -5 & 3 & -1 \\ 4 & 2 & -1 \end{bmatrix} = \begin{bmatrix} 2 & 4 & 6 \\ -10 & 6 & -2 \\ 8 & 4 & -2 \end{bmatrix}$$

d) Mátrix szorzása mátrixszal

Két mátrix csak akkor szorozható össze, ha az *A* mátrix oszlopainak a száma azonos a *B* mátrix sorainak a számával. Ha ez feltétel igaz az *A*, *B* mátrixokra, akkor a két mátrix az adott sorrendben *konformábilis*. Vigyázzunk, mert a szorzás általában nem kommutatív művelet, vagyis az $A \cdot B = B \cdot A$ nem mindig igaz. Ezalól csak a diagonál mátrixok szorzása kivétel, mert az ilyen mátrixokra a szorzás művelete kommutatív. A műveletnél az *A* mátrix megfelelő sorait szorozzuk a *B* mátrix megfelelő oszlopaival:

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

1. Az A mátrix 1. sora * a B mátrix 1. oszlopával, utána a 2. sor* az 1. oszloppal, majd a 3. sor*az 1. oszloppal stb.
2. Az A mátrix 1. sora * a B mátrix 2. oszlopával, utána a 2. sor* a 2. oszloppal, majd a 3. sor*a 2. oszloppal stb.
3. az eljárást a fentieknek megfelelően minden sorra és oszloppal elvégezzük.

Példa:

$$A \cdot B = \begin{bmatrix} 1 & 0 & -1 \\ -2 & 2 & -2 \\ 2 & 3 & 2 \end{bmatrix} \cdot \begin{bmatrix} 2 & 2 \\ 1 & -1 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} (1\cdot 2)+(0\cdot 1)+(-1\cdot 0) & (1\cdot 2)+(0\cdot -1)+(-1\cdot 3) \\ (-2\cdot 2)+(2\cdot 1)+(-2\cdot 0) & (-2\cdot 2)+(2\cdot -1)+(-2\cdot 3) \\ (2\cdot 2)+(3\cdot 1)+(2\cdot 0) & (2\cdot 2)+(3\cdot -1)+(2\cdot 3) \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ -2 & -12 \\ 7 & 7 \end{bmatrix}$$

1.2.2. Azonosságok

$$E \cdot A = A \cdot E = A$$

Egységmátrixszal szorozva az eredeti mátrixt kapjuk.

$$0 \cdot A = A \cdot 0 = 0$$

Zérusmátrixszal való szorzás zérusmátrixt eredményez.

1.2.3. Többtényezős mátrix szorzás

A két tényezős konformabilitást tetszőleges tagra is kiterjeszthetjük és szorzás ilyen sorrendben elvégezhető:

$$\underset{(m,n)}{A} \cdot \underset{(n,k)}{B} \cdot \underset{(k,l)}{C} \cdot \underset{(l,p)}{D}$$

Speciális eset a mátrix hatványozása, amit a mátrix *n*-szeri ismételt szorzásával kapunk meg:

$$A \cdot A \cdot A \cdot \ldots \cdot A = A^n$$

Megállapodás szerint $A^0 = E$. Az egységmátrix n-edik hatványa szintén egységmátrix, a zérusmátrix *n*-edik hatványa pedig zérusmátrix.

Nilpotensnek nevezzük az A mátrixt, ha *n*-edik hatványára igaz, hogy a zérusmátrixt adja eredményül:

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

$$A^n = 0$$

Idempotens az A mátrix (önnmagát visszaadó), ha teljesül rá:

$$A^n = A$$

1.3. Determinánsok

A két ismeretlent *(x, y)* tartalmazó un. elsőfokú (az ismeretlen tényezők az elsőhatványon szerepelnek) egyenletrendszerek megoldására három lehetőségünk van: a) *helyettesítő módszer alkalmazása* b) *egyenlő együtthatók módszerének alkalmazása* c) *determinánsok módszerének alkalmazása.* Tekintsük az általános egyenletrendszer alakját:

$$a_1x+b_1y=c_1$$
$$\underline{a_2x+b_2y=c_2}$$

Képezzük az együtthatókból az alábbi másodrendű determinánsokat és adjuk meg az értéküket meghatározó formulákat is:

$$D=\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = (a_1 \cdot b_2) - (b_1 \cdot a_2)$$

$$D_x=\begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix} = (c_1 \cdot b_2) - (b_1 \cdot c_2)$$

$$D_y=\begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix} = (a_1 \cdot c_2) - (c_1 \cdot a_2)$$

Az egyenletrendszer megoldása a determinánsok segítségével:

$$x = \frac{D_x}{D} \text{ illetve } y = \frac{D_y}{D}$$

Nyilván $D \neq 0$ esetén van csak megoldás.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Példa:

Oldjuk meg az alábbi elsőfokú egyenletrendszert a determinánsok segítségével:

$$4x+3y=6$$
$$\underline{2x+y=4}$$

Vegyük az egyenletrendszer másodrendű determinánsát, amit az együtthatókból képzünk (főátló szorzata – mellékátló szorzata):

$$D=\begin{vmatrix}4 & 3\\ 2 & 1\end{vmatrix}=(4\cdot 1)-(2\cdot 3)=-2$$

Mivel a D≠0, ezért az egyenletrendszer megoldható.

$$x=\frac{D_x}{D}=\frac{\begin{vmatrix}6 & 3\\ 4 & 1\end{vmatrix}}{\begin{vmatrix}4 & 3\\ 2 & 1\end{vmatrix}}=\frac{(6\cdot 1)-(4\cdot 3)}{-2}=\frac{-6}{-2}=3$$

$$y=\frac{D_y}{D}=\frac{\begin{vmatrix}4 & 6\\ 2 & 4\end{vmatrix}}{\begin{vmatrix}4 & 3\\ 2 & 1\end{vmatrix}}=\frac{(4\cdot 4)-(2\cdot 6)}{-2}=\frac{4}{-2}=-2$$

Tehát a keresett megoldások: x=3 és y=-2.

A fenti elv három vagy több ismeretlenes egyenletrendszer esetén is alkalmazható, de ilyenkor figyelemmel kell lenni az aldeterminánsok előjelére.

1.3.1. Mátrix determinánsa

Csak kvadratikus mátrixnak van determinánsa, amit a mátrix elemeiből képzünk. Ha a mátrix determinánsa detA ≠ 0, akkor a mátrix *reguláris*, ha detA = 0, akkor a mátrix *szinguláris*. Vizsgáljuk meg a következő mátrix determinánsát:

$$A=\begin{bmatrix}1 & 0 & -1\\ -2 & 2 & -2\\ 2 & 3 & 2\end{bmatrix}$$

Fejtsük sorba a mátrixt az első sora szerint. A determinánst bármelyik sora vagy oszlopa szerint kifejthetjük, csak vegyük figyelembe az együtthatók előjelszabályát. Az előjel szabály

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

(sakktábla szabály) pl. egy harmadrendű determinánsra (de ez értelemszerűen kibővül a feladatnak megfelelően)

$$\begin{matrix} + & - & + \\ - & + & - \\ + & - & + \end{matrix}$$

A kifejtés azt jelenti, hogy a kiszemelt sor vagy oszlop együtthatóival szorozzuk a hozzátartozó aldeterminánsokat. Most fejtsük ki a determinánst az első sora szerint (a kifejtés technikája: pl. $a_{11}=1$-hez tartozó aldeterminánst megkapjuk, ha letakarjuk az első sort és az első oszlopot, a megmaradt elemek lesznek az a_{11}-hez tartozó determináns elemei):

$$\det \mathbf{A} = \begin{vmatrix} \mathbf{1} & \mathbf{0} & \mathbf{-1} \\ \mathbf{-2} & \mathbf{2} & \mathbf{-2} \\ \mathbf{2} & \mathbf{3} & \mathbf{2} \end{vmatrix} = \mathbf{1}\begin{vmatrix} \mathbf{2} & \mathbf{-2} \\ \mathbf{3} & \mathbf{2} \end{vmatrix} - \mathbf{0}\begin{vmatrix} \mathbf{-2} & \mathbf{-2} \\ \mathbf{2} & \mathbf{2} \end{vmatrix} - \mathbf{1}\begin{vmatrix} \mathbf{-2} & \mathbf{2} \\ \mathbf{2} & \mathbf{3} \end{vmatrix} = \mathbf{10} + \mathbf{0} + \mathbf{10} = \mathbf{20}$$

Mivel a determináns $\neq 0$, ezért a mátrix reguláris.

1.3.2. Mátrix rangja

Az *A* mátrix rangja az a $\rho(A) = r$ természetes szám, ha az *r*-edrendű kvadratikus minormátrixai között van legalább egy olyan, amely reguláris, de az összes *r+1*-edrendű már szinguláris. Következésképp, az $m \cdot n$ mátrix rangja nem lehet nagyobb sem sorainak, sem oszlopainak számánál. A rang fontos szerepet játszik pl. a lineáris egyenletrendszerek megoldásánál.

Az előbbi mátrix rangja $\rho(A) = 3$, mivel a determinánsa láttuk, hogy $\neq 0$.

1.3.3. Inverz mátrix

Nagyon fontosak a lineáris egyenletrendszerek megoldásában vagy egyes többváltozós statisztikai módszerek elméletében. Vezessük be az adjungált mátrix fogalmát: egy négyzetes mátrix adjungáltján azt a transzponált *hipermátrixt* értjük, amelynek elemei szintén mátrixok, mégpedig az a_{ij} elemeihez tartozó *előjeles aldeterminánsok* (lásd a fenti előjelszabályt) alkotják a mátrix elemeit.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

$$\mathbf{adjA} = \begin{bmatrix} \mathbf{A_{11}} & \mathbf{A_{21}} & \cdots & \mathbf{A_{n1}} \\ \mathbf{A_{12}} & \mathbf{A_{22}} & \cdots & \mathbf{A_{n2}} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ \mathbf{A_{1n}} & \mathbf{A_{2n}} & \cdots & \mathbf{A_{nn}} \end{bmatrix}$$

Nézzük meg a fenti mátrix adjungáltját. Vegyük sorba az egyes elemekhez tartozó előjeles aldeterminánsokat:

$$\mathbf{A_{11}} = \begin{vmatrix} 2 & -2 \\ 3 & 2 \end{vmatrix} = 10 \qquad \mathbf{A_{12}} = -\begin{vmatrix} -2 & -2 \\ 2 & 2 \end{vmatrix} = 0 \qquad \mathbf{A_{13}} = \begin{vmatrix} -2 & 2 \\ 2 & 3 \end{vmatrix} = -10$$

$$\mathbf{A_{21}} = -\begin{vmatrix} 0 & -1 \\ 3 & 2 \end{vmatrix} = -3 \qquad \mathbf{A_{22}} = \begin{vmatrix} 1 & -1 \\ 2 & 2 \end{vmatrix} = 4 \qquad \mathbf{A_{23}} = -\begin{vmatrix} 1 & 0 \\ 2 & 3 \end{vmatrix} = -3$$

$$\mathbf{A_{31}} = \begin{vmatrix} 0 & -1 \\ 2 & -2 \end{vmatrix} = 2 \qquad \mathbf{A_{32}} = -\begin{vmatrix} 1 & -1 \\ -2 & -2 \end{vmatrix} = 4 \qquad \mathbf{A_{33}} = \begin{vmatrix} 1 & 0 \\ -2 & 2 \end{vmatrix} = 2$$

A kapott adjungált mátrix (transzponált mátrix):

$$\mathbf{adjA} = \begin{bmatrix} 10 & -3 & 2 \\ 0 & 4 & 4 \\ -10 & -3 & 2 \end{bmatrix}$$

Négyzetes mátrix esetén, ha $|A| \neq 0$, akkor megtudjuk határozni az *inverz vagy reciprok* mátrixt, az A^{-1} mátrixt. Ekkor igazak a következő azonosságok:

$$A \cdot A^{-1} = E \text{ és } A^{-1} \cdot A = E$$

Vagyis akár jobbról vagy balról szorozzuk az *A* mátrixt az inverzével, mindig az *E* egységmátrixt kapjuk eredményül. Az inverz mátrixt a következő módon határozzuk meg, ha $|A| \neq 0$ (ellenkező esetben *A*-nak nem létezik inverze) feltétel esetén:

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

$$\mathbf{A}^{-1} = \frac{\mathbf{adjA}}{|\mathbf{A}|}$$

Mivel az előbbi mátrix determinánsára igaz, hogy $|A| \neq 0$, ezért létezik az inverze. Ismerjük az adjA mátrixt, végezzük el az osztás műveletét, hogy megkapjuk az inverz mátrixt:

$$\mathbf{A}^{-1} = \frac{\mathbf{adjA}}{|\mathbf{A}|} = \frac{\begin{bmatrix} 10 & -3 & 2 \\ 0 & 4 & 4 \\ -10 & -3 & 2 \end{bmatrix}}{20} = \begin{bmatrix} 0.5 & -0.15 & 0.1 \\ 0 & 0.2 & 0.2 \\ -0.5 & -0.15 & 0.1 \end{bmatrix}$$

Önadjungált mátrix: az *A* önadjungált, ha *A**=*A* (lásd szimmetrikus mátrixok).

1.3.4. Sajátérték, sajátvektor

A két fogalom központi helyet foglal el a biostatisztikában. Számos statisztikai módszer alapszik ezeken a számításokon pl. PCA (főkomponens analízis), faktoranalízis. Ezeknél a többváltozós módszereknél az alapmátrix az *R* (korrelációs mátrix). A téma tárgyalása előtt nézzünk meg néhány alapfogalmat:

Skaláris mennyiség (skalármennyiség): olyan mennyiség, amely jellemzésére a számérték is elegendő pl. térfogat, terület, hosszúság stb.

Vektoriális mennyiség (vektormennyiség): olyan mennyiség, amely jellemzésére a számértéken felül a mennyiség irányára és irányítására is szükség van. Ezt megfelelő irányú egyenes szakasszal ábrázoljuk, melyen az irányítást a nyíl jelzi. Vektorok például a rendezett zámpárok, számhármasok stb., azaz a sík- illetve térbeli koordináták. Pl. erő, gravitációs térerősség stb.
Vektor-tér (vagy lineáris tér): a lineáris algebra legalapvetőbb strukturális fogalma. A vektorokkal végezhető műveletek legelemibb tulajdonságait axiomatikusan definiálja. A lineáris tér a mi szokásos síkunk és terünk általánosítása többdimenziós terekre.

Euklideszi tér: azon *T* számtest feletti vektortereket, melyekben a vektorterek axiómáin felül értelmezve van egy ún. skaláris szorzat (euklideszi norma).

Legyen *V* egy vektortér egy *T* test felett (pl. a valós számok halmaza, *R*), és legyen *A* egy *n*-edrendű kvadratikus mátrix, amely a *V* vektortér egy lineáris leképezését adja:

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

$$A: V \rightarrow V$$

és legyen $v \in V$ egy nem nulla tetszőleges vektor ($v=[v_1, v_2, v_3,\ldots,v_n]$). A v vektort az A leképezés *sajátvektorának* nevezzük, ha létezik olyan λ skalárérték ($\lambda=0$ is lehetséges), hogy $\lambda \in T$, és teljesül a következő egyenlőség:

$$A \cdot v = \lambda \cdot v$$

A λ érték az A egy v sajátvektorához tartozó sajátértéke.

Legyen A egy kvadratikus mátrix. A sajátérték egyenlet az előzőek alapján:

$$A \cdot v = \lambda \cdot v$$

Használjuk fel az E egységmátrixt, amely nem változtatja meg az egyenletet:

$$A \cdot v = \lambda \cdot E \cdot v$$

Rendezzük át az egyenletet:

$$A \cdot v - \lambda \cdot E \cdot v = 0$$

ahonnan

$$(A - \lambda \cdot E) \cdot v = 0$$

Az adott *A mátrix karakterisztikus* polinomja (det a determinánst jelöli):

$$P(\lambda) = \det(A - \lambda \cdot E)$$

A polinom fokszáma megegyezik a mátrix rendjével, így legfeljebb n sajátérték lehetséges, amiknek a megkeresése magasrendű mátrixok esetén különösen nehéz. Az alábbi determinánst kifejtve (A karakterisztikus determinánsa), λ-ra pontosan n-edfokú polinomot kapunk, amelynek a gyökei lesznek a keresett sajátértékek:

$$\begin{vmatrix} \mathbf{a}_{11} - \lambda & \mathbf{a}_{12} & . & \mathbf{a}_{1n} \\ \mathbf{a}_{21} & \mathbf{a}_{22} - \lambda & . & \mathbf{a}_{21} \\ . & . & . & . \\ \mathbf{a}_{n1} & \mathbf{a}_{n1} & . & \mathbf{a}_{nn} - \lambda \end{vmatrix} = |\mathbf{A} - \lambda \mathbf{E}| = \mathbf{0}$$

A λ_i-hez tartozó sajátvektorokat a

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

$$(A - \lambda \cdot E) \cdot v = 0$$

egyenlet alapján határozzuk meg.

***Megjegyzés**:*
- A sajátértékek összege, $\lambda_1 + \lambda_2 + \ldots + \lambda_n = Sp(A)$, ami a mátrix nyoma.
- A sajátértékek szorzata, $\lambda_1 \cdot \lambda_2 \cdot \ldots \cdot \lambda_n = \det(A)$.

Példa:

Határozzuk meg az egyik ún. Pauli mátrix sajátértékeit és vektorait.

Megjegyzés: a Pauli mátrixok 2x2-es hermitikus mátrixok, amelyek nyoma 0. Három féle ilyen mátrix van.
A mátrix alakja:

$$[\mathbf{A}] = \begin{bmatrix} \mathbf{0} & \mathbf{1} \\ \mathbf{1} & \mathbf{0} \end{bmatrix}$$

Írjuk fel a karakterisztikus egyenletet:

$$|\mathbf{A} - \lambda \mathbf{E}| = \begin{vmatrix} \mathbf{0} - \lambda & \mathbf{1} \\ \mathbf{1} & \mathbf{0} - \lambda \end{vmatrix} = \lambda^2 - \mathbf{1} = \mathbf{0}$$

A sajátértékek: $\lambda = \pm 1$. A kapott sajátértékek teljesítik a következőket:

$$Sp(A) = 1 + (-1) = 0 \text{ és}$$
$$\det(A) = 1 \cdot (-1) = -1$$

A keresett saját vektorok:

$\lambda_1 = 1$ esetén:

Felhasználva a fenti egyenletet:

$$\begin{pmatrix} -\mathbf{1} & \mathbf{1} \\ \mathbf{1} & -\mathbf{1} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{v}_1^{(1)} \\ \mathbf{v}_2^{(1)} \end{pmatrix} = \mathbf{0}$$

Végezzük el a beszorzást.

$$-v_1^{(1)} + v_2^{(1)} = 0$$
$$v_1^{(1)} - v_2^{(1)} = 0$$

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Amiből a $v_1^{(1)} = v_2^{(1)}$ egyenlőség adódik (a felső index a szóbanforgó sajátértéket jelöli). A keresett vektor alakja:

$$\mathbf{v}^{(1)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Ennek normált alakja ($s = \sqrt{1^2 + 1^2} = \sqrt{2}$ felhasználásával):

$$\mathbf{v}^{(1)} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$

$\lambda_2 = -1$ esetén:

Felhasználva a fenti egyenletet:

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} v_1^{(2)} \\ v_2^{(2)} \end{pmatrix} = \mathbf{0}$$

Végezzük el a beszorzást.

$$-v_1^{(2)} + v_2^{(2)} = 0$$
$$v_1^{(2)} - v_2^{(2)} = 0$$

Amiből a $v_1^{(2)} = v_2^{(2)}$ egyenlőség adódik. A keresett vektor alakja:

$$\mathbf{v}^{(2)} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

Ennek normált alakja ($s = \sqrt{1^2 + 1^2} = \sqrt{2}$ felhasználásával):

$$\mathbf{v}^{(2)} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}$$

A saját vektorok mátrixa tehát:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

A főátlóban a sajátértékek állnak.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

1.4. Fontosabb speciális mátrixok

a) Sormátrix (sorvektor)

Egyetlen sorból álló mátrix:

$$a=[a_{11}, a_{12}, \ldots, a_{1n}]$$

b) Oszlopmátrix (oszlopvektor)

Egyetlen oszlopból álló mátrix:

$$\mathbf{a} = \begin{bmatrix} a_{11} \\ a_{21} \\ \cdot \\ \cdot \\ \cdot \\ a_{m1} \end{bmatrix}$$

c) Zérus-mátrix

Minden eleme 0:

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

d) Diagonál mátrix

Csak a főátlóban lévő elemek nem *0*-ák. Megadási módja

$$A=\langle a_{11}, a_{22}, \ldots., a_{mn}\rangle$$

e) Egységmátrix

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

A főátlóban minden elem *1*, a többi zérus. Megadáskor a rendszámot is feltüntetjük:

$$E_3=\begin{bmatrix}\mathbf{1} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{1}\end{bmatrix}=\langle 1,\,1,\,1\rangle$$

A mátrix egyes oszlopai (sorai) adják az egységvektorokat, pl. az oszlopmátrixok:

$$e_1=\begin{bmatrix}\mathbf{1}\\ \mathbf{0}\\ \mathbf{0}\end{bmatrix} \qquad e_2=\begin{bmatrix}\mathbf{0}\\ \mathbf{1}\\ \mathbf{0}\end{bmatrix} \qquad e_3=\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\\ \mathbf{1}\end{bmatrix}$$

f) Szimmetrikus mátrix

Olyan négyzetes mátrix, amelynek elemei szimmetrikusak (tükörképek) a főátlóra, vagyis $a_{ij}=a_{ji}$. Ilyen pl. a korrelációs mátrix. Az ilyen mátrix azonos a transzponáltjával, azaz $A=A^*$.

$$\mathbf{S}=\begin{bmatrix}\mathbf{1} & \mathbf{-4} & \mathbf{5}\\ \mathbf{-4} & \mathbf{2} & \mathbf{-6}\\ \mathbf{5} & \mathbf{-6} & \mathbf{3}\end{bmatrix}$$

Antiszimmetrikus mátrix esetén nyilván $a_{ij}=-a_{ji}$. Az ilyen mátrix főátlójában csak 0 áll.

g) Háromszögmátrix

A főátló alatt vagy felett csak *0* elem áll. Így megkülönböztetünk alsó és felső háromszögmátrixt. Példa egy felső háromszögmátrixra:

$$H_f=\begin{bmatrix}\mathbf{-2} & \mathbf{4} & \mathbf{-4}\\ \mathbf{0} & \mathbf{3} & \mathbf{2}\\ \mathbf{0} & \mathbf{0} & \mathbf{5}\end{bmatrix}$$

h) Minormátrix

Tetszőleges sor(oka)t és oszlop(oka)t elhagyva a mátrixból kapjuk az A mátrix minormátrixát. Például vegyük a fenti *S* mátrixt. Hagyjuk az első sort és a harmadik oszlopot. A kapott *S* mátrix minormátrixa a következő:

$$\mathbf{S}=\begin{bmatrix}\mathbf{-4} & \mathbf{2}\\ \mathbf{5} & \mathbf{-6}\end{bmatrix}$$

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

i) Konjugált mátrix

Az *A* mátrix elemeinek (komplex számok) konjugálásával kapott mátrix: $\overline{\mathbf{A}}=\left|\overline{\mathbf{a}_{ij}}\right|$. Ha az *A* elemei valós számok, akkor $\overline{\mathbf{a}_{ij}} = \mathbf{a}_{ij}$.

j) Unitér mátrix

A komplex *A* unitér mátrix kvadratikus mátrix, melyre igazak az alábbiak:

$$\mathbf{A} \cdot \overline{\mathbf{A}^*} = \overline{\mathbf{A}^*} \cdot \mathbf{A} = \mathbf{E}$$

Vagyis, ha az *A* mátrixt megszorozzuk a konjugált mátrix transzponáltjával (akár balról vagy jobbról), akkor az *E* egységmátrixt kapjuk eredményül. Továbbá

$$\mathrm{A}^{-1} = \overline{\mathbf{A}^*}$$

A mátrix transzponáltja egyben inverze is.

k) Ortogonális mátrix

Ha az $A = R$ (az unitér mátrix elemei valós számok), akkor $\overline{\mathbf{A}^*} = \mathbf{A}$, és igaz a következő azonosság:

$$A \cdot A^* = A^* \cdot A = E$$

l) *Hipermátrix:* amelynek elemei szintén mátrixok.

2. Kombinatorika

A kombinatorika (kapcsolástan) az elemek csoportosításával foglalkozó önálló tudományága a matematikának. Elsődleges feladata az elemek csoportjainak előállítása, valamint a csoportok számának meghatározása. Az elemek egy elrendezését komplexiónak nevezzük.

Az elemek elrendezésének három legfontosabb fogalma a permutáció, a variáció és a kombináció témaköréhez tartozik.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

2.1. Permutációk

Ha n db egymástól különböző elemünk van és ezeket az elemeket az összes lehetséges módon sorba rendezzük (sorba rakjuk őket), akkor azt mondjuk, hogy az elemeket permutáljuk. Az egyes elrendezések a komplexiók. Ha az elrendezendő elemek mind különbözők, akkor *ismétlés nélküli*, ha az elemek között azonosak is vannak, akkor *ismétléses permutációról* beszélhetünk. Megegyezés szerint az azonos elemek felcserélését nem tekintjük különböző sorrendnek.

Az ismétlés nélküli permutációk száma:

$$P_n = 1 \cdot 2 \cdot 3 \ldots \cdot n = n! \qquad \text{vagy röviden } P_n = n!$$

az n elem 1–től n–ig terjedő egész számok szorzata. Jelölésben n! (ejtsd: n faktoriális), ami az n elem faktoriális értékét jelöli. Megállapodás szerint 0! = 1.

Ismétléses permutációk száma:

$$p_n^{k_1,k_2,k_3,\ldots,k_n} = \frac{n!}{k_1!, k_2!, k_3!, \ldots, k_n!}$$

ahol $k_1,k_2,k_3,\ldots,k_n$ az egymás közt megegyező elemek számát jelöli.

2.2. Variációk

Ha n számú különböző elemből kiválasztunk k($k \leq n$) számú elemet úgy, hogy figyelembe vesszük ezek sorrendjét is, akkor n elem k–ad osztályú variációjáról beszélünk. Az összes variáció számát a

$$V_n^k = \frac{k!}{(n-k)!} = n(n-1) \cdot (n-2) \cdot (n-3) \cdot \ldots \cdot (n-k+1)$$

kifejezés adja.

Ha az n elemből úgy választunk k elemet tartalmazó csoportokat, hogy a csoportban egy elem többször is szerepelhet és az elemek sorrendje is fontos, akkor az n elem k–ad osztályú ismétléses variációját határozzuk meg:

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

$$V_n^{k,i} = n^k$$

A felső indexben az i betű jelöli az ismétléses variációt.

2.3. Kombinációk

Ha az n számú különböző elemből úgy választunk ki k (k≤n) számút minden lehetséges módon, hogy a kiválasztás során a csoportokon belül az elemek sorrendje nem fontos, akkor n elem k–ad osztályú kombinációjáról beszélünk. Az összes lehetséges kiválasztás száma:

$$C_n^k = \binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\cdot(n-2)\cdot(n-3)\cdot\ldots\cdot(n-k+1)}{k(k-1)\ldots 2\cdot 1}$$

Az $\binom{n}{k}$ jelölést úgy olvassuk, hogy “n alatt a k”.

Ha a k elem között egy elem többször is előfordulhat, akkor n elem k–ad osztályú ismétléses kombinációjáról beszélünk. Az összes kiválasztási lehetőségek száma:

$$C_n^{k,i} = \binom{n+k-1}{k}$$

Az Excelben a COMBIN függvénnyel lehet kombinációt számítani.

2.4. Binomiális együtthatók tulajdonságai

Az olyan kifejezéseket amelyek két tagból állnak binomiális kifejezéseknek nevezzük, pl. (a+b) vagy (a–b). Nagyon érdekes tulajdonságot fedezett fel Pascal (1623–1662) francia matematikus a binomok hatványozásával kapcsolatban. Vegyük az (a + b) binom hatványait sorba egészen az 5. hatványig (n = 0,1,2,3,4,5):

$(a + b)^0 = 1$

$(a + b)^1 = a + b$

$(a + b)^2 = a^2 + 2ab + b^2$

$(a + b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$

$(a + b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4$

$(a + b)^5 = a^5 + 5a^4b + 10a^3b^2 + 10a^2b^3 + 5ab^4 + b^5$

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Ha az egyes tagok együtthatóit egymás alá írjuk, akkor az ún. Pascal háromszöget kapjuk, ahol a külső szárak mentén csak 1–es áll. A háromszög belsejében álló bármely szám a közvetlen felette lévő és attól balra álló két szám összege:

n = 0	1					
n = 1	1	1				
n = 2	1	2	1			
n = 3	1	3	3	1		
n = 4	1	4	6	4	1	
n = 5	1	5	10	10	5	1

Pascal–háromszög

A Pascal–háromszög kitöltését tovább lehet folytatni az n értékének megfelelően (az n tetszőleges, nem negatív egész szám). A Pascal–háromszög révén bármely $(a\pm b)^n$ binom kifejtett polinomiális alakját fel lehet írni, mivel az egyes sorok a kívánt polinom tagjainak együtthatóit tartalmazza. Az egyes tagok hatványainak a meghatározása úgy történik, hogy az első tagnak az a–nak a hatványai balról–jobbra 1–gyel csökkennek, n–től 0–ig, a b együttható hatványai balról–jobbra 1–gyel nőnek. (0–tól n–ig). Vegyük figyelembe a hatványozásnál, hogy $a^0 = 1$ és $b^0 = 1$, így ezen tagokat nem is írjuk ki a hatványozás során. Pl. a teljes alak az $(a+b)^2$ kifejezésnél a következő lenne:

$$(a+b)^2 = a^2b^0 + 2a^1b^1 + b^2a^0 = a^2 + 2ab + b^2$$

Vezessük be az $\binom{n}{0} = 1, \quad \binom{n}{n} = 1$ jelöléseket és írjuk fel a Newton–féle binomiális tételt:

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

ahol az $\binom{n}{k}$ együtthatókat binomiális együtthatóknak nevezzük.

A binomiális együtthatókra igaz az alábbi kifejezés:

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

$$\binom{n}{k-1}+\binom{n}{k}=\binom{n+1}{k}$$

A tételt a kifejtett binomiális együtthatókkal is felírhatjuk:

$$(a+b)^n = a^n + \frac{n}{1!}a^{n-1}b + \frac{n(n-1)}{2!}a^{n-2}b^2 + \ldots + x^n$$

A tételnek egy következménye az alábbi kifejezés:

$(1+x)^n \approx 1+nx$ (nx közel van a 0–hoz).

3. Valószínűség-számítás

3.1. Kísérlet, esemény

Véletlen kísérlet: olyan folyamatot, jelenséget értünk, amelynek a kimenetele előre bizonyosan nem mondható meg, de az igen, hogy milyen módon fejeződhet be. Azaz előre tudható, hogy milyen végállapotok lehetnek. A véletlen kísérletet azonos feltételek mellett, függetlenül meg lehet figyelni, akárhányszor végre lehet hajtani.

Esemény: a véletlen kísérlettel kapcsolatos eseménynek nevezünk minden olyan logikai állítást, melynek igaz vagy hamis értéke egyértelműen megállapítható a kísérlet befejezésekor. Az esemény bekövetkezik, ha az állítás igaz értéket kap a kísérlet végén, és nem következik be, ha logikai érték hamis. Jelölésük. A, B stb.

Eseménytér: az elemi események halmaza. Jelölés: **Ω**

Definíció: az A esemény maga után vonja a B eseményt, ha az A esemény bekövetkezéséből a B esemény bekövetkezése is következik. Jelölés: **$A \subseteq B$**

Axióma:

A véletlen kísérlettel kapcsolatos összes események Ω rendszere (eseménytér)

a) $I \in \Omega \quad \Rightarrow O \in \Omega$

b) ha $A \in \Omega \quad \Rightarrow A^C \in \Omega$

c) Ha $A_1, A_2, A_3, \ldots, A_n \in \Omega \Rightarrow \sum_i A_i \in \Omega$

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

3.2. Eseményalgebra

HALMAZOK		ESEMÉNYEK	
Unio:	$A \cup B$	Összeg:	A+B
Metszet:	$A \cap B$	Szorzás:	AB
Komplementer:	A^C	Ellentett esemény:	A^C
Alaphalmaz:	H	Biztos esemény:	I
Üres halmaz:	$\varnothing$	Lehetetlen esemény:	O
Részhalmaz:	$A \subseteq B$	A maga után vonja B-t:	$A \subseteq B$

Egymást kizáró események: ha A és B-re igaz, hogy AB=O

Elemi esemény: a K véletlen kísérlet egy $A \neq O$ eseménye, ha nincs olyan B esemény, amely A-t maga után vonná. Azaz $\forall B$ ($\neq O$ és $\neq A$) olyan, hogy $B \not\subset A$. Az elemi események jelölése ω.

Esemény algebra

Összeadás	Kivonás	Szorzás	Komplementer	Több művelet
A+B=B+A (A+B)+C=A+(B+C) A+A=A A+I=I A+O=A	A-B=AB'	AB=BA (AB)C=A(BC) AO=O AA=A AI=I	$(A^C)^c$ =A I^C=O O^C=I $A+A^C$=I AA^C=O De Morgan: $(A+B)^C=A^CB^C$ $(AB)^C=A^C+B^C$	A(B+C=AB+AC

3.2.1. Teljes esemény rendszer

Az $A_1, A_2, A_3, A_4, \ldots A_n$ események teljes esemény rendszert képeznek, ha

a) $A_1+A_2+A_3+A_4+\ldots+A_n = I$
b) $A_iA_j = O$, ha $i \neq j$ ($i=1, 2, 3, \ldots, n$ és $j=1, 2, 3, \ldots, n$)

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

3.3. Valószínűség fogalma

Valószínűség axiómája

- Adott P: $\Omega \rightarrow [0, 1]$ valószínűségi függvény. A P kielégíti az alábbiakat:
- 1. P(I)=1
- 2. Ha $A_1, A_2, A_3, \ldots A_\infty \in \Omega$, és $A_i A_j = O$ akkor igaz a σ-additívitás (ha $n \neq \infty$, akkor véges additivitás):

$$P\left(\sum_i A_i\right) = \sum_i P\left(A_i\right)$$

ahol

P(I): Biztos esemény valószínűsége
P(O): Lehetetlen esemény valószínűsége

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Kolmogorov-féle valószínűségi mező:

(I, Ω, P)

Valószínűségi alapfogalmak

1. Valószínűség:
Eseményeken értelmezett számértékű függvénymérték.
Jelölésben P(A)=p

Kolmogorov axiómák:

- $0 \leq P(A) \leq 1$
- $P(O)=0$ és $P(I)=1$
- Ha $AB = O \Rightarrow P(A+B) = P(A) + P(B)$

2. Valószínűségszámítás: *klasszikus valószínűségi modell:*

$$P(A) = p = \frac{k}{n} = \frac{\text{kedvező események száma}}{\text{összes események száma}}$$

3. Statisztikai próba (teszt):
A mért adatokon értelmezett függvény.

4. Szignifikancia értelmezése : $p < 0.05$

A valószínűség másik ismert megadási módja a százalékos forma, amikor pl. p = 0.60 helyett 60 %–os esélyt mondunk egy esemény bekövetkezésére. Ha magát az *A* eseményt is jelöljük a valószínűségével együtt, akkor a P(A) jelölést használjuk.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Feltételes valószínűség

$$P(A \mid B) = \frac{P(AB)}{P(B)}$$

Teljes valószínűség tétele

Ha $B_1, B_2, B_3, \ldots\ldots, B_n$ események teljes esemény rendszert alkotnak és $P(B_l) \neq 0$, akkor egy tetszőleges A esemény valószínűsége

$$P(A) = \sum_{i=1}^{N} P(A \mid B_i) \cdot P(B_i)$$

Bayes elmélet

Ha a $B_1, B_2, B_3, \ldots\ldots, B_n$ események teljes esemény rendszert alkotnak és $P(Bi) \neq 0$, valamint egy tetszőleges A eseményre igaz, hogy $P(A) \neq 0$, akkor a B_i eseményekre igaz

posteriori valószínűség

$$P(B_i \mid A) = \frac{P(A \mid B_i) \cdot P(B_i)}{\sum_{k=1}^{N} P(A \mid B_k) \cdot P(B_k)}$$

a priori valószínűség

Markov-egyenlőtlenség

- Legyen ξ pozitív valószínűségi változó véges $M(\xi)$ várható értékkel. Ekkor tetszőleges $\lambda > 0$ valós számra igaz az alábbi egyenlőtlenség:

$$P(\xi \geq \lambda \cdot M(\xi)) \leq \frac{1}{\lambda}$$

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Csebisev-egyenlőtlenség

- Legyen ξ tetszőleges valószínűségi változó, melynek van szórása. Ekkor $\varepsilon > 0$ esetén igaz:

$$\mathbf{P}(|\xi - \mathbf{M}(\xi)| \geq \varepsilon) \leq \frac{\mathbf{D}^2(\xi)}{\varepsilon^2}$$

- Ha ξ ismeretlen (várható érték és szórás igen), akkor felső korlátot tudunk megadni a várható érték körüli szimmetrikus intervallumokba esés valószínűségeire.

Nagy számok Bernoulli-féle gyenge törvénye

- Legyen ξ binomiális eloszlású valószínűségi változó, mely $x_k=k$ $(k=0, 1, 1, \ldots,n)$ értéket vesz fel, ha az A esemény az n kísérlet során k-szor következett be. Legyen $\frac{\mathbf{k}}{\mathbf{n}}$ az *A* esemény relatív gyakorisága, P(A) = p az esemény valószínűsége.
- Ekkor $\varepsilon > 0$ esetén igaz: $\mathbf{q} = \mathbf{1} - \mathbf{p} = \mathbf{P}(\overline{\mathbf{A}})$

$$P\left(\left|\frac{k}{n} - p\right| \geq \varepsilon\right) \leq \frac{p \cdot q}{\varepsilon^2 \cdot n} \qquad P\left(\left|\frac{k}{n} - p\right| < \varepsilon\right) \geq 1 - \frac{p \cdot q}{\varepsilon^2 \cdot n}$$

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Nagyszámok gyenge és Erős törvényei

$$\lim_{n\to\infty} P\left(\left|\overline{X}_n - \mu\right| < \varepsilon\right) = 1$$

$$P\left(\lim_{n\to\infty} \overline{X}_n = \mu\right) = 1$$

3.3.1. Valószínűségi változók jellemzése

A valószínűségi változó egy olyan függvény, amely az eseménytér elemeihez valós számokat rendel:

$$\xi: \Omega \to R$$

Valószínűségi változó: ha az elemi események mindegyikéhez egyértelműen hozzárendelünk egy számot, akkor az eseménytéren egy függvényt értelmezünk, és ezzel megadunk egy valószínűségi változót.

- *Diszkrét eloszlások:* értékkészletük megszámlálhatóan véges vagy ∞.

  Eloszlásfüggvénye:

A ξ diszkrét valószínűségi változó F(x) eloszlás *lépcsős* függvénye:

$$F(x) = P(\xi < x) = \sum_{k<x} p_k$$

Az F(x) eloszlásfüggvény tulajdonságai:

— balról folytonos,

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

— monoton növekedő,

— értéke 0 és l közötti.

• *Folytonos eloszlások:* értékkészletük megszámlálhatatlanul ∞.

Sűrűségfüggvénye: a ξ adateloszlását, sűrűségét jellemző folytonos függvényi. Jelölése: *f(x)*

Eloszlásfüggvénye: F(x) = P(ξ < x) és értékkészlete a [0 , 1] közötti intervallum.

Grafikonja folytonos: *görbe*

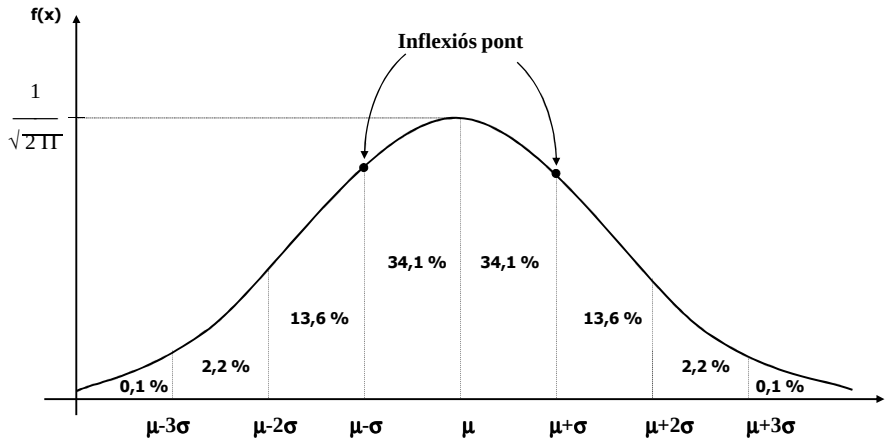
A sűrűségfv. "görbe alatti területét" egy [- , x] intervallumban az eloszlásfv. adja meg. *x*

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

A sűrűségfüggvény tulajdonsága, hogy

— értéke ≥ 0 (hiszen a valószínűség nem lehet negatív értékű),

— a függvény görbe alatti területe = l.



Normáleloszlás sűrűségfüggvénye

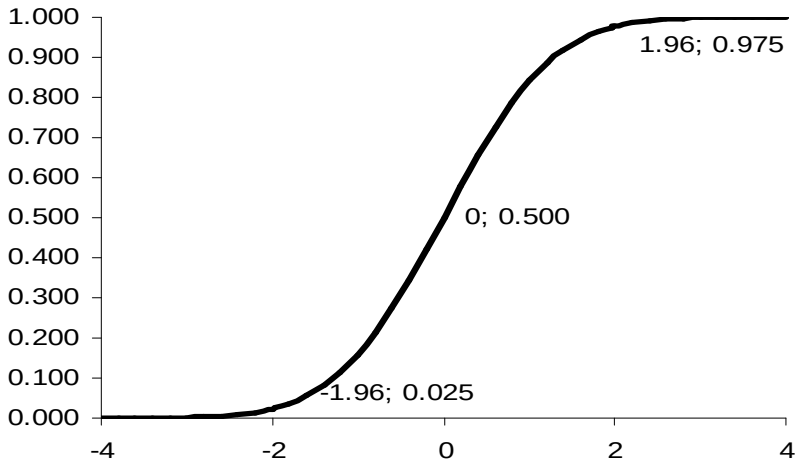
Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Sűrűségfüggvénye

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

Normáleloszlás eloszlásfüggvénye



Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Eloszlásfüggvénye

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

3.3.2. Valószínűségi változók várható értéke és szórása

Várható érték (M(ξ)): az a szám, amely körül megfigyelt értékeinek átlaga ingadozik.

- *Diszkrét esetben:*

$$\mathrm{M}(\xi) = \sum_{k=1}^{n} \mathrm{p}_k \mathrm{x}_k$$

- *Folytonos esetben:*

$$M(\xi) = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

Szórás (D(ξ)): a ξ várható értékétől való átlagos eltérését jellemzi.

Négyzete a variancia: $V(\xi) = D^2(\xi)$

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

A szórásnégyzet:

$$\mathrm{Var}(\xi) = D^2(\xi) = M[(\xi - M(\xi)]^2 = M(\xi^2) - M^2(\xi)$$

• Diszkrét esetben:

$$\mathrm{Var}(\xi) = D^2(\xi) = \sum_{k=1}^{n} p_k x_k^2 - \left(\sum_{k=1}^{n} p_k x_k \right)^2$$

• Folytonos esetben:

$$D^2(\xi) = \int_{-\infty}^{\infty} x^2 \cdot f(x)dx - \left(\int_{-\infty}^{\infty} x \cdot f(x)dx \right)^2$$

3.4. Eloszlások

3.4.1. Nevezetes diszkrét eloszlások

3.4.1.1.Binomiális eloszlás

Végezzünk el egy kísérletet *n*–szer egymástól függetlenül. A kísérlet során egy ***A*** esemény bekövetkezésének valószínűsége legyen P(A) és az ellentett esemény valószínűsége pedig $P(\overline{A}) = q = 1-p$. A *p*–ről feltesszük, hogy konstans a kísérlet folyamán. A ξ valószínűségi változó az ***A*** esemény bekövetkezéseinek a számát jelenti. Ekkor annak valószínűsége, hogy a kísérlet során az *A* esemény *k*–szor következik be a következő alakban adható meg:

$$p_k = P(\xi = k) = \binom{n}{k} p^k \cdot q^{n-k} \qquad (k = 0, 1, 2, ..., n)$$

A ξ valószínűségi változó eloszlását binomiális eloszlásnak nevezzük, amelynek várható értéke:

$$M(\xi) = n \cdot p$$

és szórása:

$$D(\xi) = \sqrt{n \cdot p \cdot q}$$

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

3.4.1.2.Hipergeometrikus eloszlás

Az N számú elemből jelöljünk meg M darabot. Random módon visszatevés nélkül válasszunk ki n darabot az N számú elemből úgy, hogy teljesüljön a választásra az $n \leq M$ és $n \leq N–M$ feltétel. Jelölje ξ azoknak a megjelölt elemeknek a számát, amelyek az n kiválasztott elemek között előfordulnak. Ekkor ξ értékeire az alábbi valószínűségek adódnak.

$$p_k = P(\xi = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \qquad (k = 0,1,2, .., k)$$

A ξ valószínűségi változó eloszlását *hipergeometrikus* eloszlásnak nevezzük. Az eloszlás várható értéke és szórása:

$$M(\xi) = n \cdot p$$

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

3.4.1.3. Negatív binomiális eloszlás

Végezzünk el több egymástól független kísérletsorozatot, amelyben egy A esemény valószínűsége $P(A)$ konstans a kísérlet folyamán és az ellentett esemény valószínűsége a $P(\overline{A}) = 1 - p$. Legyen r egy természetes szám és ξ olyan valószínűségi változó, amely – ha az A esemény r–szer éppen az r+k–adik kísérletben következik be – az x_k = k+r értéket veszi fel. Nyilván az ezt megelőző kísérletekben az A esemény r–1–szer, az $\overline{A}$ esemény k–szor következik be. Ekkor annak valószínűsége, hogy az A esemény a $k+r$ kísérletsorozatban r–szer következik be.

$$p_k = P(\xi = x_k) = \binom{k+r-1}{r-1} p^r (1-p)^k \qquad (k = 0,1,2,...)$$

A ξ eloszlását r–ed rendű negatív binomiális eloszlásnak nevezzük. Az eloszlás várható értéke és szórása:

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

$$M(\xi) = \frac{r}{p}$$

$$D(\xi) = \frac{\sqrt{r(1-p)}}{p}$$

3.4.1.4. Poisson–eloszlás

A

$$p_k = P(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda} \qquad (k = 0, 1, 2, ...)$$

eloszlást a ξ valószínűségi változó Poisson–eloszlásának nevezzük, ahol $\lambda>0$ egy tetszőleges valós szám. Poisson eloszlást követnek pl. adott idő alatt lejátszódó események száma, baktériumok, sejtek száma.egy adott téfogatban, balesetek száma egy időintervallumban, stb.

A Poisson–eloszlás és a binomiális eloszlás között szoros a kapcsolat. Ha a binomiális eloszlásban n nagy és a vizsgált esemény valószínűsége a p értéke 0–hoz közeli érték (az $n \cdot p$ szorzat értéke < 5), ilyenkor a $\lambda = n \cdot p$ választással a binomiális eloszlás jól közelíthető a Poisson–eloszlással:

$$\binom{n}{k} p^k q^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda}$$

A Poisson–eloszlás várható értéke és szórása:

$$M(\xi) = \lambda$$

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

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

3.4.2. Nevezetes folytonos eloszlások

3.4.2.1. Egyenletes eloszlás

Az egyenletes eloszlás *sűrűségfüggvénye:*

$$f(x) = \begin{cases} 0 & ha \quad x \le a \\ \dfrac{1}{b-a} & ha \quad a < x \le b \\ 0 & ha \quad x > b \end{cases}$$

Eloszlásfüggvénye:

$$F(x) = P(\xi<x) = \begin{cases} 0 & ha \quad x \le a \\ \dfrac{x-a}{b-a} & ha \quad a < x \le b \\ 1 & ha \quad x > b \end{cases}$$

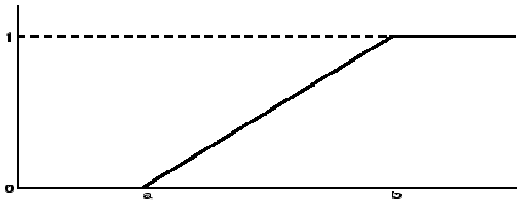
A várható érték és szórás:

$$M(\xi) = \frac{a+b}{2}$$

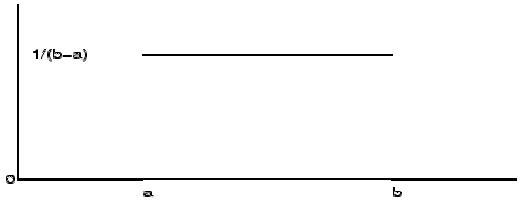
$$D(\xi) = \frac{b-a}{\sqrt{12}}$$

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.



Az egyenletes eloszlás eloszlásfüggvénye



Az egyenletes eloszlás sűrűségfüggvénye

3.4.2.2.Exponenciális eloszlás

Az exponenciális eloszlás *sűrűségfüggvénye:*

$$f(x) = \begin{cases} 0 & ha \quad x \le 0 \\ \lambda e^{-\lambda x} & ha \quad x > 0 \end{cases}$$

ahol x>0 tetszőleges pozitív szám.

Az exponenciális *eloszlásfüggvény alakja*

$$F(x) = P(\xi<x) = \begin{cases} 0 & ha \quad x \le 0 \\ 1 - e^{-\lambda x} & ha \quad x > 0 \end{cases}$$

A várható érték és szórás:

$$M(\xi) = \frac{1}{\lambda}$$

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

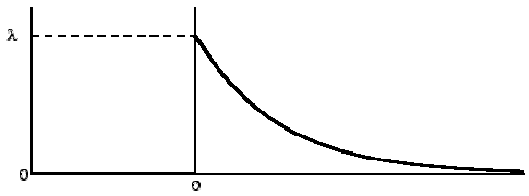
Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Exponenciális eloszlás

Az exponenciális eloszlás sűrűségfüggvénye:

$$f(x) = \begin{cases} 0, & x \le 0, \\ \lambda e^{-\lambda x}, & x > 0. \end{cases}$$

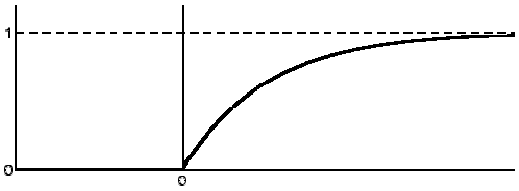


Exponenciális eloszlás

A
ξ
valószínűségi változót
λ
paraméterű **exponenciális eloszlásúnak** nevezzük, ha eloszlásfüggvénye:

$$F(x) = \begin{cases} 0, & x \le 0, \\ 1 - e^{-\lambda x}, & x > 0. \end{cases}$$

ahol $\lambda > 0$ rögzített



Az exponenciális eloszlásfüggvény

Az exponenciális eloszlás általánosított alakja a Weibull–eloszlás, amelynek sűrűségfüggvénye ($c > 0$ és $\alpha > 0$ állandók):

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

$$f(x) = \begin{cases} c \cdot \alpha \cdot x^{\alpha-1} \cdot e^{-c \cdot x^{\alpha}} & ha \quad x \geq 0 \\ 0 & ha \quad x > 0 \end{cases}$$

Eloszlásfüggvénye:

$$F(x) = \begin{cases} 1 - e^{-c \cdot x^{\alpha}} & ha \quad x \geq 0 \\ 0 & ha \quad x < 0 \end{cases}$$

A Weibull–eloszlás egyik sajátságos felhasználási területe a gyógyszerkinetikai vizsgálatok.

3.4.2.3.Gamma–eloszlás

A ξ valószínűségi változó λ paraméterű, Γ–edrendű λ–eloszlás sűrűségfüggvénye az alábbi formában adható meg:

$$f(x) = \begin{cases} \frac{\lambda^p \cdot x^{p-1}}{\Gamma(p)} e^{-\lambda x} & ha \quad x \geq 0 \\ 0 & ha \quad x < 0 \end{cases}$$

ahol $\lambda>0$ és $p>0$ állandók. Ha p egész szám, akkor:

$$\Gamma(p) = (p-1)!$$

A várható érték és szórás:

$$M(\xi) = \frac{p}{\lambda}$$

$$D(\xi) = \frac{p}{\lambda^2}$$

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

3.4.2.4.Béta–eloszlás

A ξ valószínűségi változó (p, q)–rendű béta–eloszlású, ha sűrűségfüggvénye:

$$f(x) = \begin{cases} \dfrac{\Gamma(p+q)}{\Gamma(p)\cdot\Gamma(q)} x^{p-1}(1-x)^{q-1} & ha \quad 0 < x < 1 \\ 0 & \text{egyébként} \end{cases}$$

ahol p > 0 és q > 0 állandók.

Az eloszlás várható értéke és szórása:

$$M(\xi) = \frac{p}{p+q}$$

$$D(\xi) = \frac{1}{p+q}\sqrt{\frac{p\cdot q}{p+q+1}}$$

A szabadságfok fogalmát Sir R.A. Fisher vezette be. Egy statisztika szabadságfokát – amelyet *df*–el (degrees of freedom) jelölünk a továbbiakban –, úgy definiáljuk, hogy az *N* mintaszámból levonjuk az adott statisztika kiszámításhoz szükséges, az adatokból már meghatározott paraméterek k számát.

$$df = N - k$$

3.4.2.5.F–eloszlás

Legyen az összehasonlítani kívánt két minta normális eloszlású, elemszámuk N_1 és N_2, az egyes populációk varianciája (szórásnégyzete) σ_1^2 és σ_2^2. Az F–statisztikát a következőképpen definiáljuk:

$$F = \frac{\dfrac{N_1 s_1^2}{(N_1-1)\sigma_1^2}}{\dfrac{N_2 s_2^2}{(N_2-1)\sigma_2^2}}$$

ahol s_1^2 és s_2^2 a mintákból számolt korrigált varianciák (lásd később). Az eloszlás szabadságfokai:

$$df_1 = N_1 - 1 \quad \text{és} \quad df_2 = N_2 - 1$$

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Az eloszlás sűrűségfüggvénye:

$$f_{f1,\,f2} = \frac{K \cdot F^{\frac{f_1}{2}-1}}{\left(f_1 \cdot F + f_2\right)^{\frac{f_1+f_2}{2}}}$$

ahol K a df_1 és df_2 szabadságfokoktól függő konstansérték.

Az eloszlás görbe alatti területe 1. Az eloszlás alakja az df_1 és df_2 értékektől függ.

Az F–eloszlás várható értéke:

$$M = \frac{N_2}{N_2 - 2} \qquad \text{ha } N_2 \geq 3$$

és szórása

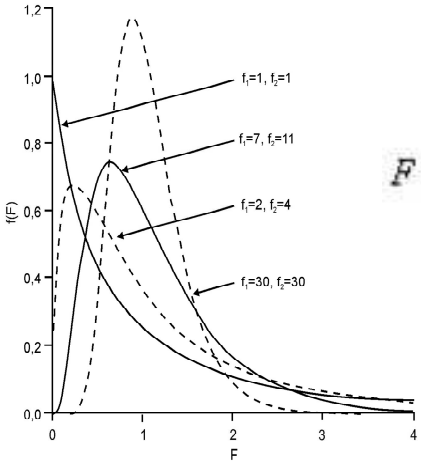
$$D^2 = \frac{2N_2^2\left(N_1 + N_2 - 2\right)}{N_1\left(N_2 - 2\right)^2\left(N_2 - 4\right)} \qquad (\text{ha } N_2 \geq 5)$$

F eloszlás

$$F = \frac{s_1^2}{s_2^2}$$

f1 = (n1-1)

f2 = (n2-1)



$$F = \frac{s_1^2 / \sigma_1^2}{s_2^2 / \sigma_2^2}$$

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

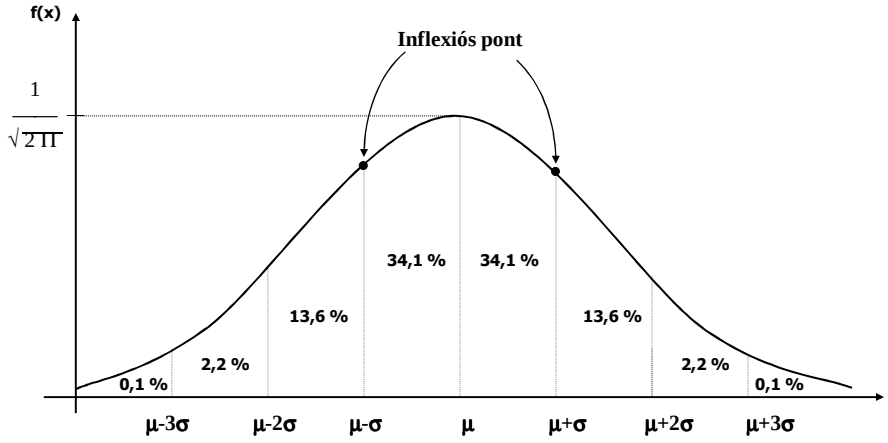
MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

3.4.2.6. Normális eloszlás

Általános jelölése: N(μ, σ). Az eloszlást Gauss-görbének vagy harang görbének is hívjuk.

Sűrűségfüggvény

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



Normális eloszlás tulajdonságai

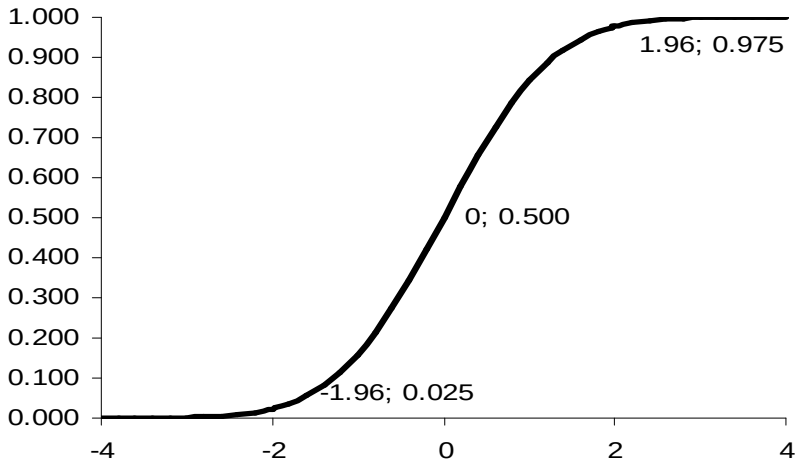
Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Eloszlásfüggvény

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

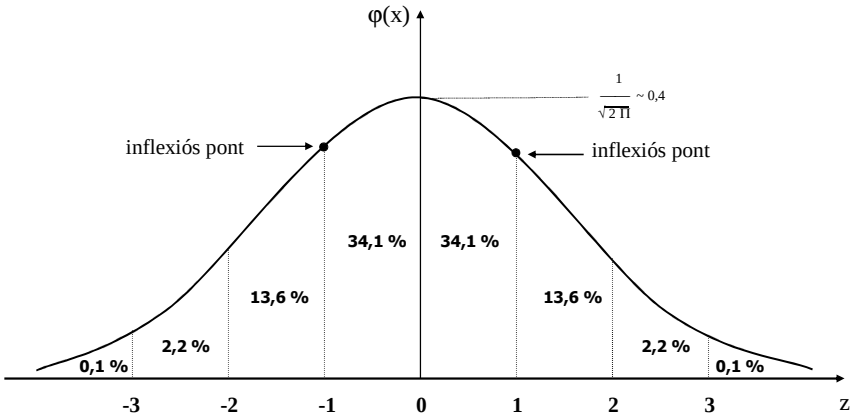
Normáleloszlás eloszlásfüggvénye



Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Standard normális eloszlás jelölése: N(0, 1)



Standard normális eloszlás

$$z_i = \frac{x_i - \mu}{\sigma}$$

z_i a transzformációs képlet, amely segítségével tetszőleges normális eloszlást standard normális eloszlásba (egyetlen ilyen alak van) transzformálhatunk.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Standard normális eloszlás
sűrűségfüggvénye

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

Standard normális eloszlás
eloszlásfüggvénye

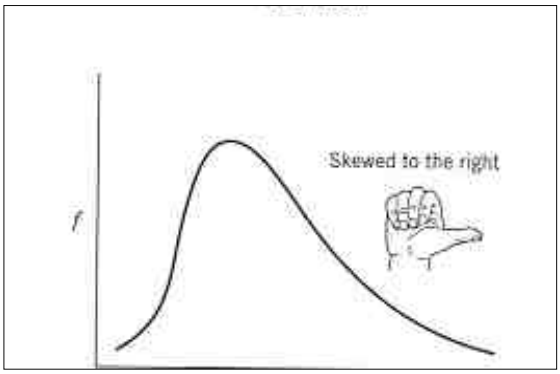
$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{x^2}{2}} dx$$

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

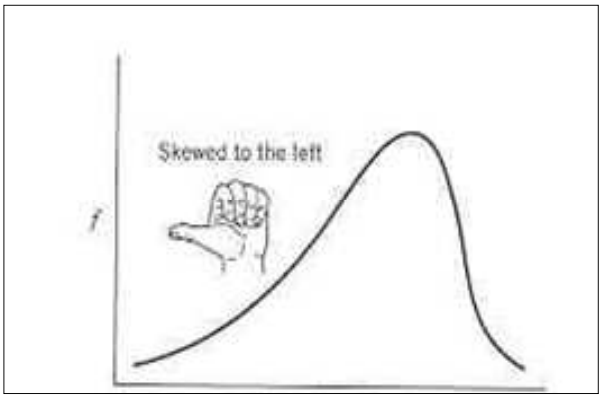
MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió
támogatásával valósulnak meg.

Aszimmetrikus normális eloszlások:

POSITIVELY SKEWED



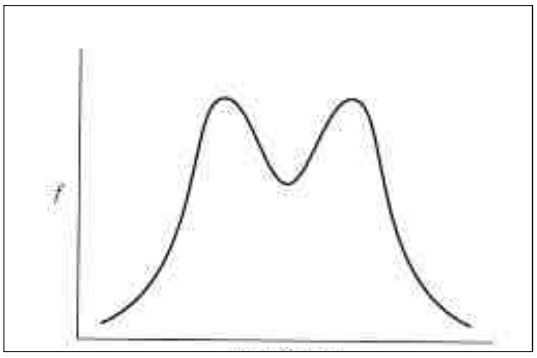
NEGATIVELY SKEWED



Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

BI-MODAL



3.4.2.6.1. Normális eloszlás aszimmetria mutatói

Pearson-féle A mutató:

A mérőszám (önmagában a számláló) előjele az aszimmetria irányát mutatja. Bal oldali, jobbra elnyúló aszimmetria esetén $A > 0$, jobb oldali, balra elnyúló aszimmetria esetén $A < 0$. Szimmetrikus eloszlás esetén $A = 0$. A mérőszám abszolút értékének nincs határozott felső korlátja, azonban már 1-nél nagyobb abszolút érték a gyakorlatban ritkán fordul elő és meglehetősen erős aszimmetriára utal.

$$A = \frac{\bar{x} - Mo}{\sigma}$$

F- mutató:

E mutatószám ugyanolyan feltételek mellett ad nulla, pozitív és negatív eredményt, mint az *A* mutató. Az *F* mutató lényegesen kisebb értékkel jelzi a már nagyfokúnak tekinthető aszimmetriát, mint az *A*.

$$F = \frac{(Q_3 - Me) - (Me - Q_1)}{(Q_3 - Me) + (Me - Q_1)}$$

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Kurtosis: a görbe csúcsosságát jellemzi. Pozitív érték esetén csúcsosabb, negatív érték esetén lapultabb a görbe. Értéke lehetőleg legyen *0* vagy *0* közeli.

Skewness: a szimmetria tengelytől való eltolás mértékét jellemzi. Pozitív érték esetén jobbra, negatív érték esetén balra eltolt az eloszlás. Értéke lehetőleg legyen *0* vagy *0* közeli.

Ha *mindkét érték* egyszerre *0* vagy *0* közeli, akkor az eloszlás *normális*.

3.4.2.7.2. Inverz normális–eloszlás (vagy Wald):

$$\sqrt{\frac{\lambda}{2\pi x^3}}\exp\left\{-\frac{\lambda}{2\mu^2 x}(x-\mu)^2\right\}$$

Nagyon sok hasonlóságot mutat a normális eloszláshoz. Balra eltolt eloszlások esetén használatos.

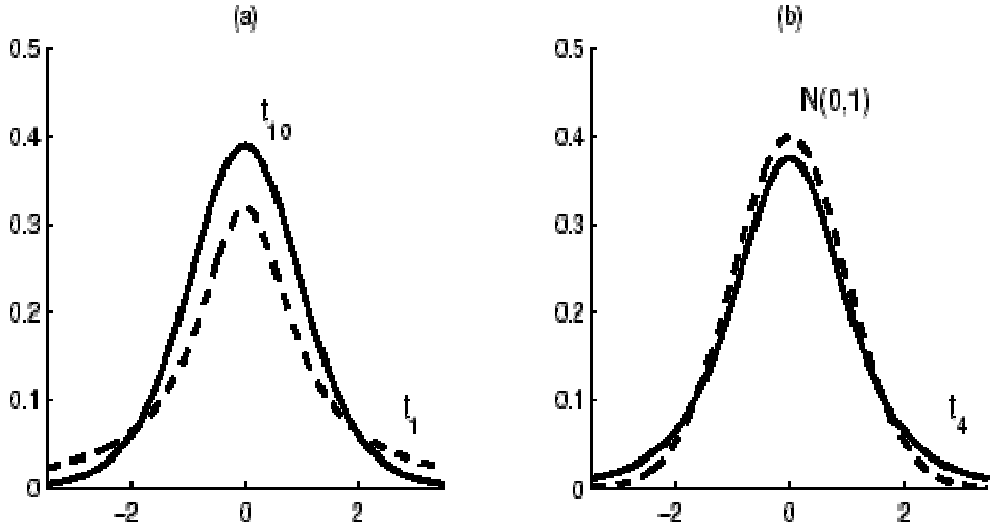
3.4.2.7. t-eloszlás

Az ξ valószínűségi változót ***n*** szabadsági fokú **Student-eloszlásúnak** (**t**-eloszlásúnak vagy $\mathbf{t_n}$-eloszlásúnak) nevezzük, ha sűrűségfüggvénye:

$$f(x)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi}\sqrt{n}\Gamma\left(\frac{n}{2}\right)\left(1+\frac{x^2}{n}\right)^{\frac{n+1}{2}}},$$

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.



Látható, hogy fenti sűrűségfüggvény a ***0***-ra szimmetrikus: **n=1** szabadsági fok esetén a Student-eloszlás a (**λ=1**, μ=0**)** paraméterű Cauchy-eloszlás.

3.4.2.8. Lognormális eloszlás

Egy ξvalószínűségi változó lognormális eloszlású, ha a változó logaritmusa:

$$\varphi = \ln \xi$$

normális eloszlású.

Az eloszlás sűrűségfüggvénye:

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\ln x - m)^2}{2\sigma^2}} & \text{ha} \quad x > 0 \\ 0 & \text{ha} \quad x \le 0 \end{cases}$$

Sűrűségfüggvénye:

$$F(x) = \Phi\left(\frac{\ln(x)}{\sigma}\right) \qquad x \ge 0; \sigma > 0$$

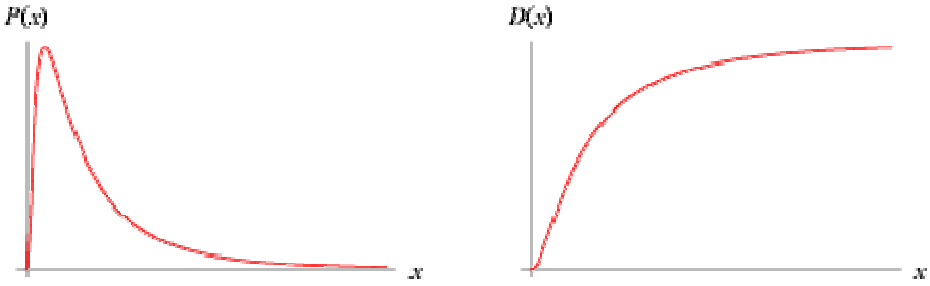
Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Az eloszlás várható értéke és szórásnégyzete:

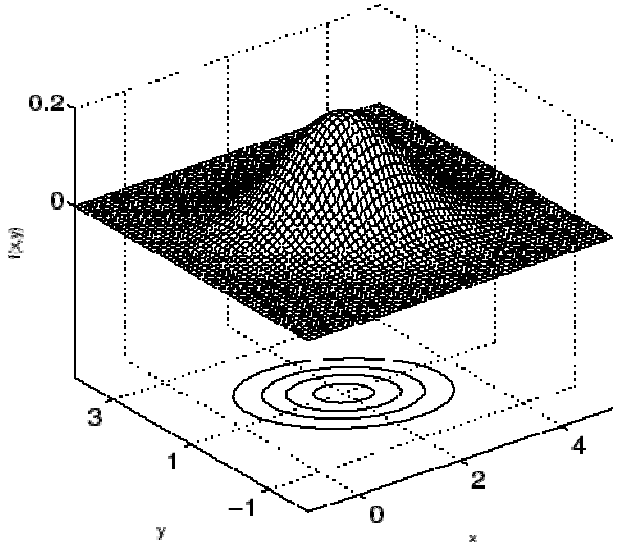
$$M(\xi) = e^{m+\frac{\sigma^2}{2}}$$

$$D^2(\xi) = e^{2m+\sigma^2}\left(e^{\sigma^2} - 1\right)$$



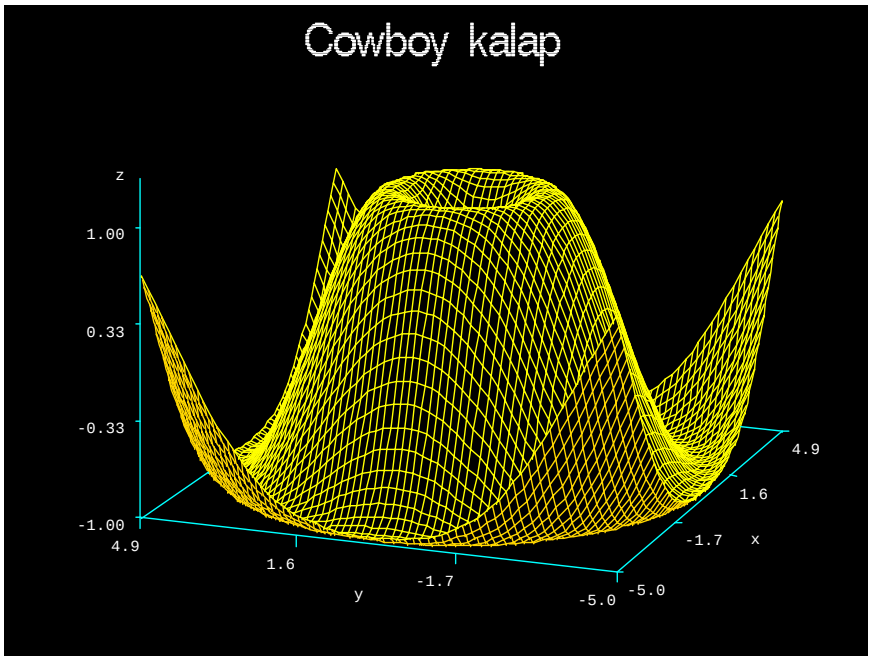
3.4.2.9. Érdekes eloszlások

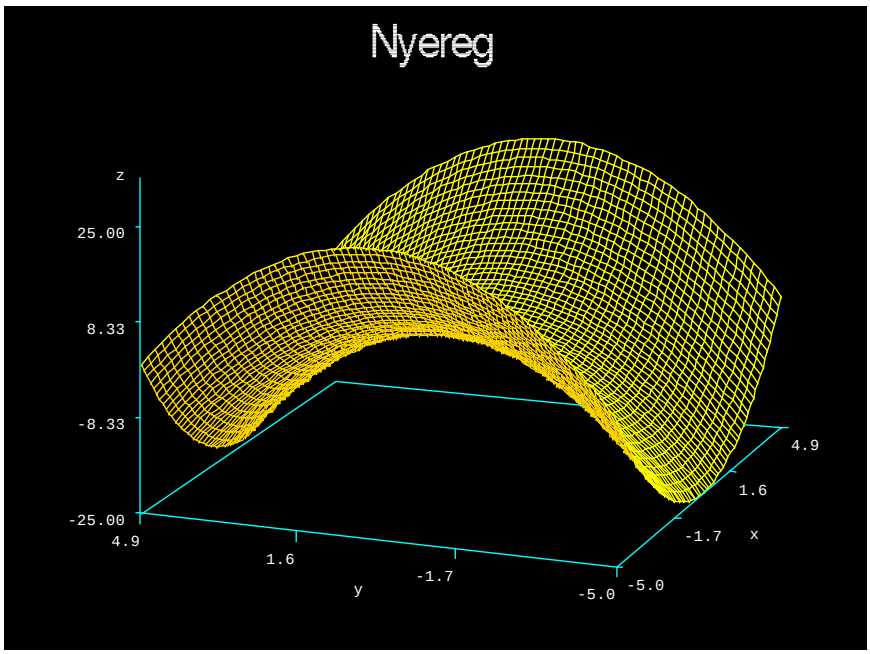
3 dimenziós normális eloszlás sűrűségfüggvénye



Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.





Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

4. Adattípusok

A statisztikában egy ξ változó mérésének a skálája olyan osztályozást jelent, amely lehetővé teszi a változón különböző matematikai műveletek végrehajtását.

A megjelenítés módszerét egyrészt a megfigyelt ξ változó természete (diszkrét vagy folytonos valószínűségi változó), illetve a vizsgálat célja határozza meg. Ennek megfelelően a következő négy fontosabb skálatípust különböztetjük meg, megjegyezvén, hogy minden következő skálatípus örökli a felette lévő műveleti tulajdonságait illetve újabbakkal bővülnek:

4.1. Nominális skála

A legegyszerűbb skálatípus, ahol a mérés eredményei között csak az egyenlőséget és a nem egyenlőséget tudjuk definiálni. A statisztikai vizsgálat eredményeit osztályokra, kategóriákra osztjuk. A nominális adatok nem számszerűsíthetőek, és így a legtöbb tárgyalt statisztikai művelet nem használható velük kapcsolatban. A skálaértékeket pusztán kódszámoknak tekintjük, amelyek között semmilyen matematikai viszonyt nem feltételezünk pl. nem=1 (férfi) és nem=2 (nő). A nominális skála esetében a skálaérték előfordulásának gyakorisága (modusz) vizsgálható, vagy kontingenciatábla is készíthető, azonban sem medián, sem átlag használatának nincs értelme a nominális skálánál.

ξ -n értelmezhető műveletek: =, ≠

4.2. Ordinális skála

Az ordinális (rendezett) adatokról nem csak egyezőségüket állapíthatjuk meg, hanem valamilyen elv szerint sorba is rendezhetjük őket. Az iskolai osztályzatok tipikus ordinális skálájú adatok. Megállapítható, hogy egy négyesnél jobb az ötös, de nem mondható, hogy a hármas és a négyes között ugyanakkora a tudáskülönbség, mint a négyes és az ötös között. Továbbá nem igaz, hogy egy négyes kétszer jobb, mint egy kettes (sem az, hogy fele annyit tud). Szintén ordinális pl. a dohányzás mértéke (nem, mérsékelt, erős dohányos). A legtöbb ordinális skálán mért adatot elvileg arány vagy intervallum skálán is mérhetnénk, de ezt valamilyen okból nem tesszük (például jegyek helyett a szerzett pontok jobban tükröznék az iskolai teljesítményt).
E skálatípus esetében a medián vizsgálható, az átlag használatának ellenben itt nincs értelme. Ordinális adatok esetében általában a nem paraméteres statisztikákat kell alkalmaznunk.

ξ -n értelmezhető műveletek: =, ≠, <, >

4.3. Intervallum skála

Az intervallum skálánál az egyes értékek közötti különbség azonos, de mivel nincs eleve adott *0* pontjuk, így arányaiknak sincs értelme. A számértékek mind a nagyság szerinti viszonyokat

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

megmutatják, mind az eltérés mértékét meghatározzák, a skálaértékek különbségét itt már értelmezni tudjuk. Legismertebb intervallumskála a Celsius-fok skála vagy Fahrenheit skála. Igaz, hogy a 20 °C és a 22 °C közötti különbség azonos a 32 °C és 34 °C közötti különbséggel. Azonban nem igaz, hogy a 10 °C kétszer olyan meleg, mint az 5 °C. Intervallum skálán adjuk meg a dátumokats vagy az IQ értéket is. Az intervallumskála nullapontjának és egységpontjának a meghatározása is megállapodás kérdése. Itt már számolhatunk átlagot, mivel a nullapont eltolása nem változtatja meg az átlag relatív helyét az átlagolt számok között.

ξ -n értelmezhető műveletek: =, ≠, <, >, –

4.4. Arány skála

Az arányskála az intervallumskála jellemzőivel rendelkezik, emellett tartalmaz egy abszolút nullapontot is. Az arányskálára igaz, hogy az értékek arányának jelentése van, például a 6 kg-os cukroszacskó kétszer annyi tömegű, mint egy 3 kg-os. Ehhez az kell, hogy legyen a skálának nulla pontja, és ezen nulla pont ne önkényes legyen. Magasságméréseknél a nullapont a *0* magassághoz tartozik, ugyanígy tömegmérésnél a *0* tömeghez. A Kelvin hőmérsékletsálának *0* pontja is adott, nem úgy a Celsius skála 0 pontja, amelyek önkényesen választottak (pl. víz fagyáspontja). A legtöbb mért adatunk aránysálán mért, a legtöbb itt tárgyalt statisztika alkalmazható arányskálára.

ξ -n értelmezhető műveletek: =, ≠, <, >, –, /

5. Adatredukció

Azt az eljárást, amelynek során az adatokból olyan számértékeket (paramétereket), statisztikai mutatókat határozunk meg, amelyek az adatok statisztikai viselkedését jól jellemzik, statisztikai *redukciónak* nevezzük. Az eljárás révén az adatok jellemzőit egyetlen számértékbe tömörítjük.

5.1. Középérték

Mi a középérték: azonos fajta számszerű adatok közös jellemzője.

Követelmények:

a) **Számított középérték:** közbenső helyet foglaljanak el az adatok között , azaz

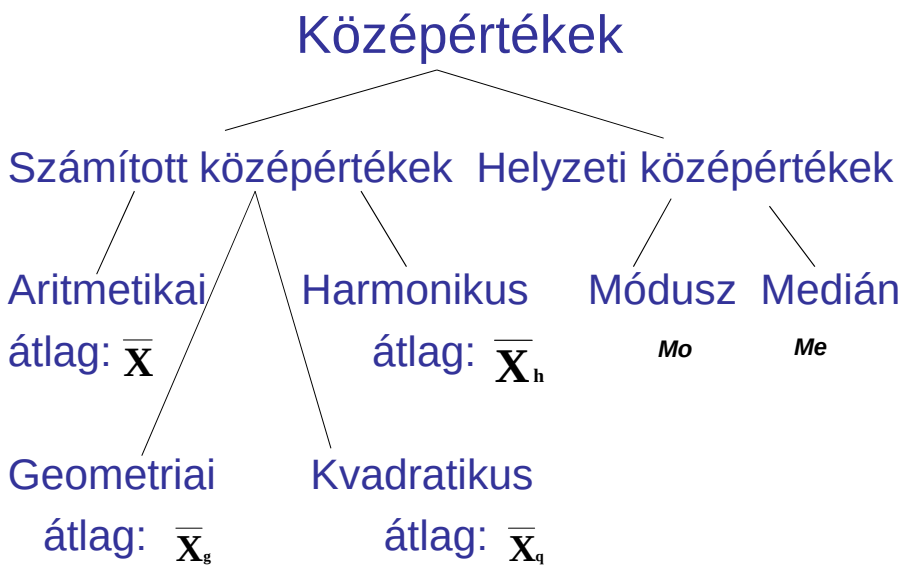
Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

$$x_{min} \leq \text{középérték} \leq x_{max}$$

b) **Helyzeti középérték:** tipikus értékek legyenek (az adatok között gyakran forduljon elő).
c) Legyenek könnyen meghatározhatók és egyértelműen definiálva.

Középérték fajták:



5.1.1. Számított középértékek

5.1.1.1. Aritmetikai átlag (Számtani átlag)

Az a szám, amelyet az átlagolandó értékek helyébe téve azok összege nem változik

$$\sum_{i=1}^{N} x_i = x_1 + x_2 + x_3 + \ldots + x_N = \bar{x} + \bar{x} + \bar{x} + \ldots + \bar{x} = \sum_{i=1}^{N} \bar{x}_i = N \cdot \bar{x} = N \cdot \frac{\sum_{i=1}^{N} x_i}{N} = \sum_{i=1}^{N} x_i$$

Súlyozott számtani átlag

A mért értékek között egyes értékek többször is előfordulnak változó gyakoriságokkal. Ebben az esetben a számtani átlag meghatározásának módja

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

$$\bar{x} = \frac{\sum_{i=1}^{N} f_i x_i}{\sum_{i=1}^{N} f_i}$$

ahol f_i az egyes értékek gyakorisága és $\sum_{i=1}^{N} f_i = N$.

Az aritmetikai átlag általános formája

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{N} = \frac{\sum_{i=1}^{N} x_i}{N}$$

Megjegyzések:

a) Az általánosan elfogadott szokás az átlag értékének megadására, hogy jegyeinek száma egy értékkel legyen több, mint a mért adatok jegyeinek száma.

b) Az átlagtól való eltérések algebrai összege 0

$$\sum_{i=1}^{N} (x_i - \bar{x}) = 0$$

mert a $\sum$-kra vonatkozó azonosságokat felhasználva irható

$$\sum_{i=1}^{N} (x_i - \bar{x}) = \sum_{i=1}^{N} x_i - \sum_{i=1}^{N} \bar{x} = \sum_{i=1}^{N} x_i - \sum_{i=1}^{N} x_i = 0$$

c) Hiányzó értékek esetén (ha számuk nem nagy), ha ezeket az értékeket az adatok átlagával helyettesítjük, akkor a helyettesítéssel elkövetett hibák négyzetösszege a minimális lesz.

d) Ha egy minta két (vagy több) részmintából állítható elő, akkor a teljes minta átlagára igaz

$$\bar{x} = \frac{N_1 \bar{x}_1 + N_2 \bar{x}_2}{N}$$

ahol N_1 és x_1 az első minta, N_2 és x_2 a második minta nagysága és átlaga, N az egyesített minta nagysága ($N = N_1 + N_2$).

e) A számtani átlag out–liers (kilógó vagy extrém) adatok esetén nem jellemzi jól a sokaságot, érzékeny az ilyen adatokra.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

5.1.1.2. Mértani átlag

A mértani átlag tulajdonsága, hogyha a megfigyelt értékeket a mértani átlaggal helyettesítjük, akkor szorzatuk az eredeti értékek szorzatával egyezik

$$\bar{x}_g \cdot \bar{x}_g \cdot \bar{x}_g \cdot \ldots \cdot \bar{x}_g = \bar{x}_g^N = x_1 \cdot x_2 \cdot x_3 \cdot \ldots \cdot x_N = \prod_{i=1}^{N} x_i$$

Az x_1, x_2, x_3, ..., x_N megfigyelt pozitív értékek mértani (geometriai) átlaga

$$\bar{x}_g = \sqrt[N]{x_1 \cdot x_2 \cdot x_3 \cdot \ldots \cdot x_N} = \sqrt[N]{\prod_{i=1}^{N} x_i}$$

ahol $\prod$ a produktum jele.

A mértani átlagot gyorsabban megkaphatjuk, ha az eredeti adatok logaritmusának összegét elosztjuk az elemszámmal

$$\log \bar{x}_g = \frac{\sum_{i=1}^{N} \log x_i}{N}$$

Innen az átlagot az antilogaritmus felhasználásával nyerjük

$$\bar{x}_g = \text{anti}\log(\log \bar{x}_g)$$

A mértani átlag kiszámításánál ügyelni kell arra, ha az értékek között az 0 érték is szerepel akkor a szorzat is és a mértani átlag is *0* lesz. Ilyen esetekben a mértani átlag meghatározásának nincs értelme.

Súlyozott mértani átlag kiszámítása a

$$\bar{x}_g = \sqrt[K]{x_1^{f_1} \cdot x_2^{f_2} \cdot x_3^{f_3} \cdot \ldots \cdot x_k^{f_k}}$$

formulával történik ahol

$$K = f_1 + f_2 + f_3 + \ldots + f_k$$

A mértani átlagot akkor célszerű alkalmazni, ha az értékek szorzata *0*–nál nagyobb szám és a mért értékek exponenciális eloszlásúak (exponenciálisan nőnek vagy csökkennek).

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Extrém adatokra kevésbé érzékeny. A számtani és a mértani átlag viszonyára a következő reláció az igaz

$$\bar{x}_g \le \bar{x}$$

5.1.1.3. Harmonikus átlag

Ha az x_i megfigyelt értékek helyébe a harmonikus átlagot tesszük, akkor reciprokaik összege az eredeti értékek reciprokainak összegével egyezik

$$\frac{1}{\bar{x}_h}+\frac{1}{\bar{x}_h}+\frac{1}{\bar{x}_h}+\ldots+\frac{1}{\bar{x}_h}=n\frac{1}{\bar{x}_h}=\frac{1}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}+\ldots+\frac{1}{x_N}=\sum_{i=1}^{N}\frac{1}{x_i}$$

A harmonikus átlag kiszámítási formulája

$$\bar{x}_h=\frac{N}{\sum_{i=1}^{N}\frac{1}{x_i}}$$

A harmonikus átlag kevésbé érzékeny a szélsőséges értékekre. Az x_h értéket mint átlagos túlélési időt, átlagsebességet, átlagteljesítményt (azonos időtartamra vonatkozóan) számítjuk. A súlyozott harmonikus átlag meghatározása

$$\bar{x}_h=\frac{\sum_{i=1}^{N}f_i}{\sum_{i=1}^{N}f_i\frac{1}{x_i}}$$

formula alapján történik.

Az $\bar{x}_h$, $\bar{x}_g$ és $\bar{x}$ értékek között érvényes a következő összefüggés

$$\bar{x}_h \le \bar{x}_g \le \bar{x}$$

5.1.1.4. Négyzetes átlag

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Meghatározása

$$\bar{x}_q = \sqrt{\frac{\sum_{i=1}^{N} x_i^2}{N}}$$

Ha az értékek helyébe az $\bar{x}_q$–t tesszük, és vesszük négyzeteik összegét, akkor fennáll a következő egyenlőség

$$N \cdot \bar{x}_q^2 = \sum_{i=1}^{N} x_i^2$$

A súlyozott négyzetes átlag a következő módon határozható meg

$$\bar{x}_q = \sqrt{\frac{\sum_{i=1}^{k} f_i x_i^2}{\sum_{i=1}^{k} f_i}}$$

A négyzetes átlag érzékeny a out–liers adatokra. Alkalmazása akkor kerül előtérbe, ha a mért értékek között pozitív és negatív értékek egyaránt előfordulnak, de csak az értékek abszolút nagyságát kívánjuk középértékkel jellemezni. Ilyen esetben az előjelek jelentőségétől eltekintünk. Jelentősége az adatok szórásánál lesz.

Az $\bar{x}_q$ és az $\bar{x}$ értékek között a kapcsolat

$$\bar{x} \leq \bar{x}_q$$

<u>Átlagokkal kapcsolatos megjegyzések</u>

a) Pozitív x_i értékek esetén, a négyfajta átlag viszonyára mindig igaz az alábbi összefüggés:

$$x_{min} < \bar{x}_h \leq \bar{x}_q \leq \bar{x} \leq \bar{x}_q < x_{max}$$

Konstans x_i értékek esetén nyilván mindegyik átlag azonos.

b) A mértani és a harmonikus átlag a nagyon alacsony, a kvadralikus átlag a nagyon magas értékekre érzékeny.

c) Használatos az ún. *trimmed mean*, amikor kilógó értékek miatt pl. elhagyjuk a minta alsó és felső 5%-át.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

5.1.2. Helyzeti középértékek

5.1.2.1. Módusz

A módusz (M_o vagy sűrűsödési középpont) a mintában az az érték, amely a leggyakrabban fordul elő. Ha az értékek egyforma gyakorisággal fordulnak elő a mintában, akkor a móduszt nem lehet egyértelműsíteni. Elsősorban intervallum vagy arányskálán mért adatok jellemzésére szolgál, de kvalitatív adatok esetén is használható. Több csúcsú eloszlásnál szintén hasznos az adatok jellemzésére.

Folytonos eloszlás esetén (pl. normális eloszlás) a módusz a görbe maximum értékénél van. Ebben az esetben nem beszélhetünk olyan értékről, amely a leggyakrabban fordul elő az adatok között. Meghatározása az osztályközös gyakorisági intervallumok alapján becsléssel történik.

Csoportosított adatok (egyenlő hosszúságú intervallumok) esetén a módusz meghatározása a

$$Mo = x_{i_0} + \frac{Mf_1}{Mf_1 + Mf_2} * h_i$$

formulával történik, ahol

x_{i_0} : a modális osztályköz alsó határa

Mf_1 : a modális osztályköz és az azt megelőző osztályköz gyakoriságának különbsége

Mf_2 : a modális osztályköz és az azt követő osztályköz gyakoriságának különbsége

h_i : a modális osztályköz hossza

5.1.2.2. Medián

A medián (Me) a nagyság szerint növekvő (csökkenő) sorrendbe rendezett adatok között a középső érték, az az 50%-os metszési pont vagy az adatok felező pontja (2. kvartilise), mivel a nálánál kisebb illetve nagyobb értékek gyakorisága azonos.

A medián a kiugró értékekre nem érzékeny, mivel a szélső értékek nem befolyásolják nagyságát. A medián a számtani közepet pótolja ferde (aszimmetrikus) eloszlásoknál vagy extrém értékek előfordulása esetén. Ordinális, intervallum vagy arányskálán mért adatok

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

jellemzésére használatos. Nevezetes tulajdonsága, hogy az adatoknak egy *c* konstanstól vett eltéréseinek összege akkor minimális, ha a konstans a mediánnal azonos:

$$\sum_{i=1}^{N} |x_i - c| = \text{minimum} \qquad \text{ha } c = Me$$

Értékét (racionális szám) a nagyság szerint rendezett adatokból kétféle módon lehet meghatározni

a) Ha az adatok száma páros:

akkor a két középső érték számtani közepe lesz a medián értéke.

b) Ha az adatok száma páratlan:

akkor a középső érték a medián.

Csoportosított adatok esetén kiszámítása a

$$Me = x_{i_0} + \frac{\frac{N}{2} - f'_{i-1}}{f_i} \bullet h_i$$

képlettel határozható meg, ahol

x_{i_0} : a mediánosztály alsó határa,

f'_{i-1} : az előző osztályközhöz tartozó kumulált gyakoriság,

f_i : a mediánosztályba eső elemek száma,

h_i : az osztályköz hossza,

n : a minta elemszáma.

5.1.2.3. Kvantilisek

A kvantilis értékek a méréssel, megfigyeléssel nyert elsősorban kvantitatív adatok rendezésére, azok eloszlásának megismerésére szolgálnak. Ezek az értékek az adatok elhelyezkedésének tömör leírását adják.

A különböző kvantilis értékek meghatározása úgy történik, hogy az első lépésben az adatokat nagyság szerint növekvőleg rendezzük, majd a minimum és maximum értékek által meghatározott tartományt *k* számú, egyenlő részre osztjuk. Az egyes tartományok felső határának értékei lesznek a kvantilis értékek.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Nevezetes kvantilis értékek:

a) medián (Me): a rendezett adatokat két részre osztjuk a medián alatt és fölött az értékek 50–50%–a szerepel.

b) kvartilisek (Q_{1-4}): a rendezett adattartományt 4 részre osztjuk, így 3 kvartilis értéket kapunk:

Q_1 = 25%–os rész értéke (alsó kvartilis)

Q_2 = 50%–os rész értéke (medián)

Q_3 = 75%–os rész értéke (felső kvartilis)

c) kvintilisek (K_{1-5}): a rendezett tartományt 5 részre osztjuk

d) decilisek (D_{1-10}): a rendezett tartományt 10 részre osztjuk

e) percentilisek (P_{1-100}): a rendezett tartományt 100 részre osztjuk. Ennek különösen az epidemiológiában van jelentős szerepe (5% és 95%-os értékek tartománya).

5.2. Szóródási mutatók

Az adatok egymástól való eltéréseit, variabilitását nevezzük szóródásnak vagy diszperziónak, amelyet egyetlen számmal fejezzük ki. Meghatározására több statisztikai mutató használatos

– a terjedelem (R)

– az interkvartilis terjedelem (IQT)

– átlagos abszolút eltérés

– szórás (s)

– relatív szórás (V).

5.2.1. Terjedelem

Az adatok közt előforduló legnagyobb és legkisebb érték különbséget nevezzük a szóródás terjedelmének

$$R = x_{max} - x_{min}$$

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Az érték az outlier adatokra nagyon érzékeny.

5.2.2. Interkvartilis terjedelem

A nagyság szerint sorbarendezett adatok tartományát negyedelve kapjuk meg a *4* db egyenlő elemszámot tartalmazó intervallumot. Az egyes intervallumokat elválasztó értékeket (Q_1, Q_2 = Me, Q_3) nevezzük kvartilisnek. A felső (Q_3) és alsó (Q_1) kvartilis különbsége az interkvartilis terjedelem

$$IQT = Q_3 - Q_1$$

Adataink minél kevésbé variábilisak, annál közelebb vannak egymáshoz a kvartilisek, illetve az ellentetje is igaz: minél távolabb vannak egymástól annál nagyobb az eltérés az adatok között.

Az IQT annak az intervallumnak a hossza, amelyben az adatok középső 50%–a helyezkedik el. A szóródásnak ez a mutatója az outlier értékekre nem érzékeny, de segítségükkel ezen értékek kideríthetők.

5.2.3. Átlagos abszolút eltérés

A szóródásnak ez a mérőszáma az outlier értékekre kevésbé érzékeny. A mérőszám az átlagtól számított eltérések abszolút értékeinek a számtani átlaga

$$\delta = \frac{\sum_{i=1}^{N} |x_i - \bar{x}|}{N} = \frac{\sum_{i=1}^{N} |d_i|}{N} \qquad \text{ahol } d_i = x_i - \bar{x}$$

Csoportosított adatok esetén

$$\delta = \frac{\sum_{i=1}^{N} f_i |x_i - \bar{x}|}{N} = \frac{\sum_{i=1}^{N} f_i |d_i|}{N}$$

5.2.4. Szórás

A leggyakrabban használt szóródási mutató, a statisztikai módszerek zöme ugyanis a *szórásanalízisre* épül. A szórás (standard deviation, SD) az adatoknak az átlagtól vett átlagos

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

eltérését jellemzi. A szórást *s*–el, ennek négyzetét a szórásnégyzetet (varianciát) s^2–el jelöljük. Az s^2 meghatározására két lehetőség van

a) Tapasztalati szórásnégyzet a mintaátlagtól való eltérések négyzetének az átlaga

$$s^{*2} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N}$$

Négyzetgyöke a tapasztalati szórás s*.

A tapasztalati szórásnégyzet egy véletlentől függő valószínűségi változó, amelytől azt várjuk el, hogy a várható értéke a populáció szórásnégyzetével (σ^2) legyen azonos. Gyakorlatilag azonban ez nem teljesül, ezért a s^{*2} értékét módosítani kell.

b) Korrigált empirikus szórásnégyzet várható értéke az elméleti szórásnégyzet (σ^2) lesz, ha a nevezőben az N helyett N–1 szerepel (df=N-1)

$$s^2 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1} = \frac{\sum_{i=1}^{N}x_i^2 - \frac{\left(\sum_{i=1}^{N}x_i\right)^2}{N}}{N-1}$$

és

$$M(s^2) = \sigma^2$$

Nagy mintaszám esetén s^{*2} és s^2 közötti eltérés nem jelentős, gyakorlatilag elhanyagolható. Ha a mintabeli elemek gyakoriságukkal adottak (csoportosított adatok esetén), akkor a korrigált empirikus szórásnégyzet

$$s^2 = \frac{\sum_{i=1}^{N}f_i(x_i - \bar{x})^2}{N-1}$$

5.2.5. Variációs együttható

Arányskálán mért adatok szóródásának relatív nagyságát méri. Dimenzió nélküli mutató, amely a szórás átlaghoz viszonyított nagyságát fejezi ki%–os formában:

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

$$V = \frac{s}{\bar{x}} \cdot 100\%$$

A relatív szórás az adatokon végrehajtott transzformációknak megfelelően az alábbiak szerint változik:

5.2.6. Relatív variációs együttható

A variációs együttható másik használt formája

$$V_r = \frac{s}{\bar{x} \cdot \sqrt{N}} \cdot 100\%$$

Normális eloszlástól való eltérés esetén a relatív szórás használata kerülendő.

5.2.7. Átlag szórása

Az átlag szórása vagy standard hibája (standard error of mean, SEM)

$$s_{\bar{x}} = \sqrt{\frac{s^2}{n}} = \frac{s}{\sqrt{N}}$$

5.2.8. Medián szórása

Nem normális eloszlás esetén az adatok jellemzésére az átlag helyett a mediánt és annak szóródását használhatjuk. Ez az adatoknak a mediántól vett abszolút eltéréseinek mediánja.

Például tekintsük az alábbi adatokat:

30, 40, 50, 80, 90

Az adatok mediánja az Me = 50. Vegyük az adatok abszolút eltéréseit a mediántól

$$|30–50| = 20, \quad |40–50| = 10, \quad |80–50| = 30, \quad |90–50| = 40$$

Rendezett formában a különbségek:

10, 20, 30, 40

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Ennek mediánja $\tilde{\text{Me}} = \frac{20+30}{2} = 25$

Így a medián szórása

$$\text{Me} \pm \tilde{\text{Me}} = 50 \pm 25$$

Normális eloszlás esetén a medián szórására

$$\frac{3\tilde{\text{Me}}}{2}$$

robusztus becslést használhatjuk. Normális eloszlás mellett ha a mintaszám nagy, akkor a medián szórása

$$s_{Me} = \frac{1.253 \cdot s}{\sqrt{N}}$$

amely kifejezés pontosabb eredményt ad.

5.3. Grafikus ábrázolás

5.3.1. Átlag±szórás ábrázolása

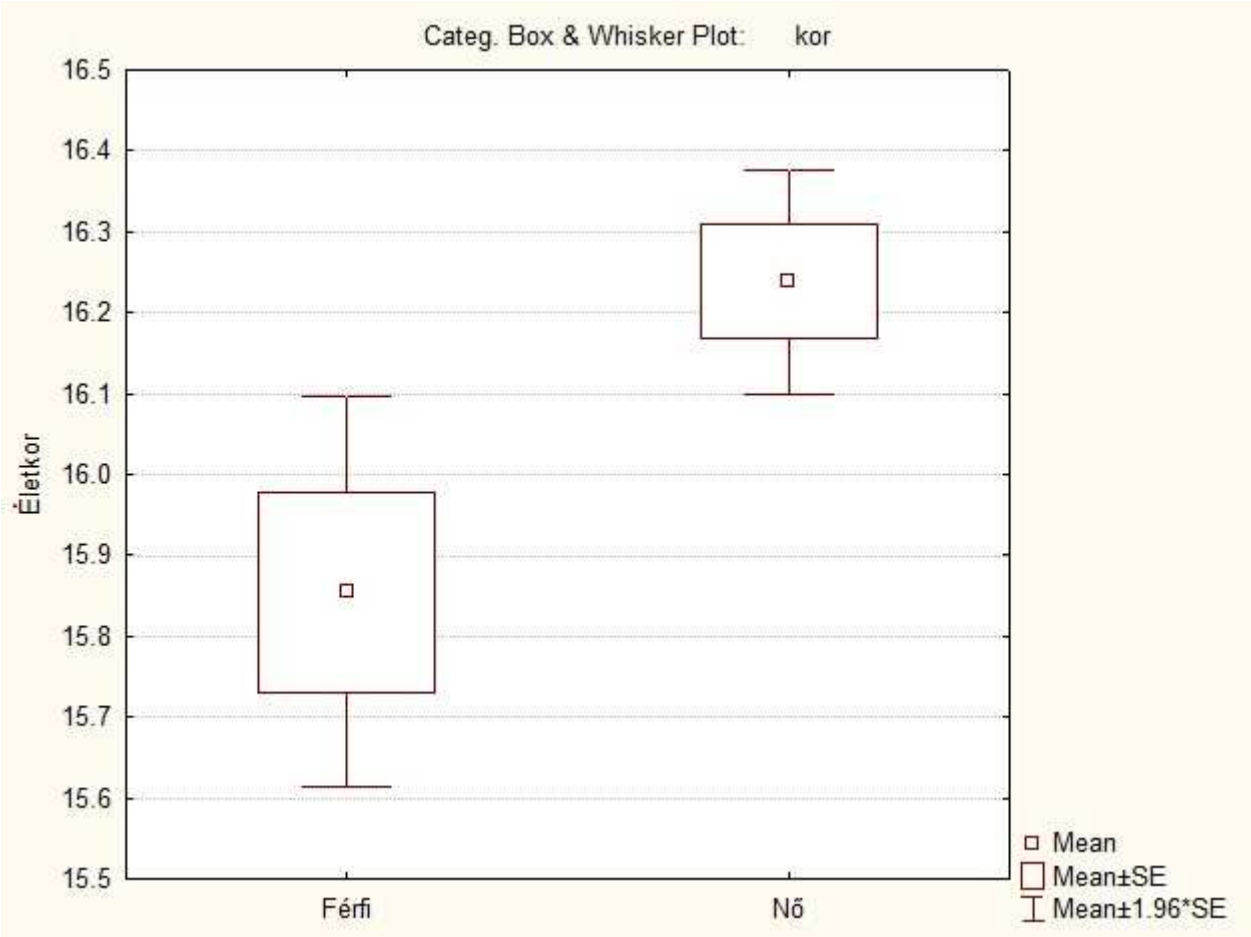
A tudományos szaklapokban megjelenő közlemények szinte mindegyike alkalmazza az adatok tulajdonságainak bemutatására az átlag és a szórás egyidejű ábrázolását. A közös ábra neve az ún. kalapácsos ábrázolás: az átlagot oszlopdiagramal ábrázoljuk, s erre helyezzük rá az adatok szórásértékét kis kalapács formájában. Az ilyen ábráknál jól érzékelhető az átlag és szórás viszonya és különösen csoportok megadása esetén vizuálisan összehasonlíthatók az egyes csoportok átlagai és szórásai.

5.3.2. Box and whiskers plot ábrázolás

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

A box and whiskers ábrázolás mintegy kiterjeszti az átlag±szórás által nyújtott információkat, átfogóbb, teljesebb áttekintést ad. Nagy előnye ennek az ábrázolási technikának, hogy az összes információt egy ábra tartalmazza.



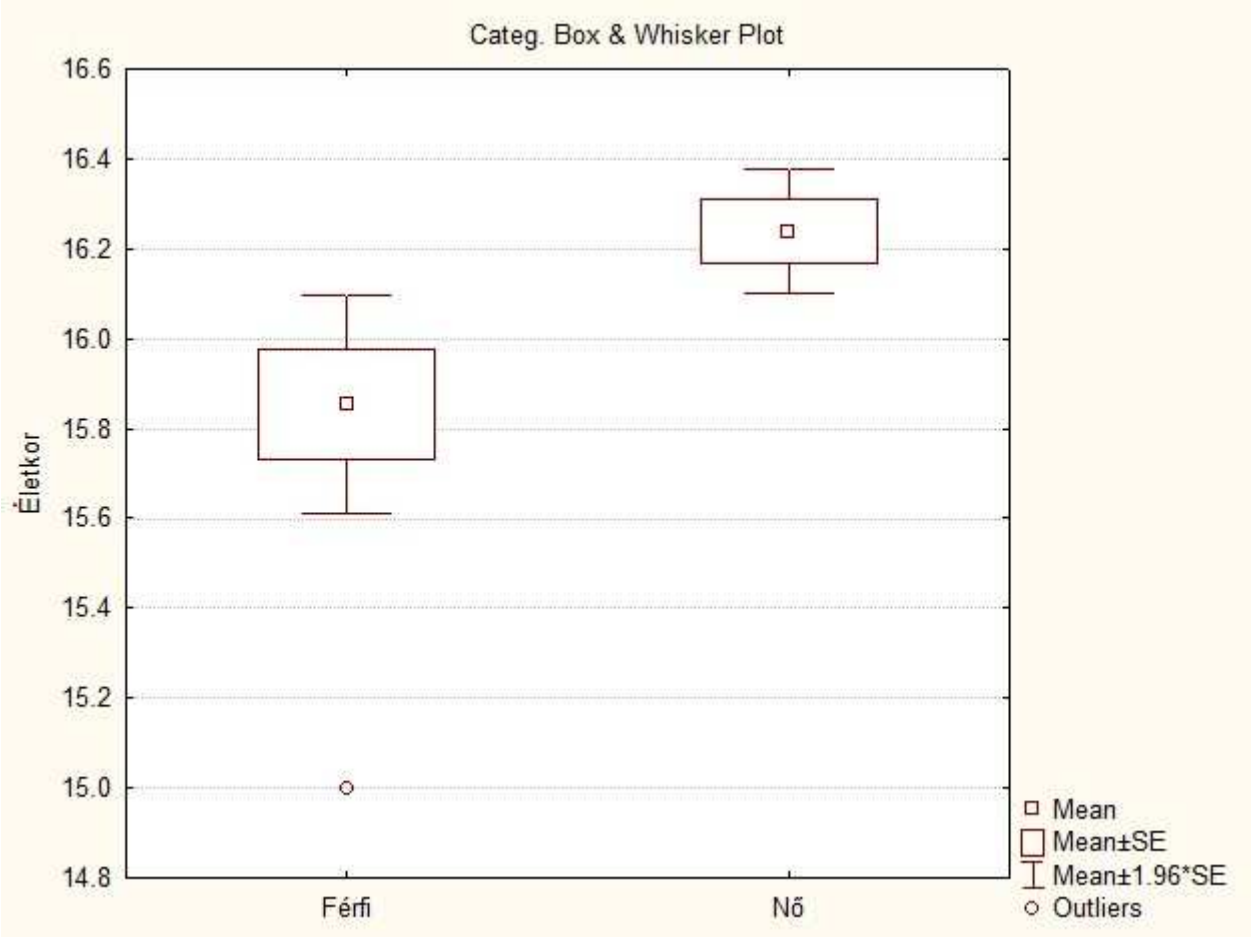
5.3.3. Kiugró értékek (outlier) vizsgálata

Az olyan adatokat, amelyek az eloszlás közepétől távol helyezkednek el kiugró, vagy extrém (szélsőséges) értékeknek nevezzük. Az alapsokaságtól való "elszakadás" származhat téves egyén mérésekor, megfigyelési hibából, a műszer téves leolvasásakor, de származhat olyan egyedi tulajdonságból is (ami az élő szervezetben nem ritka), amely nincs meg a többi egyednél. Ilyen adat(ok) esetén a helyes eljárás az, ha dupla statisztikai számítást végzünk: egyszer benthagyva az adatok között, egyszer pedig elhagyva végzünk statisztikai próbát,

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

hogy befolyásának hatását megismerjük. Az ilyen értékek grafikus ellenőrzésére a box-plot eljárást alkalmazzuk, amely a box and whiskers ábrázolás továbbfejlesztésének is tekinthető.
Az ábrázolás során kiugró értéknek bizonyult adatokat mindig meg kell vizsgálni. Ha csak adathiba lépett fel azt egyszerű korrekcióval javítani lehet. Ha a kiugró adat egyedi hatásból adódik, amely a többi egyedre nem lehet jellemző, akkor az ilyen értéket célszerű kihagyni a további elemzésből. A biometriai vizsgálatok során általában nem teszünk különbséget az enyhe és extrém kiugró értékek között. Éppen az élettani vizsgálatok fontossága miatt csak a belső határokat hagyjuk meg, és az azon kívüli értékek mindegyikét kiugró értéknek tekintjük.



Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

6. Konfidencia-intervallum

6.1. Megbízhatósági tartomány jelentősége

Konfidencia-intervallum (jelölésben: CI) adott szignifikancia szinten a becsült változó (pl. populációs átlag, a μ) alsó és felső korlátja: olyan intervallum értékű becslést ad egy paraméterre nézve, hogy az 1-α valószínűséggel esik ezen korlátok közé. Ez sok esetben jobb, mint egyetlen becsült értéket megadni. Ezt az *1-α* szintet sokszor százalékban adják meg; például 95% a tipikus érték.

Az intervallumbecslés szembeállítható a pontbecslésekkel. A pontbecslés egyetlen értékkel becsli meg az adott paramétert: azt mondja, hogy pl. 95%-os valószínűséggel közel van ehhez az értékhez. Ilyen paraméter pl. várható érték *(μ)* vagy a szórás *(σ)*.

A konfidencia-intervallum elemzésnek három előnye van a később ismertetendő hipotézis vizsgálatokkal szemben:

a) a vizsgálatok eredményei értelmezhetők a használt mértékegységre nézve,

b) tájékoztatást ad a hipotézis relatív pontosságáról,

c) az eredmények vizuálisan is megjeleníthetők.

A grafikus ábrázolás célja az, hogy mindegyik csoportra (pl. a vizsgált paraméter átlagára) meghatározzuk a 95%–os konfidencia-intervallumot és ábrázoljuk a tartományokat. Az alábbi három eset valamelyike lehetséges (függetlenül a konfidencia-intervallum megbízhatósági valószínűségétől)

a) Ha a két intervallum egyáltalán nem fedi egymást, akkor a két csoport átlaga között szignifikáns különbség van.

b) Ha az egyik intervallum a másik átlagát is tartalmazza, akkor az átlagok között nincs szignifikáns különbség.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

6.2. Átlag megbízhatósági tartománya

a) **Ha ismert a σ értéke**

$$CI_L = \bar{x} - z^* \frac{\sigma}{\sqrt{N}} \leq \mu \leq \bar{x} + z^* \frac{\sigma}{\sqrt{N}} = CI_U$$

ahol

$\bar{x}$: a mintaátlag

σ: az alapsokaság szórása

N: a mintanagyság

z*: a megbízhatósági valószínűséghez tartozó standard normális eloszlásból származó z érték (leggyakoribb értékek)

Megbízhatósági szint	z*
80%	1.28
90%	1.64
95%	1.96
99%	2.58
99.9%	3.29

Az $\bar{x}$–re szimmetrikus

$$\left(\bar{x} - z \cdot \frac{\sigma}{\sqrt{N}}, \quad \bar{x} + z \cdot \frac{\sigma}{\sqrt{N}} \right)$$

konfidenciaintervallum az eseteknek 100(1–α) %–ban tartalmazza az alapsokaság ismeretlen μ várható értékét. Ha α = 0,05 (5%), akkor μ 95%–ban ebben a konfidenciaintervallumban lesz benne és 5%–ban pedig ezen kívül. A számolás elvégzéséhez szükségünk van az elméleti szórás (σ) ismeretére, amit irodalom kutatás alapján megbecsülhetünk, vagy követelményként definiálunk.

b) **Ha csak a mintabeli szórás (s) ismert**

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

$$CI_L = \bar{x} - t^* \frac{\sigma}{\sqrt{N}} \le \mu \le \bar{x} + t^* \frac{\sigma}{\sqrt{N}} = CI_U$$

vagy

$$CI_L = \bar{x} - t^* \sqrt{\frac{s^2}{N}} \le \mu \le \bar{x} + t^* \sqrt{\frac{s^2}{N}} = CI_U$$

A *t** értékét a Student–eloszlás alapján határozzuk meg *df = N–1* szabadsági fok ismerete mellett.

A *CI* számításokra vonatkozóan, ha a megbízhatósági intervallumok pl. 95%–a tartalmazza az alapsokaság becsült, ismeretlen μ várható értékét az alábbi módon irható fel: $\alpha = 0{,}05$ értéke mellett

a) *ha σ ismert*

$$P(\bar{x} - 1.96\frac{\sigma}{\sqrt{N}} \le \mu \le \bar{x} + 1.96\frac{\sigma}{\sqrt{N}}) = 1 - \alpha = 0.95$$

b) *ha csak az s ismert*

$$P(\bar{x} - t_{0.05}\frac{s}{\sqrt{N}} \le \mu \le \bar{x} + t_{0.05}\frac{\sigma}{\sqrt{N}}) = 1 - \alpha = 0.95$$

6.3. A t*-eloszlás tulajdonságai:*

1) várható értéke *0*
2) a variancia nagyobb 1-nél, határértékben 1-hez közelít
3) szimmetrikus
4) értelmezési tartománya: - ∞ , ∞
5) eloszlás-család, *n*-1 a szabadsági fok, minél kisebb a minta (*n*), annál nagyobb a bizonytalanság, nagyobb a szórás
6) a *t*-eloszlás a normális eloszláshoz tart, ha $n \to \infty$. Minél nagyobb mintából becslünk annál jobb lesz az átlag szórásának becslése is. Általában, *n*>30 esetre a konfidencia intervallumhoz a normális eloszlás táblázata megfelelő.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

A konfidencia-intervallum általános módja mintából számolt szórással (s)

$$CI_L = \bar{x} - t_{(\alpha,N-1)} \frac{s}{\sqrt{N}}$$

$$CI_U = \bar{x} + t_{(\alpha,N-1)} \frac{s}{\sqrt{N}}$$

ahol

N-1: szabadsági fok

t (α, N-1): az ún. *t*-kritikus érték, amelyet a *t*-táblázat α oszlopából és N-1 sorából lehet kiolvasni.

7. Hipotézis vizsgálat

7.1. Hipotézis fogalma

Hipotézis: az alapsokaság paramétereire vagy az alapsokaság eloszlására vonatkozó feltevés. A gyakorlatban két hipotézissel dolgozunk:

H_0: null - hipotézis
H_1: alternatív hipotézis.

Pl. két minta átlagokra vonatkozóan formailag megfogalmazva

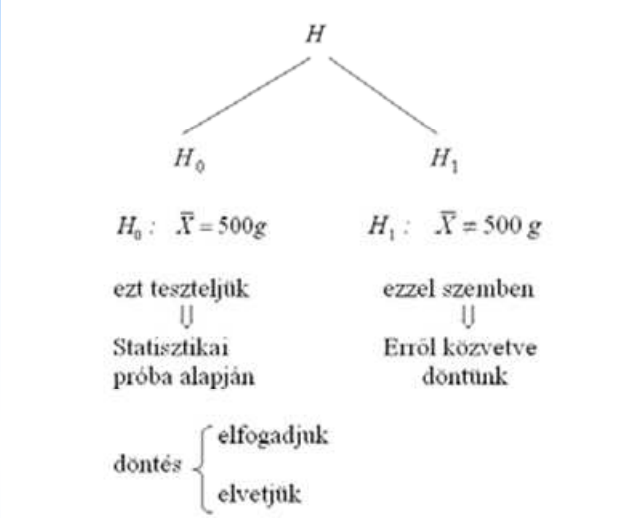
H_0: $\mu_1 = \mu_2$ vagy $\mu_1 - \mu_2 = 0$
H_1: $\mu_1 \neq \mu_2$

Hipotézisellenőrzés: az a statisztikai módszer, amelynek segítségével egy véletlen minta alapján eldöntjük, hogy az adott hipotézis *(H_0)* elfogadható-e vagy sem. Az olyan eljárást, amelyik a minták alapján dönt, *statisztikai próbának* nevezzük.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Hipotézisek megfogalmazása



7.2. Szignifikancia-szint

p-érték (empirikus szignifikancia-szint)

Az a legkisebb valószínűség, amely mellett a vizsgált H_0 hipotézist elutasíthatjuk a H_1 hipotézissel szemben, azaz, ahol éppen az elfogadásból az elutasításba váltunk.

Döntés a p értéke alapján:

$p < \alpha$: H_0-t elvetjük (elfogadjuk H_1-t)

$p \geq \alpha$: H_0-t elfogadjuk

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Döntés

Ha a **próbafüggvény értéke az E tartományba esik**, a tapasztalati adatok α szignifikancia szinten nem mondanak ellent a nullhipotézisnek. (H_0-t elfogadjuk)

Ha a **próbafüggvény értéke a K tartományba esik**, a nullhipotézist elvetjük és az alternatív hipotézist fogadjuk el. (H_1-t elfogadjuk)

7.3. Statisztikai próbák fajtái

Próbák

- **Kétoldali próba:** két oldalról állít alsó és felső korlátot (a feltételtől való eltérés tényét vizsgáljuk, irányát nem).
- **Egyoldali próba:** csak az egyik irányban állít korlátot (csak ilyen irányú eltérés lehetséges vagy fontos számunkra).

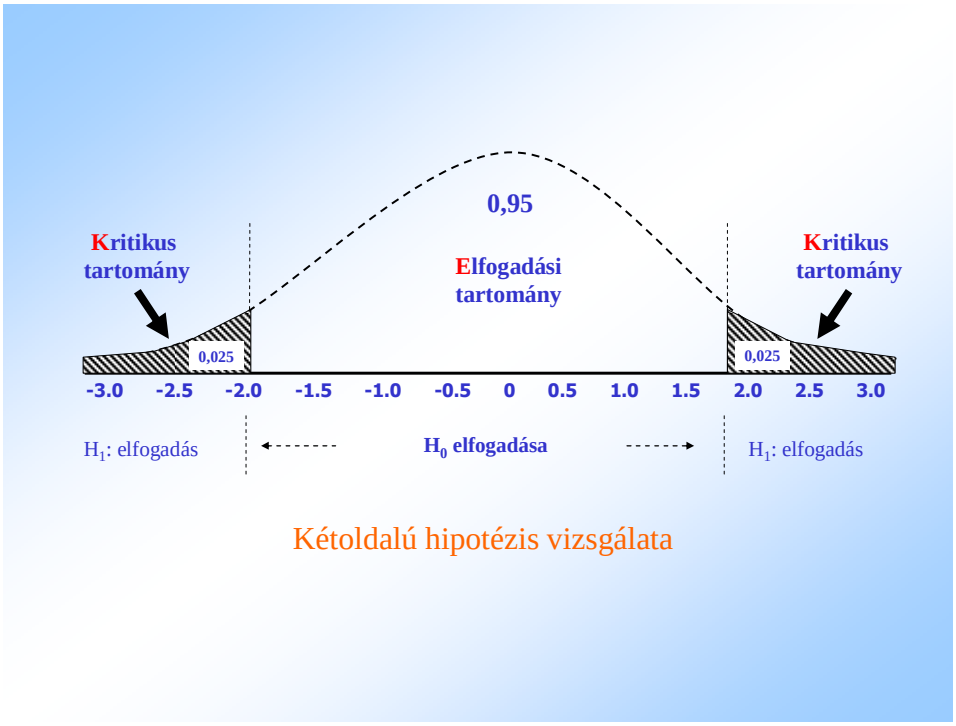
Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Kétoldalú próba:

H_0: $\overline{X_1} = \overline{X_2}$ (nincs változás az átlagok között)

H_1: $\overline{X_1} \neq \overline{X_2}$ (az átlagok nem egyenlőek (van változás: növekedés vagy csökkenés a beavatkozás után)



Kétoldalú hipotézis vizsgálata

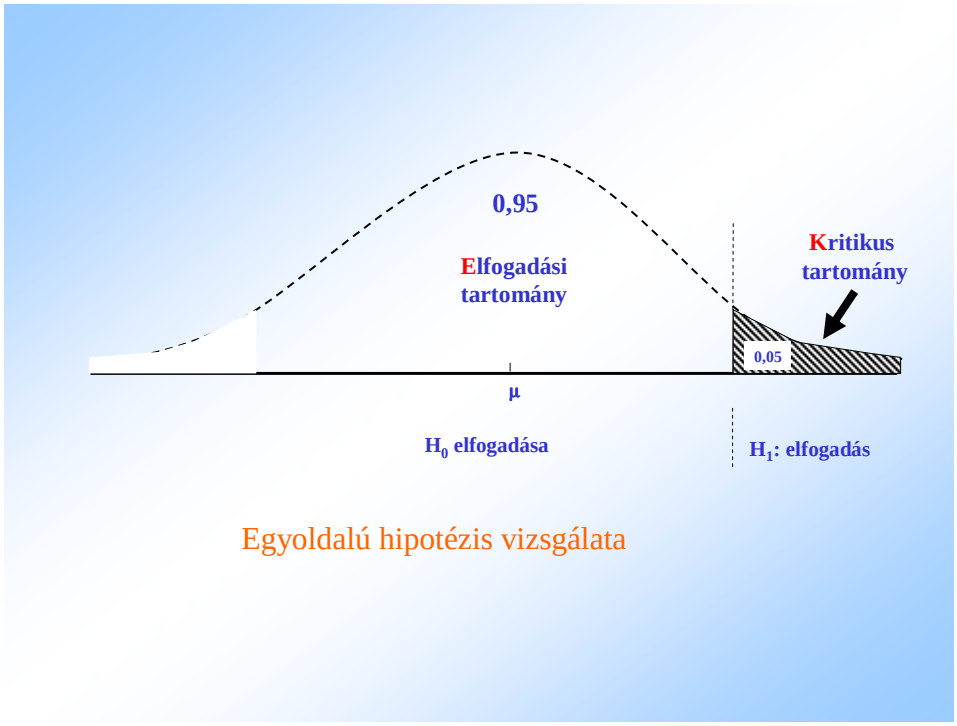
Egyoldalú próba:

H_0: $\overline{X_1} = \overline{X_2}$ (nincs változás az átlagok között)

H_1: $\overline{X_1} < \overline{X_2}$ (az átlag nő a beavatkozás után)

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.



Egyoldalú hipotézis vizsgálata

7.4. Hipotézis vizsgálat döntési táblázata

Meg kell határozni az alábbi hipotéziseket a vizsgálat indítása előtt:

H_0: Null-hipotézis
H_1: Alternatív hipotézis

Döntési táblázat

Valóshelyzet

	H_0 igaz	H_0 hamis
H_0 elfogadása	Helyes döntés (1-α)	Másodfajú hiba (β hiba)
H_0 elutasítása	Elsőfajú hiba (α hiba)	Helyes döntés (Power = 1-β)

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Elkövethető hibafajták

- **Type I error (α hiba vagy szignifikancia érték):** annak valószínűsége, hogy elutasítjuk a valós H_0 hipotézist.
- **Type II error(β hiba):** a hibás H_0 hipotézis elfogadásának valószínűsége.
- **Power:** a téves H_0 elutasításának valószínűsége. **Power = 1 - β.**

Értelmezések

- **1 - α: elfogadom a H_0 mikor az igaz, és elutasítom a nem igaz**

  H_1: μ1 ≠ μ2
- **α: elutasítom H_0 mikor az igaz, és elfogadom a nem igaz**

  H_1: μ1 ≠ μ2
- **1 - β: elutasítom a H_0 mikor az hamis, és elfogadom az igaz**

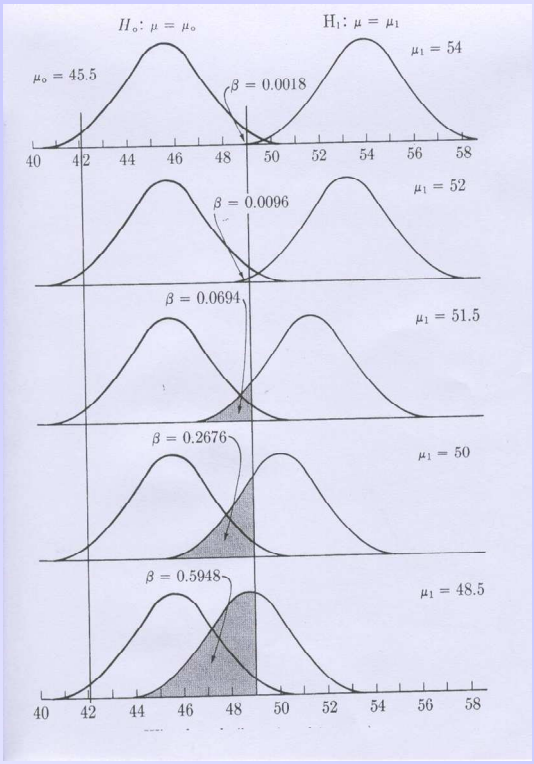
  H_1: μ1 ≠ μ2
- **β: elfogadom a H_0 mikor az hamis, és elutasítom az igaz**

  H_1: μ1 ≠ μ2

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Type II error (β)



Sokal: Biostatistics, 1982, 164. oldal

1. ábra: ha az alternatív hipotézisben (H_1) megjelölt várható érték (μ_1=54) távol esik a H_0-ban megjelölt várható értéktől (μ_0=45.5), akkor kicsi az átfedés, kicsi a β értéke is.

5. ábra: ha az alternatív hipotézisben megjelölt várható érték (μ1=54) közel van a H_0-ban megjelölt várható értékhez (μ_0=45.5), akkor nagy az átfedés, annál nagyobb a β értéke.

A mintaelemszám növelése csökkenti a β-t.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

7.5. Power-fogalma

A statisztikai próba ereje

- A valódi különbség kimutatásának valószínűsége P=1- β.
- Gyakorlatilag egy igaz munkahipotézis vagy alternatív hipotézis elfogadásának a valószínűsége.
- Minél kisebb az α, annál ritkább, hogy H_0 -t tévesen elutasítjuk, de annál gyakoribb, hogy H_0-t tévesen elfogadjuk (másodfajú hiba)

Az elso- és másodfajú hiba csökkentése

- Minta elemszámának növelése.
- Pontosabb mintavételezés (szórás csökken).
- Lehet-e az elso- és másodfajú hibát nullára csökkenteni? Válasz: NEM.
- A véletlen hatásokat nem tudjuk kiiktatni.

7.6. Hipotézis vizsgálat menete

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

A hipotézis vizsgálat menete

- A null- és alternatív hipotézis megfogalmazása.
- Próbafüggvény keresése/szerkesztése.
- Előre rögzített szignifikanciaszint mellett az elfogadási és elutasítási tartomány megszerkesztése.
- A próbafüggvény empirikus értékének meghatározása.
- Döntés: az eredmény klinikailag releváns-e?!

8. Power analízis

8.1. Mintaszám meghatározása

A mintaválasztásnál arra törekszünk, hogy olyan mintát válasszunk, amely szükséges és elegendően nagy a szignifikáns különbség biztos kimutatásához. Ugyanis, ha a mintánk indokolatlanul nagy, akkor csökkentjük a CI-t (megbízhatósági tartományt) és klinikailag érdektelen különbségeket is szignifikáns különbségként mutatunk ki. Fordítva: a szükségesnél kisebb mintaszám pedig nem alkalmas ténylegesen meglévő különbségek kimutatására.

8.1.1. Az átlag becsléséhez szükséges mintaszám

Az elsőfajú (α) és a másodfajú (β) hiba értéke lehetőséget ad a mintaszám meghatározására. Kétoldalú próbát feltételezve, meghatározhatjuk a z_α és z_β értékeket (a β–ra mindig egyoldalú próbát alkalmazunk):

Az α és β értékekhez tartozó kritikus értékek

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

$$z_\alpha = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{N}}} \quad \text{és} \quad z_\beta = \frac{\bar{x} - \mu_1}{\frac{\sigma}{\sqrt{N}}}$$

Az egyenletekből $\bar{x}$–t kifejezve és egyenlővé téve a két oldalt az N meghatározható

$$N = \left[\frac{(z_\alpha - z_\beta) \cdot \sigma}{\mu_1 - \mu_0} \right]^2$$

8.1.2 Két átlagérték összehasonlításához szükséges mintaszám

A két minta középértékei közötti szignifikáns eltérés kimutatásához (azonos varianciájú és normális eloszlásból származó független minták esetén) szükséges mintaszám

$$N = 2 \cdot \left[\frac{(z_\alpha - z_\beta) \cdot \sigma}{\mu_1 - \mu_0} \right]^2$$

Az *N* értéke mindkét csoportra vonatkozik, (azonos létszámú csoportok).

A szükséges mintaszám megállapításánál vegyük figyelembe az alábbiakat:

a) A számításnál feltesszük, hogy a két populáció szórása azonos. A becsült szórást irodalmi adatok vagy pilot–study (elővizsgálat) alapján meghatározhatjuk. Minél nagyobb a szórás, annál nagyobb lesz az *N* értéke is.

b) A szignifikancia szint (α) értéke általában 0.05. Alacsonyabb érték az *N* értékét növeli.

c) A másodfajú hiba (β) értéke szintén befolyásolja az *N* értékét. Minél alacsonyabbra választjuk értékét (vagyis emeljük a *Power* nagyságát (P=1–β), annál nagyobb a mintaszám. A leggyakrabban a 80% vagy 90%–os *Power* értékeket használjuk vizsgálatainkban, ami β= 0.2 és β= 0.1 értékeknek felel meg.

d) Minél kisebb eltérést akarunk kimutatni szignifikáns értékként, annál nagyobb lesz az *N* értéke.

A mintaszám meghatározása során még a vizsgálat megkezdése előtt el kell dönteni,

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

hogy a z_α esetében egyoldalú vagy kétoldalú próbát alkalmazunk–e. A z_β értéke mindig az egyoldalú próbának megfelelő éréket veszi fel.

Az alábbi táblázat azt mutatja, hogy különböző power és α értékek mellett a szükséges mintaszám hogyan változik

		α(kétoldalú próba)			
power	β	0.10	0.05	0.02	0.01
80	0.20	18	23	29	34
90	0.10	25	31	38	43
95	0.05	32	38	46	52

8.1.3. Adott arány különbségéhez szükséges mintaszám

A szükséges mintaszám meghatározásához az alábbi kérdésekre kell a választ előzetesen megadni:

a) Mi a nullhipotézis (Π_0) és a hozzátartozó α érték?

b) Ki kell jelöni az alternatív hipotézist (Π_1) és a power nagyságát

c) A két arány különbsége (Π_1–Π_0) klinikailag elég jelentős–e?

A kérdések megválaszolása után a mintaszámot az alábbi módon határozhatjuk meg:

$$N = \left[\frac{z_\alpha \sqrt{\Pi_0(1-\Pi_0)} - z_\beta \sqrt{\Pi_1(1-\Pi_1)}}{\Pi_1 - \Pi_0} \right]^2$$

ahol z_α az α–hoz tartozó kétoldalú teszt z értékét, z_β a β–hoz tartozó egyoldalú teszt z értékét jelenti.

Az alábbi táblázat ugyanerre a problémára vonatkozó mintaszámokat mutatja különböző power és α értékek esetén:

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

		α (kétoldalú próba)			
Power	β	0.10	0.05	0.02	0.01
80	0.20	26	34	44	52
90	0.10	33	42	53	62
95	0.05	39	49	61	71

8.1.4. Két arány összehasonlításához szükséges mintaszám

A számítások egyszerűsége végett a csoportokban azonos mintaszámot tételezünk fel. Π_1 jelenti az egyik csoportban, Π_2 a másik csoportban a vizsgált arányszámot. A két arány különbségének kimutatásához szükséges mintaszám:

$$N = \frac{2 \cdot \bar{\pi}(1-\bar{\pi})(z_\alpha + z_\beta)^2}{(\Pi_1 - \Pi_2)^2} \quad \text{ahol} \quad \bar{\pi} = \frac{\Pi_1 + \Pi_2}{2}$$

8.1.5. Mintaszám meghatározás konfidenciaintervallum alapján

A *CI* meghatározásánál láttuk, hogy az intervallum hossza a mintaszám nagyságától függ: nagy mintaszám esetén a *CI* rövidebb.

8.1.5.1. Átlagra vonatkozó mintaszám

Az *N* értékének meghatározásához három adatra van szükség: a *CI* megbízhatósági valószínűségére, a becsült szórásra (s) és az előre definiált intervallum hosszának a felére (HCl). Ezen adatok alapján a kívánt *CI*–hez szükséges mintaszám nagysága:

$$N \approx z_\alpha^2 \left(\frac{s}{HCI} \right)^2$$

A kapott *N* érték csak egy becslése a tényleges mintaszámnak, hiszen az s értékét pontosan nem ismerjük. Az s értékét irodalmi adatok vagy pilot–study alapján határozhatjuk

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

8.1.5.2. Két átlag különbségére vonatkozó mintaszám

A mintaszám meghatározása során feltesszük, hogy a szórás mindkét populációban közös, valamint a számított mintaszám mindegyik csoportra vonatkozik. Két átlag különbségének konfidenciaintervallumára vonatkozó mintaszám az alábbi módon határozható meg

$$N \approx 2 \cdot z_{\alpha}^{2} \left(\frac{s}{HCI} \right)^{2}$$

Az alábbi táblázat a különböző CI értékekhez tartozó mintaszámot tünteti fel.

Konfidenciaintervallum (CI) értékek				
80%	90%	95%	98%	99%
21	34	49	68	83

8.1.5.3. Arányra vonatkozó mintaszám

A kérdést úgy fogalmazhatnánk meg, hogy adott pontosságú becslés mellett hány elemű mintára van szükségünk egy arány értékének meghatározásához. A szükséges mintaszám:

$$\mathrm{N} \approx \mathrm{z}_{\alpha}^{2} \frac{\Pi(1-\Pi)}{\mathrm{HCI}^{2}}$$

Ha a π értékét nem tudjuk megbecsülni a számításhoz, akkor legyen a π=0.5, mivel akkor lesz maximális értékű a $\pi(1-\pi)$ kifejezés. Ebben az esetben legfeljebb a szükséges mintaszámot felülbecsüljük.

Konfidenciaintervallum (CI) értékek				
80%	90%	95%	98%	99%
370	609	865	1218	1493

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

8.1.5.4. Két arány különbségre vonatkozó mintaszám

Ahhoz, hogy két csoport arányainak különbségét adott pontossággal becsülni tudjuk az alábbi mintaszámokra van szükségünk csoportonként:

$$N = 2 \cdot z_{\alpha}^{2} \frac{\overline{\Pi}(1-\overline{\Pi})}{HCI^{2}} \quad \text{ahol} \quad \overline{\pi} = \frac{\Pi_1 - \Pi_2}{2}$$

Az alábbi táblázat különböző CI értékekhez tartozó mintaszámokat mutatja.

Konfidenciaintervallum (CI) értékek				
80%	90%	95%	98%	99%
899	1480	2101	2307	3629

A szükséges mintaszám meghatározását könnyebbé tehetjük olyan esetekben, amikor nem tudjuk a $\overline{\Pi}$ átlag értékét megbecsülni, mivel a csoportokra vonatkozólag nincsenek ismereteink. Ilyen esetekben célszerű a $\overline{\Pi}$=0.05 értéket venni, ugyanis ebben az esetben maximális mintaszámot kapunk, ami legfeljebb felülbecslést eredményez.

8.1.6. Nem egyenlő mintaszámú csoportok

Ezen számítások komplikáltabbak és erre a célra számítógépes programokat használunk. Ilyen esetben a számításhoz vagy az egyik csoportra vonatkozólag adjuk meg a mintaszámot és a program kiszámítja adott *Power* mellett a másik csoporthoz tartozó mintaszámot, vagy a tervezett arányát adjuk meg a csoportoknak. További lehetőségként megadjuk a két csoport összegét, s ezt bontja fel a program két nem egyenlő számú csoportra.

9. Paraméteres eljárások

A csoportba tartozó statisztikai eljárások *közös* jellemzője, hogy a vizsgált valószínűségi változók eloszlása normális eloszlást követ. A számítási eljárások erre a tulajdonságra épülnek.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

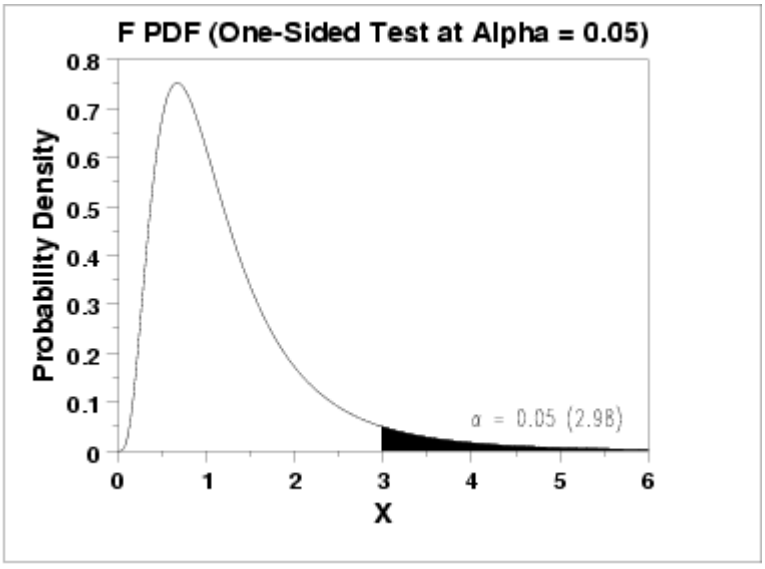
9.1. F - próba

Igen gyakran használt eljárás két variancia homogenitásának (homoszcedaszticitás) eldöntésére, azaz a kétminta azonos varianciájú alapsokaságból származik-e.

H_0: $S_1^2 = S_2^2$ azaz $S_1^2 - S_2^2 = 0$ (varianciák azonosak)

H_1: $S_1^2 \neq S_2^2$ azaz $S_1^2 - S_2^2 \neq 0$ (varianciák nem azonosak)

A két minta elemszámai: N_1 és N_2 a két szabadsági fok $df_1 = n_1 - 1$ és $df_2 = n_2 - 1$.
Az $F_{krit, (N1-1, N2-1)}$ két szabadsági foktól függ valamint α-tól..Minden *F* eloszlás aszimmetrikus, ezért az *F*-táblázatok küszöbértékei *egyoldalas tesztre* vonatkoznak. Az F_{krit} értékek közvetlenül használhatóak egyoldalú alternatív hipotézis esetén, pl. kétmintás t-tesztnél.



Az F-próba esetében (alapeset) *kétoldalú* alternatív H_1 hipotézist vizsgálunk ($S_1^2 - S_2^2 \neq 0$).

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

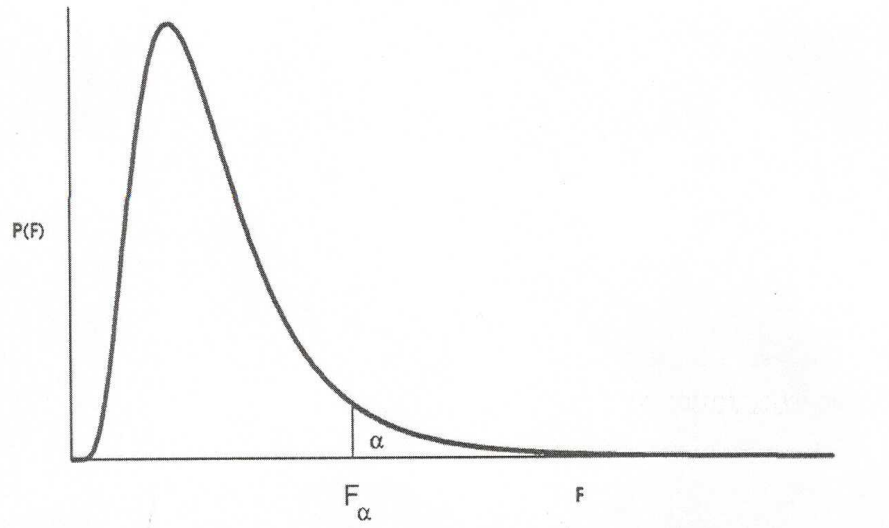


Figure K.1: The F distribution

Az F értékének kiszámítása egyszerű

$$F_{df\,1,df\,2} = \frac{s_1^2}{s_2^2}$$

Ahol df_1 jelenti a számláló szabadságfokát ($df_1 = N_1 - 1$), df_2 a nevező szabadságfokát ($df_2 = N_2 - 1$). A számításnál mindig a nagyobb szabadságfokú tagot tesszük a számlálóba.

Kétoldalú *F* próba esetén az α szignifikancia szinthez tartozó értéket az egyoldalú próbához megadott *F*-táblázat α/2 jelű sorából keressük ki, vagyis általában 0,025-nél. (Ha külön táblázatok vannak a különböző szignifikancia-szintekhez, akkor az α=0,025-höz tartozó táblázatot kell használni.).

9.1.1. Döntés a hipotézisek felől

Elfogadjuk H_0-t, ha F<Fkrit: a két mintából becsült variancia nem különbözik egymástól szignifikánsan, a minták azonos varianciájú alapsokaságból származnak.
Elvetjük H_0-t, ha F>Fkrit: a két mintából becsült variancia szignifikánsan különbözik, a minták nem származnak azonos varianciájú alapsokaságból.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu
MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Megállapíthatjuk:

- ha F-próbával a két variancia azonos, akkor pl. használhatunk *kétmintás t*-próbát;
- ha a két variancia nem azonos, akkor az ún. *d-próbát* (Welch próbát) használunk..

9.2. Egymintás t-teszt

9.2.1. Egyetlen minta várható értékének vizsgálata

Az eljárást akkor használjuk, ha azt vizsgáljuk, hogy egy a populáció várható értéke megegyezik–e egy feltételezett várható értékkel vagy pedig szignifikánsan eltér attól. Feltétel, hogy a változó legyen normális eloszlású.

Hipotézisek:

H_o: a mintából kapott átlag μ becslése
H_1: az átlag nem a μ becslése

Formálisan megfogalmazva:

$H_o: \bar{x} - \mu = 0$ (nem tér el szignifikánsan μ-től)
$H_1: \bar{x} - \mu \neq 0$ (szignifikánsan eltér)

Mivel a populációt mérni nem tudjuk, ezért a várható értékét illetően feltételezéssel kell élni. A vizsgálathoz szükségünk van az alábbi statisztikára

$$t = \frac{\bar{x}}{s_{\bar{x}}}$$

ami egy *t*–eloszlás, *df=N–1* szabadságfokkal. A táblázat használatakor *t* abszolút értékével kell számolni, azt kell hasonlítani a táblázat adott szabadsági foknál és α értéknél lévő kritikus értékhez. Amennyiben a számolt *t* értékünk abszolút értéke kisebb, mint t_{krit}, úgy a H_0 nullhipotézist α szignifikancia szinten elfogadjuk; ellenkező esetben elvetjük és a H_1-t fogadjuk el.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Általánosabb formában, amikor egy *C*-értékhez hasonlítjuk az átlagot

$$t = \frac{\bar{x} - C}{s_{\bar{x}}}$$

9.2.2. Egyoldalú egymintás t-próba

$$H_o: \bar{x} - \mu = 0 \text{ (nem tér el szignifikánsan } \mu\text{-től)}$$
$$H_1: \bar{x} - \mu > 0 \text{ (az átlag nagyobb)}$$

adott α és *df* = *N*-1 értékeknél.

Itt a t_{krit} értéket a kétoldalú próbához megadott táblázatból a 2α jelű oszlopból keressük ki: α = 5% esetén a 10%-hoz tartozó t_{krit} értéket használjuk, következésképpen kisebb eltérés is elég H_o elvetésére.

9.2.3. Párosítottt-próba

Egy kezelés hatásosságát gyakran úgy értékeljük, hogy ugyanazokon a betegeken két mérést (önkontrollos vizsgálat) végzünk különböző időpontokban: kezelés előtt (t_0) és után (t_1), így a két *N*-elemű összetartozó párokból álló mintát kapunk. A két minta, a kezelés előtti és a kezelés utáni *nem független*, hiszen ugyanazok a betegek szerepelnek a mintákban. Minden betegre kiszámítjuk a kezelés okozta különbségeket (di) és ezt a differenciát tekintjük valószínűségi változónak, erre alkalmazzuk az egymintás *t*-próbát.
Vizsgált hipotézisek:

$$H_0: \bar{d} = 0 \text{ vagy } \bar{x}_1 - \bar{x}_2 = 0$$
$$H_1: \bar{d} \neq 0 \text{ vagy } \bar{x}_1 - \bar{x}_2 \neq 0$$

kifejezést is használhatjuk.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

A különbség értékek varianciája

$$s_d^2 = \frac{\sum_{i=1}^{N}(d_i - \bar{d})^2}{N-1}$$

ahol N a párosított adatok számát, $\bar{d}$ az eltérések átlagát jelenti. Az átlagos eltérés standard hibája

$$s_{\bar{d}}^2 = \frac{s_d^2}{N}$$

A hipotézis ellenőrzéséhez szükséges t–statisztika értéke

$$t_f = \frac{\bar{d} - 0}{s_{\bar{d}}}$$

ahol a df szabadságfok értéke N–1.

A számítások egyszerűsítése végett a

$$t = \frac{\sum_{i=1}^{N} d_i}{\sqrt{\frac{\left[N \cdot \sum_{i=1}^{N} d_i^2 - \left(\sum_{i=1}^{N} d_i\right)^2\right]}{N-1}}}$$

kifejezés is használható.

9.2.4. Matched pairs- módszer

Előfordulnak olyan esetek, amikor azonos alanyokon nem végezhető el mindkét mérés valamilyen ok miatt, ilyenkor a betegeket összekapcsoljuk pl. szociális körülmények, nem, betegség súlyossága stb. alapján, és ezekből képezünk egy párt: random módon a párból az egyiket a kezelt a másikat a kezelés nélküli csoportba tesszük. Az ilyen vizsgálatot a párosított *t*-próbához hasonlóan a párok két tagja közötti különbségekkel értékeljük ki, és alkalmazzuk az *egymintás t*-próbát.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

9.3. Kétmintás t-teszt

Két független minta összehasonlítására használjuk. A függetlenség azt jelenti, hogy mindegyik csoport szeparált a másiktól. Például ilyen csoportokat kapunk, ha a férfiak és nők között végzünk összehasonlítást. A csoportok tagjai nem keveredhetnek és az egyes csoportokon belül sem szabad az adatokat megduplázni, mert ezzel megsértenénk a csoportok függetlenségét. A próba használatának feltételei:

a) csoportok függetlensége
b) adatok normalitása
c) csoportok varianciája legyen azonos (F-próba).

Az utóbbi feltétel nem teljesülése esetén is használható azonban a próba, mivel erre az esetre is van módosított eljárás.
Használt hipotézisek:

$H_0 : \bar{x}_1 = \bar{x}_2$ (csoportok között nincs szignifikáns eltérés)

$H_1 : \bar{x}_1 \neq \bar{x}_2$ (van szignifikáns eltérés)

A próba nem követeli meg a csoportok azonos elemszámát, így eltérő elemszámú csoportokra is használható. A próbának két változata van attól függően, hogy teljesül–e a csoportok közötti variancia–azonosság vagy sem.

9.3.1. Csoportok közötti variancia egyenlő

A két csoport esetén a populáció torzítatlan becslése (pooled variance) a következőképpen írható

$$s^2 = \frac{\sum_{i=1}^{N_1}\left(x_i - \bar{x}_1\right)^2 + \sum_{j=1}^{N_2}\left(x_j - \bar{x}_2\right)^2}{N_1 + N_2 - 2}$$

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

ahol a számlálóban az egyes csoportok átlagtól való eltérésének négyzetösszege, a nevezőben a szabadságfok áll. Összesen két szabadságfokot veszítünk a számolás során, mivel mindegyik minta átlagát külön–külön határozzuk meg. A két átlag eltérésének standard hibája

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s^2}{N_1} + \frac{s^2}{N_2}} = s \cdot \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}$$

Ezen értékek alapján a t–statisztika értéke

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1-\bar{x}_2}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s^2}{N_1} + \frac{s^2}{N_2}}}$$

A fenti formula a következőképpen is megadható

$$s^2 = \frac{\sum_{i=1}^{N_1} x_i^2 - \frac{\left(\sum_{i=1}^{N_1} x_i\right)^2}{N_1} + \sum_{j=1}^{N_2} x_j^2 - \frac{\left(\sum_{j=1}^{N_2} x_j\right)^2}{N_2}}{N_1 + N_2 - 2}$$

9.3.2. Csoportok közötti variancia nem egyenlő

Olyan esetekben, amikor nem teljesül a csoportok közötti variancia feltétele módosított összehasonlító eljárást alkalmazunk. Ilyen két eljárás a Cochran – Cox és a Welch eljárás. Az első módszer a *t* értékét korrigálja az adott szignifikancia szinten (leggyakrabban ez 5%), míg a Welch módszer a szabadságfokot módosítja.

Cochran – Cox módszer esetén a két minta különbségének standard hibája

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sum_{i=1}^{N_1}\left(x_i - \bar{x}_1\right)^2}{N_1(N_1-1)} - \frac{\sum_{j=1}^{N_2}\left(x_j - \bar{x}_2\right)^2}{N_2(N_2-1)}} = \sqrt{s_{\bar{x}_1}^2 + s_{\bar{x}_2}^2}$$

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Ennek megfelelően a *t* statisztika értéke

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}} \qquad *$$

A *t* értékét az adott szignifikancia szinten (a számítások során legyen 5%) a következőképpen módosítjuk

$$t_{0.05} = \frac{t_1 \cdot s^2_{\bar{x}_1} + t_2 \cdot s^2_{\bar{x}_2}}{s^2_{\bar{x}_1} + s^2_{\bar{x}_2}}$$

ahol t_1 az N_1–1, t_2 az N_2–1 szabadságfokhoz tartozó *t* kritikus érték az 5%–os szignifikancia szintnek megfelelően. A csoportok átlagai közötti eltérést akkor minősítjük szignifikánsnak, ha $t_{0.05} < t$. Ellenkező relációnál az eltérés nem szignifikáns.

Welch eljárásnál (d–próba) a *t* értékének meghatározására szintén a fenti *-al jelölt képletet használjuk, de a szabadságfok meghatározása nem a minta elemszám alapján történik

$$df = \frac{\left(s^2_{\bar{x}_1} + s^2_{\bar{x}_2}\right)^2}{\frac{\left(s^2_{\bar{x}_1}\right)^2}{N_1 + 1} + \frac{\left(s^2_{\bar{x}_2}\right)^2}{N_2 + 1}} - 2$$

A kapott *df* értéke nem lesz egész érték (a hozzá legközelebb eső egész számot vesszük).

Döntés a számított t-értékek alapján

- $|t| < |t_{krit}|$: elfogadjuk H_0-t, a két minta azonos alapsokaságból származik (a két átlag különbözősége csak a véletlennek tudható be),
- $|t| \rangle |t_{krit}|$: elvetjük H_0-t, a két minta nem azonos alapsokaságból származik; a két átlag különbözőségét szisztematikus hatásnak tudjuk be.

9.3.3. Kétmintás z–teszt

Legyen két populációnk *X* és *Y* ismert σ^2_x és σ^2_y varianciákkal, N_1 az *X*, N_2 az *Y* populációból származó minta.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Kérdés: a populációk átlagai a μ_x és μ_y azonosak–e?

Hipotézisek

H_0: $\mu_x = \mu_y$ (átlagok azonosak)

H_1: $\mu_x \neq \mu_y$ (átlagok nem azonosak)

A választ a *z*–statisztika segítségével adjuk meg

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{N_1} + \frac{\sigma_y^2}{N_2}}}$$

A *z*–statisztika standard normális eloszlást követ, így ennek táblázatot használhatjuk a *z* kritikus értékének meghatározására (adott α szignifikancia szinten). A változókról feltesszük, hogy normális eloszlásúak. Ha ez nem teljesül, akkor a mintaszámot kell megnövelni (N>30 mintaszámnál az eloszlás már közel normálisnak tekinthető).

9.3.4. t–próba ereje

A számítások a standard normális eloszláson alapulnak. Két lépésben hajtsuk végre a számítást:

a) Meghatározzuk az eloszlás z_p értékét.

$$z_p = (t - t^*)\left(\frac{1}{\sqrt{1 + \frac{t^{*2}}{2df}}}\right)$$

A *t* a már ismert módon határozható meg

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}} = \frac{\Delta_k}{\text{a különbség standard hibája}}$$

és t* az adott szabadságfokhoz tartozó kritikus érték.

A z_p értékhez kikeressük a megfelelő power értéket tartalmazó táblázatból.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

9.4.Több folytonos, normális eloszlású adatsor összehasonlítása (ANOVA)

Az ANOVA (**AN**alysis **O**f **VA**riance) módszert olyan esetkben használjuk, amikor több mintát kell egyidejűleg összehasonlítani. A *t*-tesztek az ANOVA eljárás speciális esetéinek tekinthetők:

a) *Párosított t-teszt*: ismételt méréses ANOVA lényegében, csak két időpontra vonatkoztatva.
b) *Kétmintás t-teszt*: egyszempontos ANOVA lényegében, két csoportra vonatkoztatva.

Az ANOVA lényege, hogy a mintákból számolt összvarianciát két részre osztjuk, mintákon belüli (within) és minták közötti varianciára (between). A statisztikai analízis során ezt a két részvarianciát hasonlítjuk össze F–próbával és attól függően, hogy melyik hatás (csoporton belüli vagy csoportok közötti) a domináns, döntünk a vizsgálat felől. Pl. ha négy fajta készítmény (referens és három új készítmény) terápiás hatását vizsgáljuk, akkor az előbbiek értelmében azt vizsgáljuk, hogy az összvariabilitásból milyen jelentőséggel bír az egyes csoportokon belüli egyedi variabilitás, s mennyit jelent a csoportok közötti variabilitás (a tulajdonképpeni gyógyszerhatás). Ha a kezelések (gyógyszerhatások) közötti eltérés jelentős, akkor a csoportok közötti variabilitás lesz a domináns rész a két variancia között és ilyenkor az F–próba szignifikáns eredményt ad. Azt, hogy melyik kezelések okozzák az eltéréseket a varianciaanalízis után végrehajtott post–hoc tesztek adják meg.

Attól függően, hogy hány szempont (független faktor) szerint csoportosítjuk a vizsgált változót egy és többszempontos ANOVA elrendezésekről beszélhetünk.

9.4.1. Egyszempontos ANOVA

A teszt használatának feltételei:

a. A vizsgált változó eloszlása legyen normális (az ANOVA a normalitásra robosztus, a közel normális eloszlást is “elviseli”)
b. Három vagy több független (diszjunkt) csoportunk legyen.
c. A csoportok között a variancia legyen homogén (pl. Bartlett-próba, Levene-teszt).

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

d. Legalább minimum 6 beteg adata kerüljön analízisre csoportonként.
e. Ajánlott, hogy minden csoportban az esetszám legyen azonos *(balanced)*, mert a teszt ereje ilyenkor a maximális, de *nem kritérium*. Unbalanced csoportok esetén is használható a teszt.

Ha a feltételek nem teljesülnek, akkor a megfelelő nemparaméteres statisztikát kell választani (Kruskal-Wallis teszt).

Teszteljük:

(i) az átlagértékek közötti különbség szignifikáns mértékű-e vagy sem. Formálisan a a hipotéziseknek megfelelően a következőképpen fogalmazható meg a feladat:

$$\mathbf{H_0:}\ \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_n, \text{ ha } p \geq 0.05$$

$$\mathbf{H_1:}\ \mu_1 \neq \mu_2 \neq \mu_3 \neq \ldots \neq \mu_n, \text{ ha } p < 0.05$$

(ii) a csoportok közötti variancia homogén-e (F statisztikával ellenőrizve):

$$\mathbf{H_0:}\ \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \ldots = \sigma_n^2, \text{ ha } p \geq 0.05$$

$$\mathbf{H_1:}\ \sigma_1^2 \neq \sigma_2^2 \neq \sigma_3^2 \neq \ldots \neq \sigma_n^2, \text{ ha } p < 0.05$$

A lineáris modell általános alakja:

$$y_{ij} = \mu + \alpha_i + e_{ij}$$

ahol:

y_{ij}: a függő változó értéke
μ: a kísérlet főátlaga, fix hatás
α_i: fix hatás
e_{ij}: hiba, vagy eltérés

A vizsgálat elrendezését az alábbi táblázat mutatja

1. csoport	2. csoport	3. csoport	k. csoport
x_{11}	x_{12}	x_{13}	x_{1k}
x_{21}	x_{22}	x_{23}	x_{2k}
x_{31}	x_{32}	x_{33}	x_{3k}
...	...	...	...

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

	$x_{N1}1$	$x_{N2}2$	$x_{N3}3$	$x_{Nk}k$
Mintaszám	N_1	N_2	N_3	N_k
Összeg	$\sum_{i=1}^{N_1} x_{i1}$	$\sum_{i=1}^{N_2} x_{i2}$	$\sum_{i=1}^{N_3} x_{i3}$	$\sum_{i=1}^{N_k} x_{ik}$
Átlag	$\bar{x}_1$	$\bar{x}_2$	$\bar{x}_3$	$\bar{x}_k$

A táblázatban az első index a csoporton belüli elemet, a második index a csoportot azonosítja. A teljes mintaszámot az $N=N_1+N_2+.....+N_k$ kifejezés, a teljes mintára vonatkozó átlagot (Grand Mean) az $\bar{x}$ jelöli a továbbiakban.

A mintára vonatkozó teljes variabilitást, az egyes mintaelemeknek a nagy átlagtól való eltérésének négyzetösszegeként definiáljuk (Total Sum of Squares)

$$SS(\text{teljes}) = \sum_{j=1}^{k}\sum_{i=1}^{N_j}\left(x_{ij} - \bar{x}\right)^2$$

ahol k index a csoportszámot, N_j a csoport elemszámot jelöli, vagyis x_{ij} a j–edik csoportban az i–edik elemet azonosítja.

Egyszerű számolással igazolható, hogy a teljes négyzetösszeg két részre bontható: egy csoporton belüli (Within–group) és egy csoportok közötti (Between–group) négyzetes összegre

$$SS_T = SS_W + SS_B$$

ANOVA táblázat

Source of variaton	SS (Sum of Suares)	df (Degrees of Freedom	MS (Mean Squares)	F	p
Between	SS_B^2	g-1	s_B^2	$F = \frac{s_B^2}{s_W^2}$	
Within	SS_W^2	N-g	s_W^2		
Total	$SS_B^2 + SS_W^2$	N-1			

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

ahol az egyes értékek

a) a j–edik csoport mintaelemeinek összege

$$\sum_{i=1}^{N_j} x_{ij} = S_j$$

b) a teljes minta összege

$$\sum_{j=1}^{k}\sum_{i=1}^{N_j} x_{ij} = S$$

c) teljes mintára vonatkozó négyzetes összeg

$$SS_T = \sum_{j=1}^{k}\sum_{i=1}^{N_j}\left(x_{ij} - \bar{x}\right)^2 = \sum_{j=1}^{k}\sum_{i=1}^{N_j} x_{ij}^2 - \frac{S^2}{N}$$

d) csoportokon belüli négyzetes összeg

$$SS_W = \sum_{j=1}^{k}\sum_{i=1}^{N_j}\left(x_{ij} - \bar{x}_j\right)^2 = \sum_{j=1}^{k}\sum_{i=1}^{N_j} x_{ij}^2 - \sum_{j=1}^{k}\left(\frac{S_j^2}{N_j}\right)$$

e) csoportok közötti négyzetes összeg

$$SS_B = \sum_{j=1}^{k} N_j \cdot \left(\bar{x}_j - \bar{x}\right)^2 = \sum_{j=1}^{k}\left(\frac{S_j^2}{N_j}\right) - \frac{S^2}{N}$$

A fenti számítások egyaránt használhatók egyenlő vagy nem egyenlő mintaszámok esetén.

Egyenlő elemszámú csoportok esetén az ($n=N_1=N_2=.....=N_k$)

$$SS_W = \sum_{j=1}^{k}\sum_{i=1}^{N_j} x_{ij}^{\ 2} - \frac{\sum_{j=1}^{k} S_j^2}{n}$$

és

$$SS_B = \frac{\sum_{j=1}^{k} S_j^2}{n} - \frac{S^2}{N}$$

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

9.4.2. Kétszempontos variancianalízis ismétlés nélkül

Kétszempontos varianciaanalízis esetén a vizsgált paramétert két szempont (faktor) hatásaként értékeljük. Azt vizsgáljuk, hogy az egyes faktoroknak van–e hatása az értékek alakulására. Tekintsük azt az elrendezést, amikor a faktorok által meghatározott cellákban csak egy érték szerepel. Az oszlop kezelési hatásnak **c** szintje, a sor kezelési hatásnak *r* szintje legyen:

1. faktor

	1	2	3	...	c	
1	x_{11}	x_{12}	x_{13}	...	x_{1c}	$\bar{x}_{1.}$
2	x_{21}	x_{22}	x_{23}	...	x_{2c}	$\bar{x}_{2.}$
3	x_{31}	x_{32}	x_{33}	...	x_{3c}	$\bar{x}_{3.}$
...	...	...	...	...	...	...
r	x_{r1}	x_{r2}	x_{r3}	...	x_{rc}	$\bar{x}_{r.}$
	$\bar{x}_{\bullet 1}$	$\bar{x}_{\bullet 2}$	$\bar{x}_{\bullet 3}$	...	$\bar{x}_{\bullet c}$	$\bar{x}_{\bullet\bullet}$

Az $\bar{x}_{23}$ a 2. sorban és 3. oszlopban álló értéket, az $\bar{x}_{3.}$ a 3. sor átlagát, az $\bar{x}_{2.}$ a 2. oszlop átlagát, az $\bar{x}_{..}$ az N megfigyelés átlagát jelöli (grand mean). Az 1. faktort blokknak is nevezi az irodalom. A teljes négyzetösszeg a következő alakban írható

$$\sum_{i=1}^{r}\sum_{j=1}^{c}(x_{ij}-\bar{x}_{..})^2$$

Ezt a négyzetösszeget három részre particionálhatjuk: sorok szerinti négyzetösszeg, oszlop szerinti négyzetösszeg, interakciós vagy residuális négyzetösszegre.

Az értékeknek a grand–mean–től való eltérésére igaz

$$(x_{ij}-\bar{x}_{..})=(x_{i.}-\bar{x}_{..})+(x_{.j}-\bar{x}_{..})+(x_{ij}-\bar{x}_{i.}-\bar{x}_{.j}+\bar{x}_{..})$$

Emeljük mindkét oldal négyzetre, végezzük el az összegzést i és j index–ek szerint és egyszerűsítsük a kifejezést

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

$$\sum_{i=1}^{r}\sum_{j=1}^{c}(x_{ij}-\bar{x}_{..})^2 = c\cdot\sum_{i=1}^{r}(\bar{x}_{i.}-\bar{x}_{..})^2 + r\cdot\sum_{j=1}^{c}(\bar{x}_{.j}-\bar{x}_{..})^2 + \sum_{i=1}^{r}\sum_{j=1}^{c}(x_{ij}-\bar{x}_{i.}-\bar{x}_{.j}+\bar{x}_{..})^2$$

Az első tag tehát a sor szerinti négyzetösszeg, amely a sorátlagokban fellépő variabilitást határozza meg. A második tag az oszlop szerinti négyzetösszeg, amely az oszlop átlagokban fellépő variabilitást határozza meg. A harmadik tag az interakciós tag. A számítási formulákat egyszerűsíthetjük. Vezessük be a következő jelöléseket: $S_{i.}$ az i–edik sorösszeget, $S_{.j}$ a j–edik oszlopösszeget, S_{ij} az i–edik sorban és j–edik oszlopban álló értéket, S az N megfigyelés összegét jelenti. Ennek megfelelően az előbbi formulákat a következőképpen írhatjuk:

Sor szerinti négyzetösszeg = $\frac{1}{c}\cdot\sum_{i=1}^{r}S_{i.}^2 - \frac{S^2}{N}$

Oszlop szerinti négyzetösszeg = $\frac{1}{r}\cdot\sum_{j=1}^{c}S_{.j}^2 - \frac{S^2}{N}$

Interakció = $\sum_{i=1}^{r}\sum_{j=1}^{c}x_{ij}^2 - \frac{1}{c}\cdot\sum_{i=1}^{r}S_{i.}^2 - \frac{1}{r}\cdot\sum_{j=1}^{c}S_{.j}^2 + \frac{S^2}{N}$

Teljes négyzetösszeg = $\sum_{i=1}^{r}\sum_{j=1}^{c}x_{ij}^2 - \frac{S^2}{N}$

A kétszempontos variancianalízis matematikai modellje:

$$x_{ij} = \mu + \alpha_i + \beta_j + I_{ij} + \varepsilon_{ij}$$

ahol

μ: a teljes minta átlagértéke

α_i: i–edik sorhatás $\left(\sum_i \alpha_i = 0\right)$

β_j: j–edik oszlophatás vagy blokkhatás $\left(\sum_j \beta_j = 0\right)$

I_{ij}: az i–edik sor és j–edik oszlop interakciója (a két faktor közötti interakció, amit 0–nak tételezünk fel)

ε_{ij}: hibatag (normális eloszlású valószínűségi változó 0 átlaggal az σ^2 varianciával).

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Ennek megfelelően x_{ij} is normális eloszlású valószínűségi változó μ átlaggal és σ^2 varianciával. Az ANOVA tábla

Forrás	Négyzetösszeg (SS)	df	Variancia (MS)	F
Sorok	$c\cdot\sum_{i=1}^{r}(\bar{x}_{i.}-\bar{x}_{..})^2=S_r$	r–1	$s_r^2=\frac{S_r}{r-1}$	$\frac{s_r^2}{s_{rc}^2}$
Oszlopok	$r\cdot\sum_{j=1}^{c}(\bar{x}_{.j}-\bar{x}_{..})^2=Sc$	c–1	$s_c^2=\frac{S_c}{c-1}$	$\frac{s_c^2}{s_{rc}^2}$
Interakció	$\sum_{i=1}^{r}\sum_{j=1}^{c}(x_{ij}-\bar{x}_{i.}-\bar{x}_{.j}+\bar{x}_{..})^2=S_{rc}$	(r–1)·(c–1)	$s_{rc}^2=\frac{S_{rc}}{(r-1)(c-1)}$	
Total	$\sum_{i=1}^{r}\sum_{j=1}^{c}(x_{ij}-\bar{x}_{..})^2=S_t$	r·c–1		

Az F–próbánál a két szabadságfok a számláló és a nevező szabadságfokával azonos.

Az analízis során két nullhipotézist tesztelünk.

$H_0^{(1)}$: minden sorátlag egyenlő

$$\bar{x}_{1.}=\bar{x}_{2.}=\bar{x}_{3.}=\ldots=\bar{x}_{r.}$$

$H_0^{(2)}$: minden oszlopátlag egyenlő

$$\bar{x}_{.1}=\bar{x}_{.2}=\bar{x}_{.3}=\ldots=\bar{x}_{.c}$$

9.4.3. Kétszempontos varianciaalízis ismétléssel

Azokat a kísérleti elrendezéseket, amelynek során a vizsgálatot kétszer vagy többször megismételjük, ismétléses eljárásoknak nevezzük. A variancianalízis modellje ebben az esetben

$$x_{ijk}=\mu+\alpha_i+\beta_j+I_{ij}+\varepsilon_{ijk}$$

ahol x_{ijk} a k–adik megfigyelési érték az i–edik sor és j–edik oszlophatásra vonatkozóan, ε_{ijk} a hozzátartozó hibatag. Az egyenlet többi tagja azonos a replikáció nélküli modell tagjaival. A vizsgálat során három nullhipotézist vizsgálunk

$H_0^{(1)}$: minden sorátlag egyenlő

$$\bar{x}_{1.}=\bar{x}_{2.}=\bar{x}_{3.}=\ldots=\bar{x}_{r.}$$

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

$H_0^{(2)}$: minden oszlopátlag egyenlő

$\bar{x}_{.1} = \bar{x}_{.2} = \bar{x}_{.3} = \ldots = \bar{x}_{.c.}$

$H_0^{(3)}$: nincs kereszthatás a faktorok között

$I_{ij} = 0$

A varianciaanalízis táblázata

Forrás	Négyzetösszeg	df	Variancia (MS)	F
Sorok	$nc\sum_{i=1}^{r}(\bar{x}_{i..} - \bar{x}_{...})^2 = S_r$	r–1	$s_r^2 = \frac{S_r}{r-1}$	$\frac{s_r^2}{s_e^2}$
Oszlopok	$nr\sum_{j=1}^{c}(\bar{x}_{.j.} - \bar{x}_{...})^2 = S_c$	c–1	$s_c^2 = \frac{S_c}{c-1}$	$\frac{s_c^2}{s_e^2}$
Interakció	$n\sum_{i=1}^{r}\sum_{j=1}^{c}(\bar{x}_{ij.} - \bar{x}_{i..} - \bar{x}_{.j.} + \bar{x}_{...})^2 = S_{rc}$	(r–1)(c–1)	$s_{rc}^2 = \frac{S_{rc}}{(r-1)(c-1)}$	$\frac{s_{rc}^2}{s_e^2}$
Residuál	$\sum_{i=1}^{i}\sum_{j=1}^{j}\sum_{k=1}^{n}(x_{ijk} - \bar{x}_{ij.})^2 = S_e$	rc(n–1)	$s_e^2 = \frac{S_e}{rc(n-1)}$	
Total	$\sum_{i=1}^{i}\sum_{j=1}^{j}\sum_{k=1}^{n}(x_{ijk} - \bar{x}_{...})^2 = S_t$	nrc–1		

A korábbi interakcióval kapcsolatos észrevételeknek megfelelően az ellenőrzést mindig az interakció szignifikanciájával kezdjük:

a) Ha ez nem szignifikáns akkor a $H_0^{(1)}$ és $H_0^{(2)}$ hipotéziseket ellenőrizhetjük az adott α szignifikancia érték mellett. Elvben létezik egy korrekciós hányados

$$s_k^2 = \frac{S_{rc} + S_e}{(r-1)(c-1) + rc(n-1)}$$

amivel az F–próbánál a nevezőben szereplő s_e^2–t helyettesíteni lehet. Ezt ellenőrzésképpen használjuk, ha a két nullhipotézis p értéke nagyon a szignifikanciaszint határán van.

b) Ha az érték szignifikáns, akkor a $H_0^{(1)}$ és $H_0^{(2)}$ hipotézisükre vonatkozó eredményt óvatosan kezeljük.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Ilyen esetben érdemes a hipotéziseket oly módon vizsgálni, hogy mindkét hipotézisre vonatkozó F–értéket más módon állítjuk elő

$$H_0^{(1)} \text{ esetén: } F = \frac{s_r^2}{s_{rc}^2} \quad \text{és} \quad H_0^{(2)} \text{ esetén: } F = \frac{s_c^2}{s_{rc}^2}$$

A szabadságfok értelemszerűen módosul mindkét F értéknél.

9.4.4. Többszörös összehasonlítási eljárások

Az egyszempontos variancianalízis végrehajtásakor arra nem kaptunk választ, vajon milyen csoportok átlagértékei között van eltérés. Ehhez páronkénti összehasonlítások szükségesek, aminek a száma *k(k –1)/2=n.* A többszörös összehasonlítás azzal a veszéllyel jár, hogy megnövekszik az elsőfajú hiba elkövetési valószínűsége.

Az elsőfajú hiba elkerülésének valószínűsége *n* számú összehasonlításnál $(1-\alpha)^n$. Megfordítva, annak valószínűsége, hogy *n* számú összehasonlítás során legalább egyszer hibásan döntünk és elkövetjük az elsőfajú hibát: $1-(1-\alpha)^n$.

A többszörös (Multiple Comparison) összehasonlításnál szükséges, hogy csökkentsük az elsőfajú hiba elkövetésének valószínűségét. Ezt az ún α–korrekciós eljárásokkal tehetjük meg. Ezeket az eljárásokat *post-hoc* teszteknek nevezik.

a) Szignifikancia szint csökkentése

Többszörös összehasonlítások esetén, ha a csoportok száma ≤10, a legegyszerűbb módja az α–korrekciónak, ha az α értékét elosztjuk az összehasonlítások számával. Ezt az ökölszabályt nyugodtan alkalmazhatjuk a gyakorlatban

b) Bonferroni-eljárás

Egyaránt alkalmazható ortogonális és nem ortogonális összehasonlításokra. Az eljárás ereje alacsony, és ne használjuk öt vagy ennél nagyobb számú csoportok összehasonlítására, mert növeli a tévesztések számát. Használható csak kijelölt párok vagy az összes lehetséges párok közötti összehasonlításra.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

c) Scheffé - eljárás

Az egyik legkonzervatívabb összehasonlító eljárás, mert kevesebb szignifikáns eltérést jelez, mint a többi eljárás. Egyaránt használható csak kijelölt párok vagy csoportok halmazának teljes összehasonlítására. Közel azonos esetszámú csoportokra működik jól. A normalitástól való eltérés és a csoportok közötti inhomogenitás kevésbé befolyásolja.

d) Dunett- eljárás

Arra a speciális esetre vonatkozó vizsgálati módszer, amikor a kísérleti elrendezésben egy kontroll és több kezelt csoport szerepel. A feladat a kezelt csoport és a kontroll csoport páronkénti összehasonlítása. A kezelt csoportok páronkénti összehasonlítása ilyenkor nem megengedett. A teszt viszonylag alacsony küszöbértékkel rendelkezik, de megfelelő erővel.

e) Holm-eljárás

A teszt lényegében a Bonferroni eljáráson alapul, annak a hipotéziseket szekvenciálisan vizsgáló és elutasító változata. A teszt nagy előnye, hogy egyaránt használható parametrikus és nem parametrikus modellekre egyaránt, továbbá nincs megszorítás a csoportok összehasonlítására vonatkozó tesztekre (csak a p értékekre van szükség az eljárás alkalmazásához).

f) Newman–Keuls eljárás

Gyakran használt szekvenciális összehasonlító eljárás. A vád vele szemben, hogy szignifikáns eltérést "fölöslegesen" is találhat. Ezért a szignifikáns eltéréseket (és nem szignifikánsokat úgyszintén) fokozott szakmai kritikával fogadjuk.

g) Tukey–eljárások

Két fajtája létezik: a HSD (Honest Significant Difference) ami egyenlő elemszámú csoportokra használható, és a nem egyenlő elemszámú csoportok összehasonlítására vonatkozó HSD for unequal N eljárás. Konzervatívabb mint a Neuman–Keuls eljárás, mivel kevesebb szignifikáns eredményt jelezhet.

h) Duncan–eljárás

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Azonos elvet használ mint a Newman–Keuls eljárás, de a használt szignifikancia szint nagyságában eltérnek. Hátránya is ebből fakad, már kisebb eltéréseknél is szignifikanciát jelez. Kerüljük a használatát.

i) LSD eljárás

Az egyik legrégebben kidolgozott összehasonlító eljárás (LSD = least significant difference). Gyakorlatilag kétmintás t–tesznek felel meg. Az összehasonlító tesztek közül a legkisebb hatékonyságú. Különösen nagyobb számú összehasonlításoknál hátrányos, kisebb eltéréseket is szignifikáns különbségeknek jelöl.

9.4.5. Kovarianciaanalízis (ANCOVA)

Élettanban gyakoriak az olyan vizsgálatok, amikor egy vizsgált változóra egy másik változó (pl. életkor) hatást gyakorol, azt befolyásolja. Az ilyen változók az ún. *kovariáns változók*. Ezek hatását nem szabad figyelmen kívül hagyni az analízis során, mert az eredmény nem lesz valós. A kovariáns hatását az *ANOVA* kezeli, az ilyen modell neve az ***ANCOVA*** (kovariancia hatásával bővített modell).

9.4.6. Randomizált faktoriális elrendezés

9.4.6.1. Randomizált blokkok

A randomizált blokkok elkészítéséhez tekintsük a következő példát: patkányokon végzünk kísérletet a súlygyarapodásra vonatkozóan. Összesen 36 állatunk van és háromfajta kezelés hatását akarjuk vizsgálni: osszuk testsúlyuk alapján két csoportba (blokkba) az állatokat, így mindegyik blokkba 18–18 állat kerül. Az egyik blokk legyen a közepes, a másik a nagy súlyú állatok blokkja. Nyilván a csoportosító változó korrelál az állatok súlyához. Ezután blokkon belül random módon osszuk szét az állatokat a három kezelés között. Egy blokkon belül így mindegyik kezelési csoporthoz 6 állat tartozik.

Az alábbi ábra egy lehetséges blokk elrendezést mutat

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

	Kezelés		
	1.	2.	3.
I. blokk	C	B	A
II. blokk	B	A	C

A randomizált blokkok analízise kétszempontos ANOVA módszerrel történik (a példa pl. replikációs módszerrel oldható meg). Igazából a kezelés hatását vizsgáljuk, ami csak akkor ad megbízható eredményt, ha az interakció kicsi (ez általában igaz is). A blokk elrendezéssel csökkentjük a hiba nagyságát, ezáltal a kezelés hatására vonatkozó F–próba érzékenyebbé és megbízhatóbbá válik. A biostatisztikai vizsgálatok során nagyon gyakori probléma, hogyan vegyük figyelembe pl. az életkornak a befolyásoló hatását. A kort zavaró (confounded) változónak nevezzük, mert nem lehet igazából tudni, hogy egy vizsgálatban egy adott hatásért a ténylegesen vizsgált faktor vagy a kor a felelős. Ilyen esetekben a randomizációs blokk segít a blokkok közötti eltérés kiszűrésében.

9.4,6.2. Latin–négyzetek

Az olyan véletlen blokk elrendezéseket, amellyel két confounding (két hibaforrás) változó hatását akarjuk kiegyenlíteni Latin–négyzeteknek nevezzük. Az elrendezés tulajdonsága, hogy minden kezelés csak egyszer fordul elő a sorokban és oszlopokban. Az ilyen elrendezést kiegyensúlyozott (balanced) elrendezésnek is nevezzük. Az elnevezés onnan ered, hogy a kezelések jelölésére a latin betűket használjuk.

Általános alakja

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Latin négyzet elrendezés

Önkéntesek csoportja	*Kezelési periódus*			
	I	*II*	*III*	*IV*
1	A	B	C	D
2	B	D	A	C
3	C	A	D	B
4	D	C	B	A

A…D – a vizsgálatban alkalmazott gyógyszerdózisok

A Latin–négyzet elrendezést három vagy magasabb szempontú ANOVA eljárásokra is alkalmazhatjuk replikációval vagy anélkül. A faktoriális kísérleteket, ha lehetséges akkor Latin–négyzet elrendezésbe kell szervezni.

9.4.6.3. Cross-over vizsgálat

Gyógyszervizsgálatoknál nagyon gyakran alkalmazott kisérleti elrendezés, amelynek során az alkalmazott készítményeket egy kimosási periódus uán felcseréljük (aki az *A* készítményt kapta először az később a *B* kezelést kapja és fordítva is igaz az eljárás), és a kezelés az új készítménnyel folytatódik tovább. A legegyszerűbb elrendezés a *2x2*-es crossover vizsgálat: 2 kezelési szekvenciát (AB és BA) és két periódust tartalmaz.

Lehetséges elrendezések

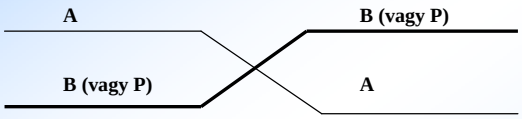
Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

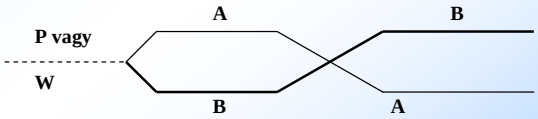
Keresztezett elrendezésű klinikai vizsgálat (cross-over)

A, B: aktív kezelés
P: placebo

a) Két kezelési típus, egyszeres keresztezés



b) Két aktív kezelés placebó vagy gyógyszer nélküli kimosási periódus bevezetéssel (W = washout)



10. Nemparaméteres eljárások

Ha egy paraméteres statisztikai eljáráshoz kapcsolódó feltételeket nem tudjuk biztosítani, akkor annak megfelelő *nemparaméteres* eljárást válasszuk. A nemparaméteres vagy eloszlásmentes (distribution free) tesztek nem igénylik a változók normalitását illetve a varianciák homogenitását, de felteszik, hogy az összehasonlítandó minták formája közel azonos. Ezek a feltételek gyengébb kritériumok mint a normalitás kritériuma. A nemparaméteres eljárások egyaránt érvényesek nominális, ordinális és intervallum skáláról származó adatokra, éppen ebből fakad a nemparaméteres eljárások népszerűsége, mivel "szabadon" használhatók. A nemparaméteres tesztek ereje gyengébb mint a neki megfelelő paraméteres teszté, ami a háttérfeltételek hiányából adódik. A nemparaméteres eljárások esetén pl. két minta összehasonlításakor nem tesztelhetjük a populáció átlagainak azonosságát, H_0: $\mu_1=\mu_2$ (H_1: $\mu_1\neq\mu_2$) mivel az eljárás eloszlásmentes. E helyett azt a hipotézist (H_0) vizsgáljuk, hogy a minták eloszlása azonos. Természetesen ha feltesszük, hogy a populációk eloszlása szimmetrikus, akkor a teszt az átlagok tesztelésére vezethető vissza a medián használatán keresztül (szimmetrikus eloszlásnál a medián és az átlag azonos). A

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

teszteket rendstatisztikáknak is hívják, mert képzésükhöz nem az eredeti adatokat használjuk, hanem az adatok növekvő sorrendbe rendezett sorszámait (rangjait).

10.1. Rangszámok tulajdonságai

A rangsorolásos eljárás viszonylag egyszerű két műveletből áll:

a) az adatokat növekvőleg nagyság szerinti sorba kell állítani

b) a rendezett adatokat a legkisebbtől kiindulva egész számokkal megszámozzuk (1, 2, 3, ..., N). Például

Eredeti adatok	5, 2, 10, 15, 7, 4, 11, 3, 9, 12
Rendezett adatok	2, 3, 4, 5, 7, 9, 10, 11, 12, 15
Rangszámok (r_i)	1, 2, 3, 4, 5, 6, 7, 8, 9, 10

Ha az adatok között azonosak is előfordulnak, akkor az ilyen értékek az ún. kapcsolt rangot (tied rank) kapják. Ez a közös rang a rájuk jutó különböző rangszámok átlaga

Eredeti adatok	1, 2, 4, 4, 4, 7, 8, 8
Kiosztott rangok	1, 2, 3, 4, 5, 6, 7, 8
Valós rang (r_i)	1, 2, 4, 4, 4, 6, 7.5, 7.5

Az adatok között 3 db 4–es szerepel, ezeknek közös rangjuk lesz, ami a különböző rangjaik átlaga

$$\frac{3+4+5}{3}=\frac{12}{3}=4$$

Hasonló a helyzet a 2 db 8–as esetén is.

$$\frac{7+8}{2}=\frac{15}{2}=7.5$$

A kapcsolt rangokhoz két megjegyzés tartozik

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

a) nem mindig egész számok

b) a nagyon sok azonos érték rontja az alkalmazott próba érzékenységét.

A rangokra vonatkozóan az alábbi műveletek érvényesek:

a) rangszámok összege

$$R = \sum_{i=1}^{N} r_i = \frac{N(N+1)}{2}$$

b) rangszámok négyzetösszege

$$R = \sum_{i=1}^{N} r_i^2 = \frac{N(N+1)(2N+1)}{6}$$

c) rangszámok átlaga és varianciája

$$\bar{x}_R = \frac{N+1}{2}$$

A nemparaméteres próbák számolástechnikailag viszonylag egyszerűek. Különösen nem számszerű adatok (pl. kérdőívek) esetén előnyös a használatuk.

10.2. Előjel teszt (sign test)

Páros minták összehasonlításának egyszerű nemparaméteres eljárása. Tulajdonképpen az egymintás *t*–teszt nemparaméteres változatának tekinthető. Az eljárás abból áll, hogy a két összetartozó minta különbségének előjelét vesszük és azt elemezzük:

a + és – előjelek különbségére kiszámítjuk a következő statisztikát

$$z = \frac{|D| - 1}{\sqrt{N}}$$

és standard normális eloszlás alapján határozzuk meg a *z*–hez tartozó *p* értéket. Ez a formula tartalmazza a folytonossági korrekciót. Mivel a binomiális eloszlás normális eloszlást követ ezért a

$$z = \frac{x - \mu}{\sigma} = \frac{x - Np}{\sqrt{Npq}}$$

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

statisztika is használható ahol *x* pl. a + jelek számát jelenti és μ = Np. A folytonossági korrekció azt jelenti, hogy az *x* értékét megnöveljük 0.5–tel ha x < Np és csökkentjük, ha x > Np, mivel az *x* diszkrét változó.

10.3. Wilcoxon párosított teszt

Igen gyakran használjuk önkontrollos (párosított minták) vizsgálatok során a két minta eltérésének tesztelésére. A vizsgálat során azt tesszük fel, hogy a minták mediánjai között nincs eltérés. Az előjel tesztnél erőteljesebb, és csaknem ugyanolyan hatékony, mint az egymintás *t*–teszt. A teszt teljes neve Wilcoxon signed–ranks teszt, mivel az adatok rangsorát és különbségük előjelét használja fel a számítások során.

10.4. Mann–Whitney U – teszt

Akkor használjuk, ha a kétmintás *t*-teszt feltételei (a normalitás vagy a varianciák homogenitása) nem teljesülnek, de a *d-teszt* helyett is használhatjuk. A teszt több lépésből áll:

1. A két minta elemeit összevonjuk, növekvő sorrendbe állítjuk és minden értékhez hozzárendeljük a megfelelő rangszámot.

2. Minden mintára meghatározzuk a csoportok rangszámainak összegét (R_1, R_2). Ha a csoportok nem azonos elemszámúak, akkor a kisebb csoportot jelölje N_1, a nagyobb létszámú csoportot pedig N_2.

3. Számoljuk ki a kisebb mintához tartozó U-statisztikát

$$U = N_1 \cdot N_2 + \frac{N_1(N_1+1)}{2} - R_1$$

ami egy szimmetrikus eloszlás. Az eloszlás átlaga és szórása:

$$\bar{x}_U = \frac{N_1 \cdot N_2}{2} \qquad s_U^2 = \frac{N_1 \cdot N_2 (N_1 + N_2 + 1)}{12}$$

Ha N_1 és $N_2 \geq 8$, akkor az U közelítőleg normális eloszlású. A standard normális eloszlás

$$z = \frac{U - \bar{x}_U}{s_U}$$

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

értéke alapján dönthetünk a H_0: $R_1 = R_2$ hipotézis felől. Ha 5%–os szinten a $-1.96 \le z \le 1.96$ reláció igaz, akkor a H_0 hipotézist elfogadjuk, egyébként elutasítjuk. A H_0 elutasítása a minták közötti szignifikáns különbséget jelenti.

A számolást az N_2–vel is elvégezhetjük

$$U = N_1 \cdot N_2 + \frac{N_2(N_2+1)}{2} - R_2$$

a többi lépés azonos az előbbiekkel. A két U_1 és U_2 statisztikára igaz, hogy

$$U_1 + U_2 = N_1 \cdot N_2 \qquad \text{illetve} \quad R_1 + R_2 = \frac{(N_1+N_2)(N_1+N_2+1)}{2}$$

azonosságok.

10.5. Kolmogorov–Szmirnov teszt

A tesztet a két minta eloszlásának tesztelésére használjuk. Azt a H_0 hipotézist ellenőrizzük, hogy a két eloszlás azonos-e. Az eljárás a két minta kumulatív eloszlásának összehasonlításán alapul.

10.6. Wald–Wolfowitz runs teszt

Wald és Wolfowitz a tesztet a véletlenszerűség vizsgálatára fejlesztette ki az ún. sorozatok (runs) használatán keresztül. A sorozat tagjait 0–val és 1–el használták, de bármilyen más jelölés alkalmazható megkülönböztetésükre. A tesztet egyébként sorozatpróbának is nevezik.

Tekintsük a 0 és 1 jelekből álló sort

$$000\,|\,11\,|\,00\,|\,1111\,|\,0\,|\,1\,|\,0\,|\,11$$

Az x és y betűk bármilyen eseményt jelenthetnek: beteg – nem beteg; kezelt – nem kezelt stb. A sorozat definíciója a következő: a sorozat egy olyan blokk, amelyben csak azonos jelek szerepelnek. Így a példában az elválasztó vonalak egy–egy blokkot definiálnak hiszen bennük vagy csak 0 vagy csak 1 szerepel. Ennek megfelelően a példasor 8 sorozatot (blokkot) tartalmaz. Nyilvánvaló, hogy a sorozatok elhelyezkedése nem csak a véletlentől függ, hiszen a következő ciklikus sor adott számú 0 és 1 esetén a legtöbb sorozatot tartalmazza:

$$0\,|\,1\,|\,0\,|\,1\,|\,0\,|\,1\,|\,0\,|\,1$$

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Azt sem tekintjük véletlennek, ha egy sor túl kevés sorozatot tartalmaz. Ha képezzük az összes olyan N jelből álló sort, amelyben N_1 számú 0 és N_2 számú 1–es van ($N = N_1 + N_2$), akkor ezek a sorok együttes eloszlása normális eloszlást követ, amely az N_1, N_2 növelésével felgyorsul. A sor sorozatainak a számát jelöljük ψ–el. A normalitás felhasználásával meghatározható a ψ eloszlás átlaga és varianciája ($N_1, N_2 \geq 5$) :

$$\bar{x}_\psi = \frac{2N_1N_2}{N_1+N_2} + 1 \qquad s_\psi^2 = \frac{2N_1N_2(2N_1N_2 - N_1 - N_2)}{(N_1+N_2)^2(N_1+N_2-1)}$$

Az eloszlás standardizálásával

$$z = \frac{\psi - \bar{x}_\psi}{s_\psi}$$

a standard normális eloszlás táblázatával a véletlen hatása tesztelhető.

Az eljárást két minta (N_1, N_2) eltérésének a vizsgálatára is felhasználhatjuk:

a) a két mintát egyesítjük ($N = N_1 + N_2$) és növekvően sorba rendezzük, majd pl. x és y–al megjelöljük az adatokat

b) a z értéke alapján döntünk a minták különbözőségéről:

H_0: a minták ugyanabból a populációból származnak

Ha a vizsgált sor véletlenszerű, akkor a minták nem különböznek, egy populációkból valók ($1.96 < z < 1.96$). Ellenkező esetben a H_0–t elutasítjuk.

10.7. k *független minta összehasonlítása*

A próbát Kruskal–Wallis féle H próbának nevezik. A Mann–Whitney vagy a Wilcoxon rank sum teszt általánosításának is tekinthetjük *k* független mintára, vagyis a módszer az egyszempontos varianciaanalízis nemparaméteres változata. A vizsgálat során a H_0 hipotézis, hogy a *k* független minta ugyanabból a populációból való. A hipotézis ellenőrzése a következő lépéseket tartalmazza

1) a megfigyelt értékeket összevonjuk, vagyis a k db N_1, N_2, N_3, ..., N_k mintákat egy mintává egyesítjük ($N = N_1 + N_2 + N_3 + ...+ N_k$)

2) az adatokat növekvőleg sorba állítjuk, majd meghatározzuk rangszámaikat (r_i) és az egyes minták rangöszegeit (R_i)

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

3) a H_0 hipotézis tesztelésére kiszámoljuk a H–statisztikát

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \left(\frac{R_i^2}{N_i} \right) - 3(N+1)$$

amely k–1 szabadságfokú χ^2 eloszlást követ. A p értékét a χ^2 táblázatból lehet meghatározni (IV. táblázat).

Abban az esetben, ha a rangok között kapcsolt rangok is előfordulnak, akkor korrekcióra van szükség és a korrekciós tag

$$1 - \frac{\sum_j T_j}{N^3 - 1}$$

ahol $T_j = t^3 - t$ és t a kapcsolt rangok száma. A H értékét elosztjuk a H_k értékkel így kapjuk meg a korrigált H értéket

$$H_k = \frac{\frac{12}{N(N+1)} \sum_{i=1}^{k} \left(\frac{R_i^2}{N_i} \right) - 3(N+1)}{1 - \frac{\sum_j^N T_j}{N^3 - N}}$$

A korrekcióval H értéke is nő. Mivel több minta összehasonlítását végezzük, szignifikáns eltérés esetén szükségünk lehet a minták páronkénti összehasonlítására, vagyis a post–hoc teszt eredményére.

A Kruskal–Wallis teszt után post–hoc tesztként a Mann-Whitney U tesztet (vagy a Wilcoxon rank sum tesztet is lehet) és a Holm–eljárást alkalmazhatjuk a megfelelő α korrekció mellett.

10.8. k *számú összetartozó minta vizsgálata*

Az ismertetésre kerülő Friedman–teszt a kétszempontos varianciaanalízis nemparaméteres változata. A mintaelemeket (a sorok jelentik az egyéneket) random módon választjuk ki, így biztosítjuk a mintaelemek függetlenségét. A kezelések (oszlopok) sorrendjét a kezelési blokkon belül szintén randomizációval határozzuk meg.

A Friedman–teszt a kezelések (oszlopok) közötti eltérést vizsgálja a következő lépésekben:

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

a) minden sorban meghatározzuk az adatok rangját (r_i) (először az adatokat növekvő sorrendbe állítjuk majd utána határozzuk meg rangjukat)

b) meghatározzuk az oszlopok rangjainak összegét (R_i)

c) kiszámítjuk az alábbi statisztikát

$$\chi_r^2 = \frac{12}{Nk(k+1)} \sum_{i=1}^{k} R_i^2 - 3N(k+1)$$

ahol

N a sorok száma

k az oszlopok száma

R_i^2 az oszlopok rangösszegének négyzete.

Az eloszlás χ^2 eloszlást követ k–1 szabadságfokkal, ha a sorok és oszlopok száma nem túlságosan kicsi. Módosított statisztikával a sorok közötti eltérést is lehet ellenőrizni.

$$x_{rk}^2 = \frac{12}{Nk(k+1)} \sum_{j=1}^{N} R_i^2 - 3k(N+1)$$

Az így meghatározott χ^2 eloszlás szabadságfoka N–1. A Kruskal–Wallis teszt után post–hoc tesztként a Mann-Whitney U tesztet (vagy a Wilcoxon rank sum tesztet is lehet) és a Holm–eljárást alkalmazhatjuk a megfelelő α korrekció mellett.

10.9. Rangkorrelációs eljárások

10.9.1. Spearman–féle rangkorreláció

A módszer a lineáris korrelációs együttható speciális esetének tekinthető. A kapcsolat szorosságának mérésére a két változó rangszámainak különbségét használjuk fel:

$$r_s = 1 - \frac{6 \cdot \sum_{i=1}^{N} d_i^2}{N^3 - N}$$

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

ahol

$d_i = x_i - y_i$ az x és y rangjainak különbsége

N = a mintaszám.

Az együttható értékei a

$$-1 \le r_s \le 1$$

intervallumba esnek: minél közelebb vannak ezek az értékek a –1–hez vagy +1–hez, annál szorosabb a kapcsolat a két változó között. A $-r_s$ estén a kapcsolatot úgyis értelmezhető, hogy a két ismérv szerinti rangsor fordított sorrendben van.

Kapcsolt rangok estén az r_s kiszámítása a következőképpen módosul

$$r_s = \frac{\frac{1}{6}\left(N^3 - N\right) - \left(T_x + T_y\right) - \sum_i d_i^2}{\sqrt{\left[\frac{1}{6}\left(N^3 - N\right) - 2T_x\right]\left[\frac{1}{6}\left(N^3 - N\right) - 2T_y\right]}}$$

ahol

$$T = \sum_{j=1}^{i} \frac{1}{12}\left(t_j^3 - t_j\right)$$

t: a kapcsolt rangok száma

j = 1, 2, 3, ..., iaz azonos rangszámú csoportok száma.

A számítás során az a hipotézis (H_0), hogy a korrelációs koeficiens 0, az alábbi t–statisztikával ellenőrizhető

$$t = r_s\sqrt{\frac{N-2}{1-r_s^2}}$$

amely N–2 szabadságfokú t–eloszlást követ. Ha az így kiszámított t > a táblázatbeli kritikus értéknél, akkor az r_s értékét a két változó kapcsolatának a jellemzésére használhatjuk. Ellenkező esetben nincs valós kapcsolat a két változó között. Az r_s és a lineáris korrelációs együttható (r) eloszlása nagy mintaszám esetén azonos (hiszen a t–statisztika is megegyezik). Ennek ellenére a két korrelációs együtthatót nem szabad egymással helyettesíteni, mert egészen más a jelentésük.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

10.9.2. Kendall–féle rangkorreláció

Kétváltozó kapcsolatát mérő τ együttható a Spearman–féle korrelációs együttható alternatívája. A számításhoz az egyes változók rang adatainak természetes sorrendjét vizsgáljuk, pl.

X:	1	2	3	4	5
Y:	1	5	3	2	4

az X rangjai természetes sorrendben szerepelnek míg az Y rangjai nem. Az Y változóban a rangok eltéréseinek a súlyát az S értéket úgy határozzuk meg, hogy minden különböző Y rangpárhoz vagy a (+1) vagy a (–1) súlyt rendeljük annak megfelelően, hogy a párok adatai természetes sorrendben vannak–e vagy sem. Pl. az Y változó esetén az (1, 5) pár (+1) az (5, 3) pár (–1) súlyt kap. Ennek megfelelően a súlyok:

+1, +1, +1, +1, –1, –1, –1, –1, +1, +1

S a súlyok összege, így az S = 2. A súlyoknak megfelelően

$S_{max} = \frac{1}{2}N(N+1) = 1$, ha minden pár súlya (+1) és

$S_{min} = -\frac{1}{2}N(N+1) = -1$ ha minden pár súlya (–1).

A τ értékét a következő formula határozza meg

$$\tau = \frac{S}{\frac{N(N-1)}{2}}$$

A τ éréke a [–1, +1] intervallumban helyezkedik el: +1 érték jelenti, hogy a rangpárok sorrendje természetes és –1 jelenti a fordított sorrendet. A példa alapján a

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

$$\tau = \frac{2}{\frac{5(5-1)}{2}} = 0.2$$

ami jelentéktelen kapcsolatra utal.

Kapcsolt rangok esetén az olyan Y pár, amelyben azonos kapcsolt rangok szerepelnek 0 súlyt kapnak, de ha két Y pár felett az X értékei kapcsoltak, szintén 0 súlyt kap az ilyen pár pl.

X:	1	2.5	2.5	4	5.5	5.5
Y:	3	2	3.5	3.5	1	6

Rendre a súlyok: –1, +1, +1, –1, +1, 0, +1, –1, +1, 0, –1, +1, –1, +1, 0 = 2 = S

Kapcsolt rang esetén a

$$\tau = \frac{S}{\sqrt{\left[\frac{1}{2}N(N-1) - T_x\right]\left[\frac{1}{2}N(N-1) - T_y\right]}}$$

képlet alapján számoljuk, ahol T_X az X változó, T_Y az Y változóinak kapcsolt rangjainak a számát jelenti:

$$T_X = \frac{1}{2}\sum_i t_i(t_i - 1) \quad \text{és} \quad T_X = \frac{1}{2}\sum_j t_j(t_j - 1)$$

A példa alapján

$$T_X = \frac{1}{2}\left[2(2-1) + 2(2-1)\right] = 2$$

$$T_y = \frac{1}{2}\left[2(2-1)\right] = 1$$

Így a τ értékére

$$\tau = \frac{2}{\sqrt{\left[\frac{1}{2}6(6-1) - 2\right]\left[\frac{1}{2}6(6-1) - 1\right]}} = 0.148$$

A τ szignifikancia érékét a

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

$$z = \frac{|S| - 1}{\sqrt{\frac{N(N-1)(2N+5)}{18}}}$$

formula alapján határozzuk meg: 5%–os szignifikancia szinten $-1.96 \le z \le +1.96$ reláció estén a H_0 hipotézist megtartjuk, ellenkező esetben elvetjük. A H_0 hipotézis az, hogy a változók között nincs valós kapcsolat. Az utóbbi példára vonatkozóan

$$z = \frac{|2| - 1}{\sqrt{\frac{6(6-1)(2 \cdot 6+1)}{18}}} = 0.188$$

így a H_0 hipotézist megtartjuk. A két változó között nincs valós kapcsolat.

A Spearmann r_s és a Kendall–féle τ korrelációs együtthatók noha azonos feladatot látnak, mégis különböznek. Ha ugyanazon az adathalmazon számítjuk ki őket, az r_s értéke nagyobb lesz mint a τ értéke. A τ számítása bonyolultabb, különösen kapcsolt rangok esetén, ezért az ilyen problémák megoldását számítógéppel végezzük. A két értéket nem lehet összehasonlítani, mert más értelemmel bírnak.

11. Regressziós vizsgálatok

Leggyakoribbak azok a vizsgálatok, amelynek során két vagy több változó közötti kapcsolatot számszerűen akarjuk kifejezni illetve azt vizsgáljuk, hogy egy vagy több változó (független változók) milyen hatással van egy kitüntetett változóra (függő változó). A kapcsolat elemzésnek az elsőfajtáját *korrelációszámításnak* az utóbbi típusát *regressziószámításnak* nevezzük. Mindegyik elemzési módszer speciális feladatot lát el, de szoros kapcsolatban is állnak egymással. A regressziószámítás a változók közötti sztochasztikus kapcsolatban lévő törvényszerűségeket, tendenciát igyekszik kifejezni függvények formájában. A korrelációszámítás a változók közötti kapcsolat erősségét vizsgálja. Nyilvánvaló, hogy a két vizsgálati módszer egymást kiegészíti: a változók közötti erős korreláció azt jelenti, hogy nyugodtan használhatjuk a regressziószámítással nyert függvényt a változók közötti kapcsolat jellemzésére, míg gyenge korreláció épp az ellenkezőjét sugallja. Attól függően, hogy

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

egyszerre hány változó kapcsolatát vizsgáljuk beszélhetünk két vagy több változós korreláció illetve regressziószámításról. Az utóbbi esetben a korrelációt többszörös korrelációnak nevezzük.

A vizsgálatokat aszerint csoportosítjuk, hogy a változók közötti kapcsolat lineáris vagy attól eltérő. Általában lineáris kapcsolatra törekszünk, mert ez a kapcsolat a legjobban érthető, ugyanakkor matematikailag is a legjobban kezelhető függvényt ad. Ha a probléma nem lineáris kapcsolatra utal, akkor különböző transzformációk segítségével megpróbáljuk azt lineárissá tenni.

11.1. Korrelációszámítás

Minden olyan esetben, amikor feladatunk két vagy többváltozó között a kapcsolat erősségének a megállapítása, korreláció-analízist kell végeznünk. Ez két fajta lehet a változók eloszlásától függően:

a) *lineáris korreláció:* a változók normális eloszlásúak,
b) *nemlineáris korreláció:* a változók nem normális eloszlásúak.

A korrelációs együttható értéke *[-1, +1]* tartományban van, és *-1* a maximális negatív, *+1* a maximális pozitív korrelációs kapcsolatot, a *0* közeli érték a korrelálatlanságot (de nem függetlenséget) jelenti a változók között. A lineáris korrelációs együtthatók közül a *Pearson*-féle *r* együtthatót, a nemlineáris korrelációs együtthatók közül a *Spearman*-féle ρ együtthatót használjuk leggyakrabban a kapcsolatok mérésére. A korreláció-számítás szoros kapcsolatban van a regressziós eljárással, gyakran együtt is használjuk őket.

Általánosan az alábbi hipotéziseket vizsgáljuk:

H₀: nincs korrelációs kapcsolat az *x* és *y* változók között vagy H_0: $r = 0$.

H₁: van kapcsolat az *x* és *y* változók között vagy H_1: $r \neq 0$

11.1.1. Kovariancia

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Két egymástól különböző valószínűségi változó közös (együttes) eloszlására jellemző érték, amely megadja a két változó együttmozgását. Tulajdonképpen a várható értékektől vett eltérések szorzatának várható értékét fejezi ki

$$Cov(X,Y) = M[(X-M(X))\,(Y-M(Y))]$$

vagy

$$Cov(X,Y) = M(XY) - M(X)\,M(Y)$$

Értéke

$$Cov(X,Y) = \frac{\sum_{i=1}^{N} (x_i - \bar{x}) \cdot (y_i - \bar{y})}{N-1}$$

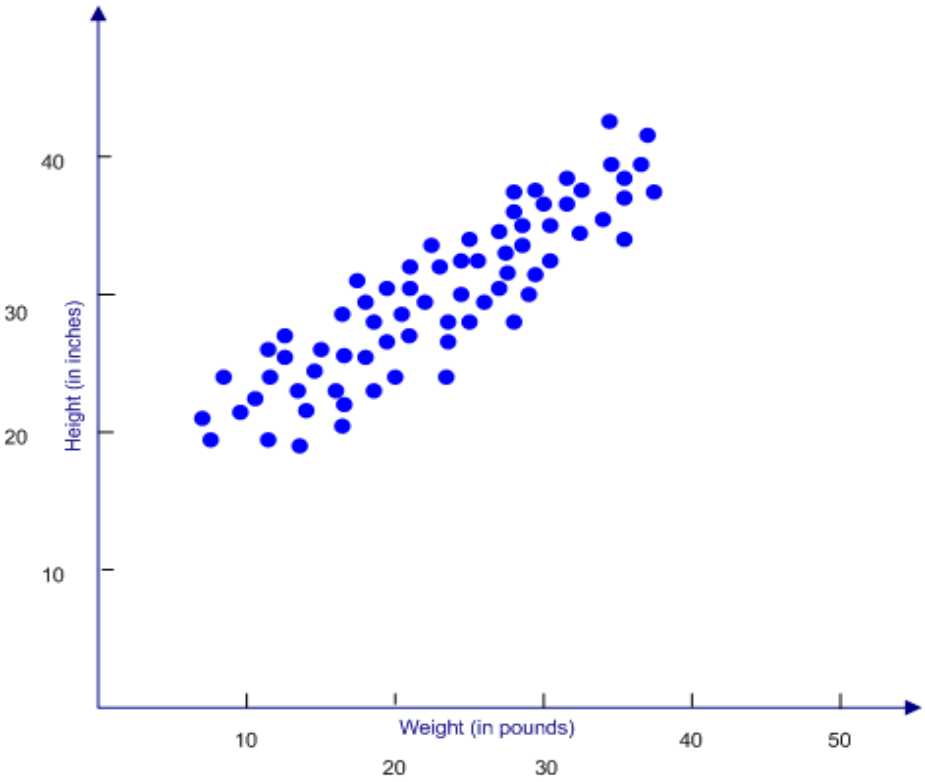
A függvény értékkészlete: (-∞ , ∞). Ha a kovariancia pozitív, akkor a két változó átlagosan ugyanabba az irányban tér el a saját átlagától, X növekedésével átlagosan Y is nő, ha negatív az X növekedésével Y csökken. Ha X=Y, akkor Cov(X,Y) = Var (X)

11.1.2. Lineáris korreláció

Az alábbi scatter-plot (felhő diagram) ábra az *X* és *Y* változók kapcsolatait mutatják

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.



A lineáris korrelációs vagy *Pearson–féle* együttható értékét a következő módon számítjuk

$$r = \frac{\sum_{i=1}^{N} (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N} (x_i - \bar{x})^2 \cdot \sum_{i=1}^{N} (y_i - \bar{y})^2}}$$

ahol $\bar{x}$ az x_i értékek, $\bar{y}$ az y_i értékek átlagai.

Számításának feltételei:

a) Az X és Y változók legyenek normális eloszlásúak.

b) Az összes kovariancia legyen lineáris.

c) Az X és Y értékeket egymástól függetlenül mérjük.

d) Kiugró (outliers) értékek erőteljesen befolyásolják *r* értékét.

Megjegyzések:

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

a) Tulajdonképpen a változók közötti kapcsolat jellemzésére a kovariancia értéke is megfelelő lenne, de nagyságát a változók értékei befolyásolják. Így a korrelációs együtthatók nem válnának összehasonlíthatóvá. Ezért szerepel a nevezőben standardizáló tényezőként a két változó szórásának szorzata: így lesz az *r* értéke standardizált érték és válik összehasonlíthatóvá.

b) Akkor értelmes, ha X és Y kapcsolata (jó közelítéssel, az adott tartományon belül) lineáris. Ha más természetű a kapcsolat, a korrelációs formula akkor is csak a lineáris komponensét méri.

c) Ha $r = 0$, (illetve ha *r* nem különbözik szignifikánsan a 0-tól) akkor korrelálatlanságról (nem függetlenségről) beszélünk.

d) A korrelációs értékeket $r \geq 0.7$ felett mondjuk erős kapcsolatnak, de az *r* értéknek valós tartalmát a szakmai megfontolások adják valójában.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

11.1.3. Determinációs együttható

Az r^2 értéket nevezzük determinációs együtthatónak, amely két variancia hányadosaként írható fel

$$r^2 = \frac{s_{y'}^2}{s_y^2}$$

ahol

$s_{y'}^2$: Y varianciájának az a része, amit az X megmagyaráz

s_y^2 : Y teljes varianciája.

Hasonlóan írható fel a korreláció szimmetriája miatt X–re is

$$r^2 = \frac{s_{x'}^2}{s_x^2}$$

Az r^2 értéke tehát azt fejezi ki, hogy az X változó a Y varianciájának hány %–át magyarázza (hány százalékért a felelős). Minél magasabb ez az érték, annál szorosabb a két változó között a kapcsolat. Az r^2 értéke 0 és 1 közötti szám és $r^2 < |r|$.

11.1.4. Korrelációs együttható szignifikanciája

Ha vesszük az *X* és *Y* változók összes populációbeli *N* számú mintáját, akkor az így kapott sokaságot kétváltozós sokaságnak nevezzük, amelyről feltételezzük a kétváltozós normális eloszlást. E kétdimenziós normális eloszlás korrelációját az elméleti korrelációs együttható méri, amit ρ – val jelölünk. A mintából meghatározott *r* ennek a ρ elméleti korrelációs együtthatónak a becslése. A ρ értéke a [–1, 1] intervallum. Az *r* eloszlása nem szimmetrikus eloszlás, a ρ – t a *–1, 0, +1* értékek kivételével csak jól közelíti. Az *r* eloszlása épp a végpontok miatt ferde eloszlás, ami $\rho = 0$ estén válik szimmetrikussá.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Az *r* szignifikancia értékének ellenőrzésére *t*–statisztikát használhatunk

$$t = r \cdot \sqrt{\frac{N-2}{1-r^2}}$$

N – 2 szabadságfokkal. Szignifikáns eltérés esetén a H_0: ρ = 0 hipotézist elvetjük és az *r* értékét valós kapcsolatnak minősítjük. A mintaszám (N) erőteljesen befolyásolja, hogy eldönthető-e a korreláltság.

Döntés a t értéke alapján:

- Ha $t<t_{krit}$, akkor H_0-t elfogadjuk, vagyis az *r* érték nem különbözik szignifikánsan a 0-tól.
- Ha $t>t_{krit}t$, akkor H_0-t elvetjük az adott szignifikanciaszinten. Ez esetben *r* olyan mértékben különbözik 0-tól, amit az adott mintaelemszám mellett a mintavételi hiba már ritkán okoz.

A ρ ≠ 0 vagy ρ = ρ_1 (ρ ≠ 0 populációs korrelációs érték) hipotézisek tesztelésénél az *r* eloszlása aszimmetrikus, de az ún. Fisher–féle Z transzformációval normális eloszlást kapunk

$$z = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right)$$

10–es alapú logaritmus használata esetén

$$z = 1.1513 \cdot \lg\left(\frac{1+r}{1-r}\right)$$

Az eloszlás átlaga és szórása

$$\mu_z = \frac{1}{2}\ln\left(\frac{1+\rho_1}{1-\rho_2}\right) \qquad \text{és} \qquad \sigma_z = \frac{1}{\sqrt{N-3}}$$

A z értékét a korrelációs együttható konfidencia intervallumának a meghatározására is felhasználhatjuk, amely 5%–os szignifikanciaszinten:

alsó érték: $z_A = z - \frac{1.96}{\sqrt{N-3}}$

felső érték: $z_F = z + \frac{1.96}{\sqrt{N-3}}$

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Az adatokat visszatranszformálva kapjuk az r_A és r_F értékeket

$$r_A = \frac{e^{2 \cdot Z_A} - 1}{e^{2 \cdot Z_A} + 1} \quad \text{és} \quad r_F = \frac{e^{2 \cdot Z_F} - 1}{e^{2 \cdot Z_F} + 1}$$

A *Z* transzformáció segítségével két korrelációs együttható (r_1, r_2) eltérésének szignifikanciáját is tesztelhetjük a

$$z = \frac{z_1 - z_2 - \mu_{z_1} - \mu_{z_2}}{\sigma_{z_1 - z_2}})$$

képlet alapján, ahol

μ_{z_1}, μ_{z_2} : az r_1 és r_2 együtthatók z eloszlásbeli átlagai

$\sigma_{z_1 - z_2}$: az r_1 és r_2 együtthatók z eloszlásbeli szórásainak különbsége:

$$\sigma_{z_1 - z_2} = \sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}$$

11.1.5. Többszörös korreláció

Ha kettőnél több változó kapcsolatrendszerét vizsgáljuk, akkor lineáris többszörös korrelációról beszélünk, de ebből nem érzékelhető a többi változó befolyásoló hatása a másik változóra. Ilyen esetben ahhoz, hogy az X_1 és X_2 változók között a kapcsolatot más változó(k) hatásától megtisztítsuk, a zavaró hatást el kell távolítani. Erre szolgál a *parciális korreláció*, amely két változó kapcsolatát úgy vizsgálja, hogy a többi változó hatását konstansnak tekinti.

Legyen három változónkX, X_2 és X_3, a közöttük lévő korrelációk r_{12}, r_{13} és r_{23}. Az r_{12} hatásából az X_3 hatását a következő módon szűrjük ki (elsőrendű parciális korrelációs együttható)

$$r_{12.3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{\left(1 - r_{13}^2\right)\left(1 - r_{23}^2\right)}}$$

Az $r_{12.3}$ együtthatót parciális együtthatónak nevezzük: az 12.3 index–ben a pont utáni szám jelenti azt a változót, amelynek hatását kiszűrjük. Az $r_{12.3}$ a reziduálok közötti korrelációt jelenti, az X_3 hatásának kiszűrése után.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

A parciális korrelációs együttható szignifikanciáját, a H_0: $r_{12.3} = 0$ hipotézist, a következő statisztikával ellenőrizhetjük.

$$t = \frac{r_{12.3}}{\sqrt{\frac{1 - r_{12.3}^2}{N - 3}}}$$

amely df = N – 3 szabadságfokú t–eloszlást követ.

Az R^2 determinációs együtthatót a kétváltozós r^2 –hez hasonlóan értelmezzük

$$R^2 = \frac{s_{1 \cdot 23}^2}{s_1^2}$$

ahol

$S_{1 \bullet 23}^2$: az x_1 varianciájának az a része amit az x_2 és x_3 változók együttesen magyaráznak

s_1^2: az x_1 változó teljes varianciája.

11.2. Lineáris regresszió

Kétváltozó közötti kapcsolatot becslő regressziós függvény alakja

$$\hat{y} = a + bx$$

ahol

$\hat{y}$: a függő változó

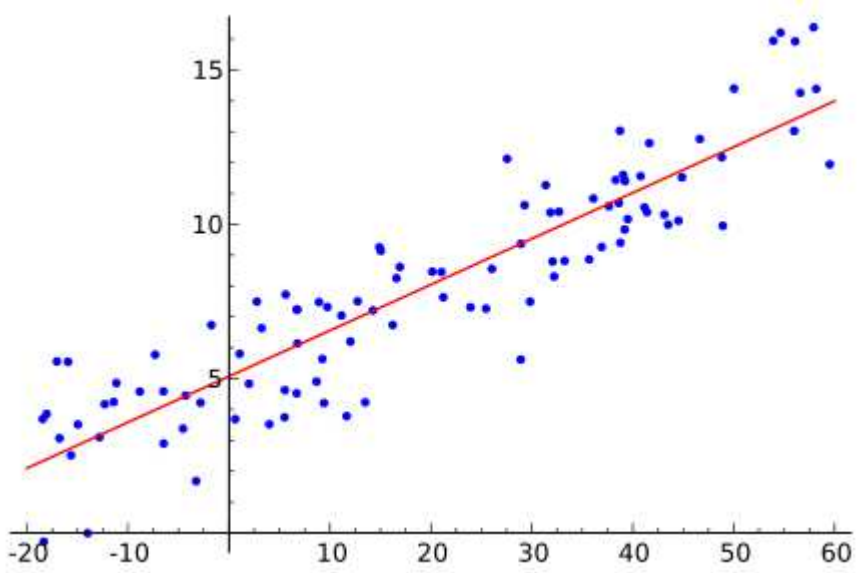
x: a független változó

a: az y tengely metszete

b: az egyenlet meredeksége (az α szög tangense).

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.



A regressziószámítás feltétele, hogy az Y változó eloszlása legyen normális és a minta legyen random módon kiválasztva (reprezentatív). Az X változóra egyedül a hibamentes adatfelvétel a kritérium.

Az egyenes paramétereinek meghatározásakor keressük azokat az *a* és *b* értékeket, amelyek mellett a mérési pontokra a regressziós egyenes a legjobban illeszkedik. A feladatot a legkisebb négyzetek módszerével végezzük el.

Határozzuk meg az egyenlet (a, b) paramétereit, hogy az y_{D_i} rezidum értékek eltérésének négyzetösszege minimális legyen:

$$y_D = \sum_{i=1}^{N} \left(y_i - \hat{y}_i\right)^2 \quad \Rightarrow \text{minimális}$$

Helyettesítsük be az egyenletbe a regressziós függvény általános alakját

$$y_D = \sum_{i=1}^{N} \left(y_i - a + bx_i\right)^2 \quad \Rightarrow \text{minimális}$$

kifejezést kapjuk. A feltételnek eleget tevő a és b értékét szélsőérték számítással kapjuk meg

$$b = \frac{\sum_{i=1}^{N} \left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sum_{i=1}^{N} \left(x_i - \bar{x}\right)^2}$$

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

és

$$a = \bar{y} - b\bar{x}$$

A *b* paraméter jelentése: az X független változó egységnyi változása milyen nagyságú változást okoz az Y függővváltozóban. Az *a* értéke a tengelymetszet magasságát adja.

A regressziós összefüggés szignifikanciáját az ANOVA táblázat alapján vizsgáljuk. A regresszió eredményének tanulmányozását is ezzel a táblázattal kell kezdeni, ugyanis a

H_0: nincs kapcsolat X és Y változók között

H_1: van kapcsolat X és Y változók között

Ha az eredmény szignifikáns az adott α érték mellett, akkor fogadhatjuk csak el a valósnak a változók közötti kapcsolatot.

Az ANOVA táblázat felépítése

Forrás	SS	df	MS	F	p
Regresszió	$\sum_{i=1}^{N}\left(\hat{y}_i - \bar{y}\right)^2 = SS_R$	1	$\frac{SS_R}{1} = s_R$	$\frac{s_R}{s_H}$	
Reziduális (hiba)	$\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2 = SS_H$	N – 2	$\frac{SS_H}{N-2} = s_H$		
Total	$\sum_{i=1}^{N}\left(y_i - \bar{y}\right)^2$	N – 1			

Az F_{krit} kritikus értéket az F-táblázatból df = 1, N – 2 szabadságfoknál keressük Ez egy *egyoldalú* próba ($s_R^2 \geq s_H^2$), az ANOVA-hoz hasonló, tehát az 5%-os F-táblázatot kell használni, ha α=5%-os szignifikancia szinten kell döntenünk.

Ha $F > F_{krit}$ akkor elvetjük H_0-t és a *b* eltérése a 0-tól szignifikáns, és predikcióra használhatjuk a lineáris egyenletet: az *y* várt értéke adott *x* érték mellett jósolható.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Megjegyzés:

a) Az egyenlet használata csak azon a tartományon belül valid, ahol a regressziót végeztük. A kívül eső tartományokban használatával óvatosan bánni.

b) Kétféle regresszió lehetséges: vagy *x* segítségével becsüljük vagy *y* segítségevel becsüljük *x* értékét.

c) Az egyenes alakjától függően lehet: *pozitív* irányú regresszió (*x* és *y* értéke együtt nő) vagy negatív irányú regresszió (*x* értéke nő az *y* értéke csökken). Lásd pozitív és negatív korreláció.

11.3. Többváltozós lineáris regresszió

A keresett egyenlet általános alakja:

$$\hat{y} = b_0 + b_1 \cdot x_1 + b_2 x_2 + \ldots + b_n x_n$$

Az alábbi hipotéziseket vizsgáljuk:

H₀: nincs kapcsolat az x_i és *y* változók között vagy H_0: $b_i = 0$.

H₁: van kapcsolat az x_i és *y* változók között vagy H_1: $b_i \neq 0$

Az eljárás arra is választ ad, hogy az x_i változók közül melyek az *y* szempontjából fontos változók, melyek azok, amelyek tényleges befolyásolják az értékét. Ki lehet szűrni a fontos x_i változókat. A módszer használatának feltétele:

a) az x_i változók és *y* között a kapcsolat lineáris
b) x_i változók legyenek függetlenek (kollinearitás vizsgálat)

A független változók között nemcsak folytonos, hanem nominális (dummy változók) változók is megengedettek. A többváltozós vizsgálatok (több változó bevonása a vizsgálatba) értékesebb, komplexebb vizsgálat, mivel a nyerhető információ is sokoldalúbb. Azonban azt szemelőtt kell tartani, hogy több változó esetén az eredmény nehezebben értelmezhető.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

A többváltozós vizsgálatnál az egyik legfontosabb szempont, hogy x_i változók függetlenek legyenek egymástól, vagyis a változók között ne legyen kapcsolat. A problémát multikollinearitásnak nevezzük. Az egymással kapcsolatban lévő változókat ki kell hagyni a vizsgálatból. A multikollinearitás vizsgálatára a változók korrelációs mátrixának determinánsa is felhasználható: $|R| = 0$ estén a változók között a kapcsolat maximális, $|R| = 1$–nél a változók függetlenek.

A számítógépes programok kiszámolják az R^2–t és az ún. módosított R^2–t (adjusted R^2). Az R^2 jelentése ebben az esetben is az, hogy az Y varianciájának a változók hány %–át magyarázzák. A módosított R^2 érték kisebb, és megbízhatóbb mértéke a regresszió jóságának, mivel ez az érték már mintafüggetlen.

Az analízis eredményét a kiugróérték erőteljesen befolyásolhatja, hasonlóan a kevés esetszám is. Többváltozós analízisnél az esetszámra vonatkozó ökölszabály: a szükséges esetszám legalább hatszorosa legyen az X változók számának.

11.4. Nemlineáris regresszió

Olyan esetekben, amikor a függő és független változók között a kapcsolat nem lineáris, az *y* becslésére a nemlineáris regressziós eljárást alkalmazzuk. Hangsúlyozni kell, hogy a probléma megoldása bonyolultabb a lineáris problémánál, és nagy segítség, ha a kapcsolat jellegéről van előzetes információnk pl. polinommal írható le a kapcsolat, ismerjük a polinom fokszámát stb. Mivel a becslő függvény bonyolult lehet, ezért arra kell törekedni:

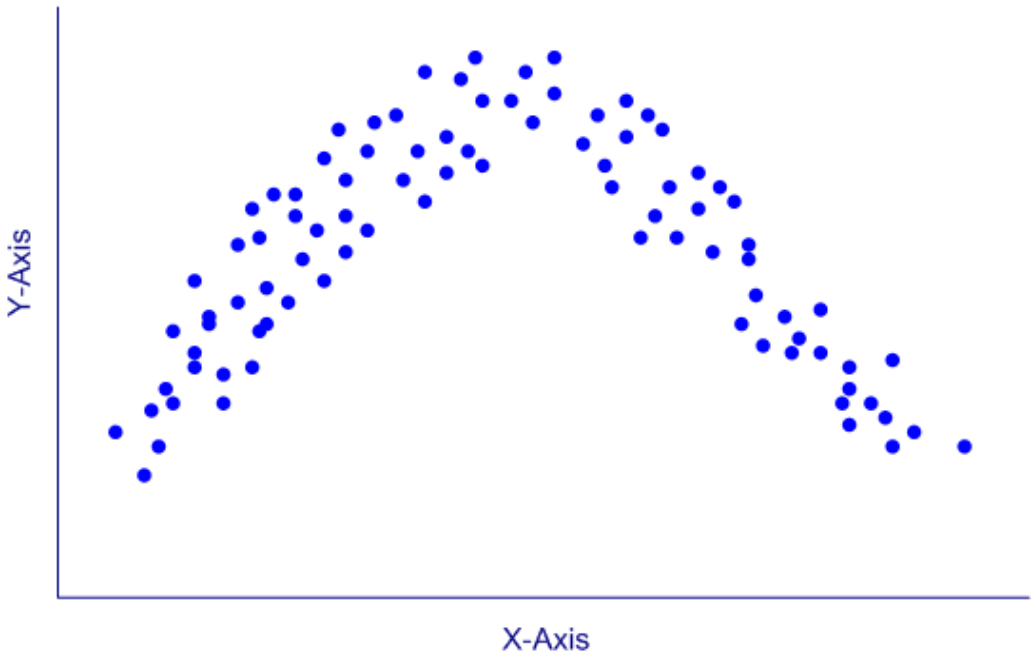
a) minél kevesebb paramétert tartalmazzon,
b) jól illeszkedjen a modell
c) a residuálisok kicsik legyenek.

H₀: nincs kapcsolat az *x* és *y* változók között.
H₁: van kapcsolat az *x* és *y* változók között.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

A feladat megoldását különböző statisztikák segítik, de előzetesen mindig ajánlatos a scatter-plot ábra tanulmányozása az analízis kezdetekor. A feladat megoldása során keressük a lehető legjobb modellt, de gyakran ennek megtalálásához több lépésben jutunk el: több modell illesztést kell elvégezni és értékelni.



12. Kontingencia táblák vizsgálata

A nominális és ordinális skáláról származó változókat, diszkrét vagy megállapítható valószínűségi változónak nevezzük. Az ilyen változók analízise más típusú vizsgálati módszereket igényelnek, mivel gyakorisági értékeket (megfigyeléseket) tartalmaznak. Ha kontingencia (gyakorisági) táblákat készítünk az adatokból és minél nagyobb a mintaszám, annál megbízhatóbb lesz a következtetésünk az ilyen táblázatok segítségével. A legegyszerűbb méretű kontingencia tábla a *2x2*-es tábla (*2* sort és *2* oszlopot tartalmaz), amely pl. influenza vírus elleni készítmény eredményét tartalmazza

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

	Készítmény		
Betegség	Kapott	Nem kapott	Total
Van	g_1	g_2	$g_1 + g_2$
Nincs	g_3	g_4	$g_3 + g_4$
Total	$g_1 + g_3$	$g_2 + g_4$	$N = g_1 + g_2 + g_3 + g_4$

Általános formában a kontigencia táblázat mérete *rxk*, és szabadságifoka a df=(r-1)·(k-1).

12.1. Pearson-féle Chi-négyzet teszt (χ^2-teszt)

A kontigencia táblák egyik leggyakoribb elemző eszköze: *függetlenség* vizsgálatra, *homogenitás* vizsgálatra, *eloszlás* vizsgálatra használhatjuk. Különösen fontosak azok az összehasonlító vizsgálatok, amelyek két független binomiális arány vizsgálatára irányul. A kontingencia táblák méretét a sorok és oszlopok száma határozza meg. Jelentőségüknél fogva a *2x2* táblák vizsgálata különösen fontos pl. diagnosztikai vizsgálatok (szenzitivitás stb.).
A khí-négyzet teszt használatának feltételei:

a) a megfigyelt értékek táblázatában bármely cella értéke lehet *0*, sőt sorok és oszlopok teljesen *0* értékűek is lehetnek,
b) a várható értékek között nem lehet *0* érték,
c) a várható értékek táblázatában az olyan cellák száma, ahol az érték *1-5* közötti, nem lehet több, mint az össz cellaszám *25%*-a,
d) a teszt ereje $N \geq 30$ mintaszámnál a legerősebb, alatta ne használjuk,
e) a teszt aszimptótikus *p* értéket ad (általában).

A kontingencia táblák analízise során a megfigyelt és a várható gyakoriságoknak az eltérését vizsgáljuk. A nullhipotézis szerint nincs eltérés az értékek közt, amit a gyakoriságokból készített Pearson–féle χ^2 statisztikával ellenőrzünk, ami kétdimenziós eloszlás esetén a

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

$$\chi^2 = \sum_{j=1}^{n}\sum_{i=1}^{k}\frac{\left(g_{ij} - e_{ij}\right)^2}{e_{ij}}$$

formulával határozható meg, ahol

g_{ij}: az i-edik sor és j-edik oszlopban lévő cella megfigyelési értéke

e_{ij}: az i-edik sor és j-edik oszlopban lévő cella várható értéke

Egy cella várható értékét úgy kapjuk meg, hogy a hozzátartozó sor– és oszlopösszeg szorzatát elosztjuk a mintaszámmal, az N – el

$$e_{ij} = \frac{g_{i.} \cdot g_{.j}}{N}$$

A statisztika értékét tehát úgy kapjuk meg, hogy minden cellára vonatkozóan a megfigyelt gyakoriságból kivonjuk a cellához tartozó várható gyakoriságot, az eredményt négyzetre emeljük, majd osztjuk a várható gyakorisággal és a kapott értékeket celláról–cellára összegezzük. Az így kapott eloszlás egy folytonos eloszlás, a jól ismert χ^2 eloszlás. Mivel a χ^2–eloszlást a szabadságfokok különítik el egymástól, így a szabadságfok meghatározása igen fontos szempont.

Kis minták esetén a χ^2–statisztika eredménye pontosítható, ha folytonossági korrekciót alkalmazunk

$$\chi^2_{korr} = \sum_{j=1}^{n}\sum_{i=1}^{k}\frac{\left(\left|g_{ij} - e_{ij}\right| - 0.5\right)^2}{e_{ij}}$$

A korrekció azt jelenti, hogy a cellák értékeiből levonunk 0.5–t. Ezt a korrekciót *Yates–féle* korrekciónak hívják. Az így módosított χ^2 érték kisebb lesz, mint a korrigálatlan érték.

Ha a változók pl. nem függetlenek egymástól, a kapcsolat szignifikáns, a két változó közötti kapcsolat erősségének megállapítására ún. szimmetrikus asszociációs mérőszámokat használunk

Kontingencia együttható:

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}}$$

Értéke [0, 1] intervallumban van: 0 a függetlenséget, 1 a "tökéletes kapcsolatot" jelenti.

Phi-együttható

$$\phi = \sqrt{\frac{\chi^2}{N}}$$

Csuprov- együttható

$$T = \sqrt{\frac{\chi^2}{N \cdot \sqrt{(k-1)(n-1)}}}$$

Értéke [0, 1] intervallumban helyezkedik el.

Cramer-együttható:

$$V = \sqrt{\frac{\chi^2}{N(k-1)}}$$

Értéke: $0 \leq V \leq 1$. A V–t egy 2x2–es tábla esetén tetrakorikus korrelációnak is nevezik.

12.2. 2x2–es kontingencia táblák

Kitüntetett szerepük van az olyan kontingencia tábláknak, amelyeket két dichotom változó határoz meg. Az ilyen táblák 2x2–es méretűek, vagyis négy cellát tartalmaznak (fourhold táblák).

Tekintsük a következő példát: egy régi és egy új diagnosztikus teszt eredményét hasonlítjuk össze. A megfigyelés eredménye általánosan a következőképpen foglalható táblázatba:

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

		B teszt		
		+	–	Total
A teszt	+	a	b	a + b
	–	c	d	c + d
	Total	a + c	b + d	N = a + b + c + d

A χ^2–statisztika kiszámítása egyszerűbb módon is lehetséges:

$$\chi^2 = \frac{N(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

A folytonossági korrekcióval módosított érték

$$\chi^2 = \frac{N\left[|ad - bc| - \frac{N}{2}\right]^2}{(a+b)(c+d)(a+c)(b+d)}$$

A szabadságfok mindkét számítási módnál=1.

12.2.1. Fisher–egzakt teszt

A tesztet azért hívják egzakt tesztnek, mert pontos *p* értéket ad, ellenben a folytonos χ^2–eloszlással, amelynek *p* értéke aszimptótikus. Minden olyan esetben amikor a mintaszám kisebb mint 30, vagy van olyan cella, amelyben a várható érték kisebb, mint 5, akkor ezt a tesztet kell használni.

Maga a számolás eléggé fáradtságos eljárás: a marginális értékek változatban hagyása mellett megkonstruáljuk az *összes* lehetséges altáblát, mindegyikhez kiszámoljuk a hozzátartozó valószínűséget és ezeket összeadjuk. Ez az érték lesz az eredeti táblázat *p* értéke. Az egyes táblák *p* értékeit a következő formulával határozhatjuk meg

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

$$p = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{N!a!b!c!d!}$$

12.2.2. Nem független minták vizsgálata

Olyan esetekben használjuk az eljárást, amikor ugyanazon személyeken végzünk vizsgálatot két különböző időpontban vagy párosított mintákat (case–control study) használunk. Pl. két orvos is megvizsgálja a betegeket és a véleményeket kontingencia táblázatba foglaljuk:

		B orvos véleménye: +	B orvos véleménye: –	
A orvos véleménye	+	a	b	$r_1 = a + b$
	–	c	d	$r_2 = c + d$
		$c_1 = a + c$	$c_2 = b + d$	

A McNemar szimmetria teszttel a főátló mellett lévő cellák egyensúlyát vizsgálhatjuk: kiegyenlített–e az eltérő vélemény az átló két oldalán. A teszt kiszámításához csak a mellékátló elemeit használjuk

$$\chi^2 = \frac{\left(|b-c|-1\right)^2}{b+c}$$

Az így kiszámított χ^2 eloszlás szabadságfoka 1.

Szignifikáns eltérés esetén a mellékátló, vagyis az eltérő vélemény a domináns. A teszt a változások irányának a tesztelésére is alkalmas. A McNemar–teszt minden k x k méretű kontingencia tábla esetén használható. Összetartozó minták esetén inkább ezt a tesztet használjuk sem mint a Pearson–féle χ^2–próbát.

A számítógépes programok az ilyen típusú kontingencia táblák vizsgálatakor az ún. κ–együtthatót (kappa), a megegyezési arányt is meghatározzák (a főátló hatását vizsgáljuk). A κ értéke [-1, 1] közötti szám. A vélemények hasonlósága 0.4 alatti κ érték esetén gyenge, 0.4 –

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

0.75 között jó és 0.75 felett nagyon jónak mondható. A κ–együttható szintén minden k x k méretű kontingencia táblára meghatározható. A κ értéke a következő módon számolható:

$$\kappa = \frac{p_0 - p_e}{1 - p_e}$$

ahol a p_0 és p_e értéke 2x2–es tábla esetén:

$$p_0 = \frac{a+d}{N} \qquad \text{és} \qquad p_e = \frac{(a+b)(a+c)+(c+d)(b+d)}{N^2}$$

k x k tábla esetén (g_{ii} jelöli a megfigyelt a, b, c, d értékeket a táblában):

$$p_0 = \sum_i \frac{g_{ii}}{N} \qquad \text{és} \qquad p_e = \sum_i \frac{r_i c_i}{N^2}$$

A κ standard hibája

$$Se_\kappa = \sqrt{\frac{p_0(1-p_0)}{N(1-p_e)^2}}$$

és 95%–os konfidencia intervalluma

$$(\kappa - 1.96 \cdot Se_\kappa;\ \kappa + 1.96 \cdot Se_\kappa)$$

12.2.3. Likelihood-becslés

kontingencia táblák esetén függetlenség vizsgálatra gyakran alkalmazott eljárás. Értéke nagyon közel van a Pearson-féle khí-négyzet értékhez. Ha lehetséges, akkor ezt a statisztikát alkalmazzuk, mivel pontosabb *p* értéket ad

$$\mathbf{G^2 = 2 \cdot \sum_i \sum_j g_{ij} \cdot \ln\left(\frac{g_{ij}}{e_{ij}}\right)}$$

A G^2 statisztika khí-négyzet eloszlást követ.

12.3. Diagnosztikai vizsgálatok

Tekintsük az alábbi vizsgálat általános *2x2*-es kontigencia táblázatát, amelyben egy régi (Gold Standard) és egy új szűrési eljárás eredménye található

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

	Standard teszt		
Új teszt	Beteg (+)	Nem beteg (–)	Total
Beteg (+)	a	b	a + b
Nem beteg (–)	c	d	c + d
Total	a + c	b + d	N = a + b + c + d

ahol az egyes betűk jelentései:

a: az új teszttel kiszűrt betegek (valódi pozitívak)

b: tévesen kiszűrt nem beteg egyének (álpozitívak)

c: betegek, de a teszt nem jelzi (álnegatívak)

d: a teszt által nem betegnek minősített egyének (valódi negatívak)

A táblákkal kapcsolatosan az alábbi fogalmakat használjuk:

Szenzitivitás: a ténylegesen beteg egyének helyesen besorolt része

$$\text{Szenzitivitás} = \frac{a}{a+c}100$$

Specificitás: a nem beteg egyének helyesen besorolt része

$$\text{Specificitás} = \frac{d}{b+d}100$$

Besorolási pontosság: a valódi pozitív és a valódi negatív besorolások aránya.

$$\text{Pontosság} = \frac{a+d}{N}100$$

Pozitív prediktív érték: azt jelzi, hogy egy valódi pozitív teszt eredményű egyén milyen valószínűséggel beteg.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

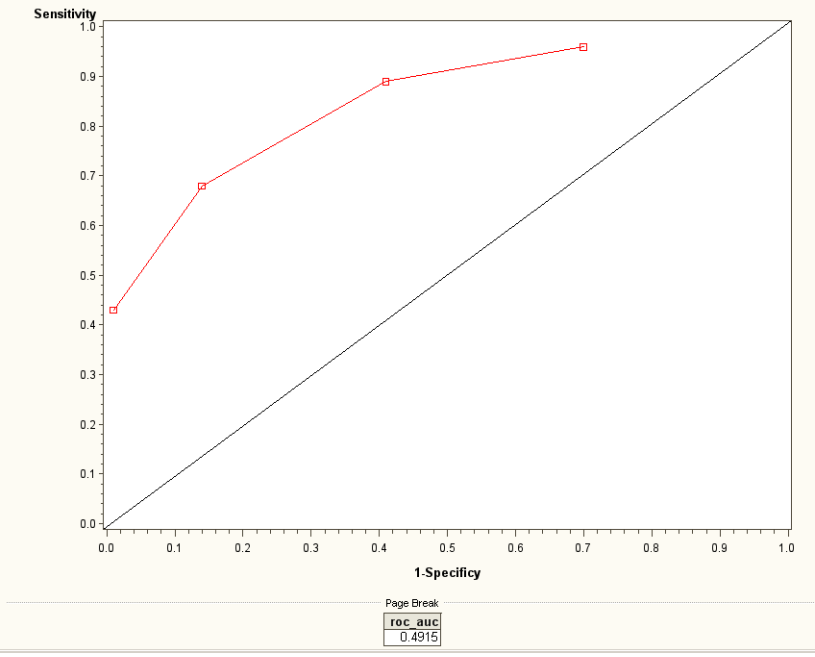
$$\text{Pozitív prediktív érték} = \frac{a}{a+b}100$$

Negatív prediktív érték: egy valódi negatív teszt eredményű egyén milyen valószínűséggel mentes a betegségtől.

$$\text{Negatív prediktív érték} = \frac{d}{c+d}100$$

12.3.1. ROC analízis

Diagnosztikus vizsgálatoknál (2 x 2-es táblák), arra is lehetőség van, hogy együttesen vizsgáljuk a specificitás, szenzitivitás együttes alakulását. Erre a feladatra a ROC (Receiver Operating Charasterics Curve) analízis szolgál. Az átfedés mértéke az osztópont (cutoff point) megválasztásától függ az X–tengelyen. A cutoff point mozgatásával változtatni tudjuk a szenzitivitás és a specificitás arányát. Az osztópontok révén 2x2–es táblázatokhoz jutunk, amelyből meghatározhatjuk a már ismert diagnosztikus paramétereket. A szenzitivitás és a specificitás kapcsolatát grafikusan is ábrázolhatjuk a ROC görbe (Receiver Operating Characteristic) megrajzolásával, amely igen szemléletesen mutatja a viszonyokat.



Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Egy teszt akkor hatásos, ha a görbe a bal felső sarokba koncentrálódik, mert ilyenkor a szenzitivitás és a specificitás is magas. Ha a görbe az átlóhoz közeli, akkor a teszt hatástalan, nem tudja a csoportokat szétválasztani.

A ROC görbe nagyon hasznos olyan esetekben, amikor több diagnosztikus tesztet kell összehasonlítani. Ilyen esetben egy ábrán ábrázoljuk a különböző tesztek ROC görbéit és a kapott ábra alapján döntünk a tesztek hatásossága felől. A másik lehetőség a görbe alatti területek összehasonlítása módosított Wilcoxon rank–sum teszt révén

12.4. Epidemiológiai vizsgálatok

Az epidemiológiai tanulmányok gondosan megtervezett, kontrollált vizsgálatok. Az értékelő módszerek attól függnek, hogy milyen formáját választjuk a vizsgálatoknak. Három vizsgálati módszert használhatunk:

a) Prospektív vizsgálat (prospective vagy cohort vagy longitudinal study): a vizsgálat N számú random módon kiválasztott egyénnel kezdődik. A mintát kétfelé osztjuk a kockázati tényező megléte vagy nem léte alapján és a két csoportot bizonyos ideig követjük miközben regisztráljuk a csoportokban az új megbetegedéseket. A két csoport összehasonlítása a kockázati tényező jelentőségére ad felvilágosítást.

b) Retrospektív vizsgálat (retrospective vagy case–control study): a betegséget előidéző kóroki tényezők hatását a betegség bekövetkezése után visszamenőleges értékeljük. Random módon pl. kórházi kórlapok alapján, kiválasztunk egy betegcsoportot és ugyancsak random módon hozzájuk rendelünk egy kontrollcsoportot. A kontrollcsoport mentes a betegségektől, de a rizikótényezőktől nem.

c) Cross–sectional study (prevalence study): random módon kiválasztunk N elemű mintát tekintet nélkül a vizsgált betegségre vagy az azt előidéző rizikófaktorra. Ezután a mintát csoportosítjuk a rizikófaktor és a vizsgált betegség alapján, majd a két csoport betegségének prevalenciáját összehasonlítjuk.

A vizsgálatok eredményeit az alábbi elrendezésű 2x2–es kontingencia táblázatba foglaljuk, amely az analízis alapját is adja:

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

		Betegség		
		+	–	Total
Rizikófaktor	+	a	b	a + b
	–	c	d	c + d
	Total	a + c	b + d	N = a + b + c + d

A betegség kockázatának mérésére két mutatót használunk a relatív kockázatot (Relative Risk, RR) és az esélyhányadost (Odds Ratio, OR).

A relatív kockázat azt fejezi ki, hogy a rizikófaktorral rendelkező egyénnek hányszor nagyobb az esélye a megbetegedésre, mint az ilyen rizikófaktorral nem rendelkező egyénnek.

$$RR = \frac{\text{Incidencia az exponált csoportban}}{\text{Incidencia a nem exponált csoportban}}$$

A táblázat alapján ez az érték (prospektív vizsgálatnál):

$$RR = \frac{\frac{a}{a+c}}{\frac{c}{c+d}} = \frac{a(c+d)}{c(a+b)}$$

Ha 1 az értéke, akkor a két csoport kockázata azonos, nincs különbség a betegség előfordulásának gyakoriságában. Ha az érték 0 és 1 közötti, akkor a “rizikófaktor” inkább gátolja a betegség kialakulását, míg RR > 1 értéknél a rizikófaktor és a betegség között pozitív a kapcsolat.

Az RR 95%–os konfidenciaintervalluma

$$\ln(RR) \pm 1.96\sqrt{\frac{\frac{b}{a}}{a+b} + \frac{\frac{d}{c}}{c+d}}$$

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Az odds–ratio, vagyis az esélyhányados

$$OR = \frac{\text{Esemény bekövetkezési valószínűsége}}{\text{Esemény benem következési valószínűsége}} = \frac{p}{1-p}$$

illetve

$$p = \frac{OR}{1+OR}$$

azt méri, hogy hányszor valószínűbben következik be a vizsgált esemény, mint az, hogy nem következik be. Ha a betegség eléggé ritka, akkor jól egyezik a RR–el. Meghatározása a következő formulával lehetséges:

$$OR=(a\cdot d)/(b\cdot c)$$

Ha az a, b, c, d értékek valamelyike 0, akkor mindegyik cellához adjunk hozzá 0.5–t és így számoljuk az OR értékét, illetve 95%–os konfidenciaintervallumát:

$$\ln(OR) \pm 1.96\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}$$

Retrospektív vizsgálat esetén az RR és OR értéke közel azonos:

$$RR = \frac{ad}{bc} \approx OR$$

12.5. Terápia hatásosságát kifejező tényezők

Abszolut rizikó (Absolut Risk, AR): Annak valószínűsége, hogy egy előre definiált kimenet (outcome) a vizsgálat időtartama alatt egy vizsgált személynél jelentkezik. Értéke 0 és 1 közötti szám, de százalékos formában is megadható.

Abszolút rizikócsökkentés (Absolute Risk Reduction, ARR): A kezelt és a kontroll csoport esetén megfigyelhető rizikó abszolút kölönbsége. Abban az esetben használatos, ha a kontroll csoport rizikója meghaladja a kezelt csoport rizikóját. Az ARR számítása során a kontroll csoport AR-jéből kivonjuk a terápiás csoport AR értékét. A megfogalmazás szerint az érték az összes kezelt beteg arányában jelentkezik és változik a kontrollcsoport AR értékével.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Esélyhányados (Odds Ratio, OR): A terápia hatékonyságának egyik mérési módszere. Annak az esélye, hogy a vizsgált esemény bekövetkezik a kezelt csoportban, a kontroll csoportban bekövetkező esemény esélyének százalékában kifejezése. Minél közelebb van az OR az 1-hez, annál kisebb a hatás a kezelt és a kontroll csoportban történt beavatkozás között. Ha az OR nagyobb (vagy kisebb) mint 1, akkor a kezelés hatása nagyobb (vagy kisebb) mint a kontroll csoportban észlelt hatás. A mért hatás lehet nemkívánatos (pl. halál, infarktus) vagy kívánatos hatás (pl. túlélés) egyaránt. Ha a vizsgált esemény ritkán következik be, akkor az OR megegyezik a relatív rizikóval (RR), de ha az esemény bekövetkezési gyakorisága nő, akkor az OR és az RR közötti távolság nő.

Relatív rizikó (Relative Risk, RR): A klinikai események kezelt csoportban mért gyakoriságának %-os aránya a kontroll csoportban mért gyakorisághoz képest. Az OR hasonló jelentésű, de némileg más matematikai formával bír. Minél kisebbhányadát teszi ki a terápiás csoport abszolút rizikója a kontroll csoport abszolút rizikójának, annál kedvezőbb a terápia hatása.

Relatív rizikó csökkentés (Relative Risk Reduction, RRR): A terápiás és a kontroll csoport közötti rizikó egymáshoz viszonyított csökkenése. Gyakran százalékos formában fejezik ki, aminek értelme az, hogy a kontroll csoport abszolút rizikóját 100%-nak tekintve ennyivel kisebb a kezelt csoport abszolút rizikója. Számítása: 1-RR.

Egy egység kimenet eléréséhez szükséges esetszám (Number Needed to Treat, NNT): A kezelés hatékonyságának egyik mérőszáma. Azoknak az embereknek a száma, akiket általában egy bizonyos módon kezelni kell egy bizonyos időperiódusban ahhoz, hogy 1 nemkívánatos kimenet elkerülhető legyen, vagy egy kívánatos kimenet elérhető legyen. Az NNT=1/ARR.

Egy egység pozitív kimenet eléréséhez szükséges esetszám (Number Needed to Harm, NNH): A kezelés ártalmasságának egyik mérőszáma. Azoknak az embereknek a száma, akiket általában egy bizonyos módon kezelni kell egy bizonyos időperiódusban ahhoz, hogy 1 nemkívánatos esemény bekövetkezzék.

Konfidencia intervallum (Confidence Interval, CI): A 95%-os konfidencia inretvallum (de lehet más %-os értékű is) a 95%-át tartalmazná azoknak az eredményeknek,

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

amelyeket az azonos módon megtervezett , azonos nagyságú, azonos populációval végrehajtott vizsgálatok eredményeképpen kapnánk. Ha a CI az RR (relative risk) vagy az OR (odds ratio) esetén tartalmazza az 1-t, akkor az adott hatásra vonatkozóan nincs elegendő bizonyíték. A CI használatának előnye, hogy a lehetséges hatásoknak egy sávját (range) adja meg.

Populációs járulékos kockázat (Population Attributable Risk, PAR): Arra a népegészségügyi kérdésre ad választ, hogy milyen mértékű többletincidenciához vezet az adott, vizsgált kockázati tényező a vizsgált populációban.

13. Túlélés analízis

A klinikai vizsgálatok során gyakran azt nézzük, hogy egy megfigyelés során a vizsgált esemény mennyi idő múlva következik be pl. kohorsz vizsgálatnál a tüdőrák kialakulása. Ezt a megfigyelt időt nevezzük túlélési időnek (survival time). A módszer minden olyan esetben használható, ahol valamilyen esemény (end point) bekövetkezésének az idejét vizsgáljuk.

Klinikai vizsgálatunknak célja tehát, hogy egy esemény (*outcome*) időbeli bekövetkezését figyeljük és rögzítjük minden beteg esetén az eseményig eltelt megfigyelési időt (*survival time*) valamint az egyén *státuszát* a vizsgálat lezárásáig. A *státusz* két értéket vesz fel: ***0*** = az esemény nem következik be vagy az egyén kiesett a vizsgálatból (dropout), ezek az egyének a *cenzorált* (*censored***)** egyének; ***1*** = az esemény bekövetkezett, az ilyen egyének a *nem cenzorált (complete**)*** egyének. A két nélkülözhetetlen változó mellé más változókat is felvehetünk a feladatnak megfelelően. Az megfigyelési időt évben, hónapban, napban stb is mérhetjük, de arra is lehetőség van, hogy kezdő és végdátumot is rögzítünk
A túlélés vizsgálat célja, hogy választ adjunk arra a kérdésre, hogy a beteg bizonyos időszakot milyen valószínűséggel élhet meg. A probléma az, hogy a túlélés idő nem normális eloszlású, továbbá a cenzorált időket is kezelni kell. Az ilyen feladatot a túlélés-analízissel oldjuk meg.

A túlélési vizsgálatok megkezdése előtt az alábbi szempontokat érdemes átgondolni

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu
MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

1) Mikor indítsuk a vizsgálatot, milyen mintaszámmal milyen hosszú ideig tartson, mi legyen a vizsgált esemény (end point). A szükséges mintaszám meghatározását a programok általában támogatják. Lehetőleg nagy mintát használjunk.

2) Hogyan kezeljük a drop out eseteket pl. ha valaki baleset következtében hal meg? Halottként vagy cenzorált adatként regisztráljuk a megfigyelés során az ilyen egyént? Mindkét eset megengedett, de erről még a vizsgálat előtt dönteni kell.

3) A mintát random módon válasszuk és a megfigyelések függetlenek legyenek egymástól.

4) Ismétlődő jelenségre ne végezzünk túlélés analízist.

5) A túlélési kritérium, a feltétel rendszer nem változhat meg a vizsgálat folyamán, mindenkire azonos kell hogy legyen (pl. vizsgált közben nem alkalmazhatunk más diagnosztikai eljárást, mint a vizsgálat elején).

6) A cenzorált adatok száma ne legyen nagy, mert rontja a vizsgálat értékét. A mintaszám kb. 10%–a még elfogadható arány a drop out miatt elmaradókat tekintve.

Feladat: tüdőrákos betegeket vizsgáltak, amelynek során rögzítették a halálig eltelt időt (nap), a *státuszt* (0=cenzorált, 1=halott), *kezelés fajtáját* (0=teszt, 1=hagyományos), *sejt-típust* (1=squamous, 2=adeno).
**Forrás: Prentice, R. L. (1973): Exponential survivals with censoring and explanatory variables. Biometrika, 60, 279-288.*

A feladatok megoldásához a SAS Enterprise Guide statisztikai programcsomagot használtam.

13.1. Life table (Halandósági tábla) analízis

Ennek lényege, hogy a megfigyelési időszakot különböző számú, egyenlő hosszúságú intervallumra osztjuk, az események részletes leírása intervallumokra történik, majd az eredményeket összegezzük.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Life Tables Analysis

The LIFETEST Procedure

Life Table Survival Estimates									
Interval		Number Failed	Number Censored	Effective Sample Size	Conditional Probability of Failure	Conditional Probability Standard Error	Survival	Failure	Survival Standard Error
[Lower,	Upper)								
0	100	34	3	60.5	0.5620	0.0638	1.0000	0	0
100	200	11	1	24.5	0.4490	0.1005	0.4380	0.5620	0.0638
200	300	4	1	12.5	0.3200	0.1319	0.2414	0.7586	0.0563
300	400	3	0	8.0	0.3750	0.1712	0.1641	0.8359	0.0498
400	500	2	0	5.0	0.4000	0.2191	0.1026	0.8974	0.0419
500	600	1	0	3.0	0.3333	0.2722	0.0615	0.9385	0.0337
600	700	0	0	2.0	0	0	0.0410	0.9590	0.0280
700	800	0	0	2.0	0	0	0.0410	0.9590	0.0280
800	900	0	0	2.0	0	0	0.0410	0.9590	0.0280
900	1000	2	0	2.0	1.0000	0	0.0410	0.9590	0.0280
1000	.	0	0	0.0	0	0	0	1.0000	0

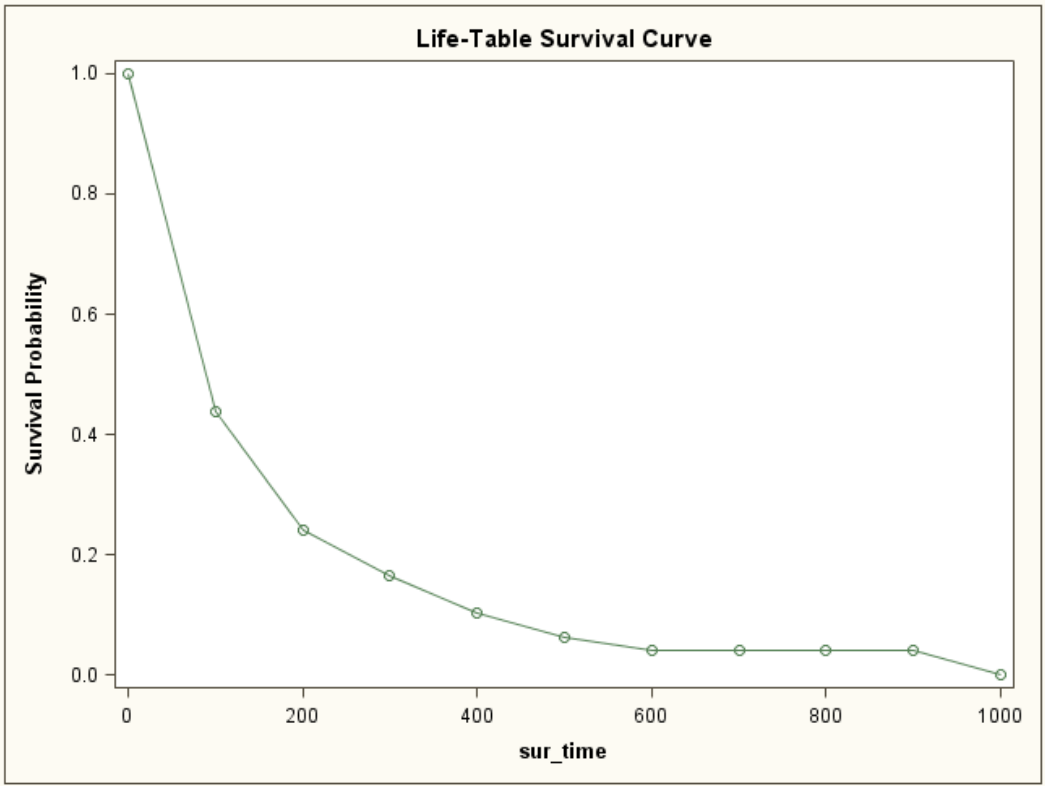
		Evaluated at the Midpoint of the Interval			
Median Residual Lifetime	Median Standard Error	PDF	PDF Standard Error	Hazard	Hazard Standard Error
88.9706	11.4385	0.00562	0.000638	0.007816	0.001234
128.9	57.2887	0.00197	0.000525	0.005789	0.001671
170.6	55.4594	0.000772	0.000366	0.00381	0.00187
150.0	70.7107	0.000615	0.000337	0.004615	0.002593
150.0	111.8	0.000410	0.000280	0.005	0.003423
425.0	43.3013	0.000205	0.000202	0.004	0.003919
350.0	35.3553	0	.	0	.
250.0	35.3553	0	.	0	.
150.0	35.3553	0	.	0	.
50.0000	35.3553	0.000410	0.000280	0.02	0
.	.	.	.	.	.

Az események leírása időintervallumokra történik: mennyi az esemény száma *(Number Failed)*, a cenzorált érték *(Number Censored)*, az aktuális mintaszám *(Effective Sample Size)* a számítások miatt nem egész érték is lehet), az intervallum túlélési valószínűsége *(Survival)*, a

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

vizsgált esemény bekövetkezésének pillanatnyi kockázatának valószínűsége *(Hazard)*, hogy az esemény éppen abban a pillanatban, az adott időintervallum közepén következik be.



Az ábra az adott időponthoz (X-tengely) tartozó túlélési valószínűségeket adja meg (Y-tengely).

Summary of the Number of Censored and Uncensored Values			
Total	Failed	Censored	Percent Censored
62	57	5	8.06

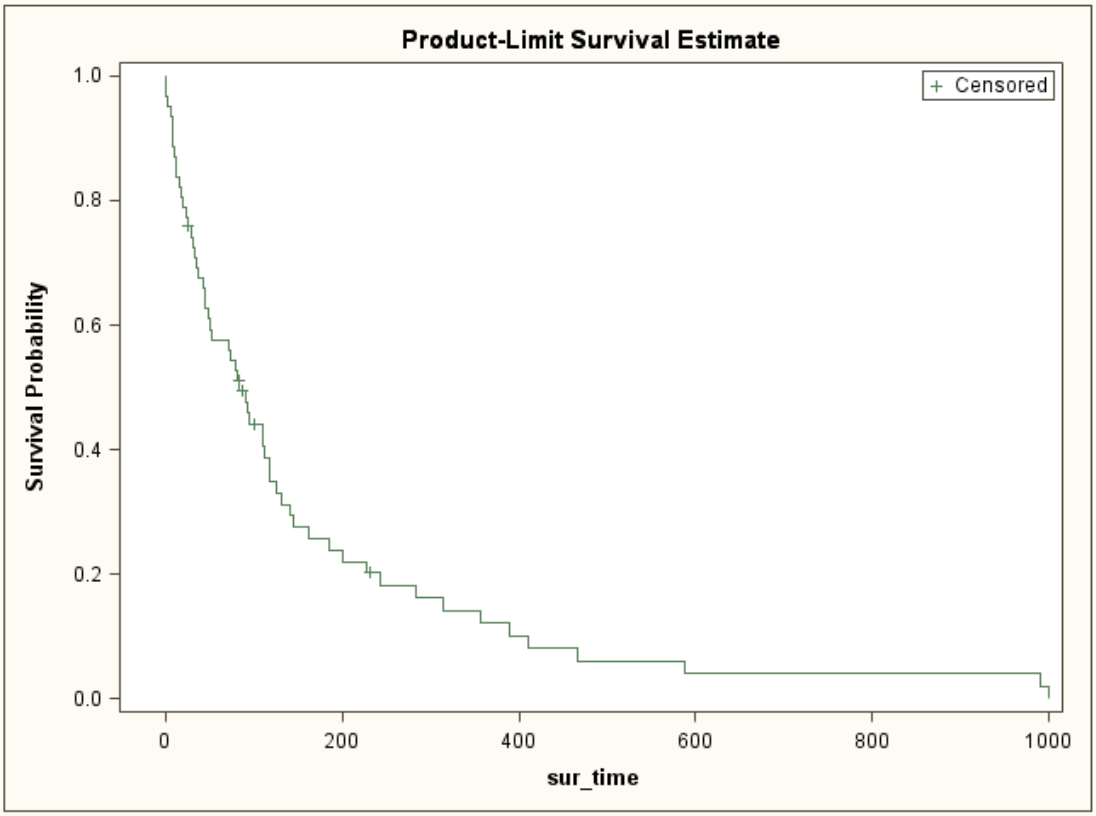
Összefoglaló statisztika a vizsgálatra vonatkozóan a *státusz* változó alapján.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

13.2. Kaplan-Meier eljárás

Az eljárást *Product-limit* módszernek is nevezik, amelynek során a jellegzetes lépcsős függvényt is megkapjuk.



Kaplan-Meier függvény

13.3. Kaplan-Meier túlélési függvények összehasonlítása. Log–rank módszer

A klinikai vizsgálatok során gyakoriak az olyan feladatok, amikor két vagy több Kaplan–Meier eljárással készített túlélési görbét kell összehasonlítani. Arra a kérdésre keressük a választ, hogy a görbék között van–e szignifikáns eltérés a sejttípus túlélését illetően. A görbére ránézve mindjárt az a benyomás alakul ki, hogy az 1-típus hatása kedvezőbb. Azt

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

viszont nem lehet egyértelműen megállapítani, hogy a két görbe eltérése szignifikáns mértékű–e.

Az egyik általánosan használt módszer a görbék összehasonlítására a log–rank vagy Mantel–Haenszel teszt. A két görbe közötti különbség kimutatására a teszt előnye akkor jelentkezik, ha a vizsgált esemény az egyik csoportban konzisztensen magasabb, mint a másik csoportban és a két csoport között a halálozási arány időben állandó.

A teszt χ^2–statisztika meghatározásán alapul, amelyet a megfigyelt esemény és a hozzátartozó várható érték alapján számolunk ki mindegyik csoportra, majd ezeket az értékeket összegezzük. A χ^2–eloszlás szabadságfoka 1. A log–rank teszt számolásához az alábbi adatok szükségesek:

t_i : a vizsgált esemény időpontja

n_1: a megfigyelt egyének száma az 1. csoportban a t_i időpontban

n_2: a megfigyelt egyének száma a 2. csoportban a t_i időpontban

n: a megfigyelések száma a t_i időpontban, $n = n_1 + n_2$

c: a cenzorált események száma a két csoportban a t_i időpontban

g_1: a megfigyelt esemény száma az 1. csoportban a t_i időpontban

g_2: a megfigyelt esemény száma a 2. csoportban a t_i időpontban

r: az össz esemény száma, $r = g_1 + g_2$

e_1: a várható esemény száma az 1. csoportban a t_i időpontban

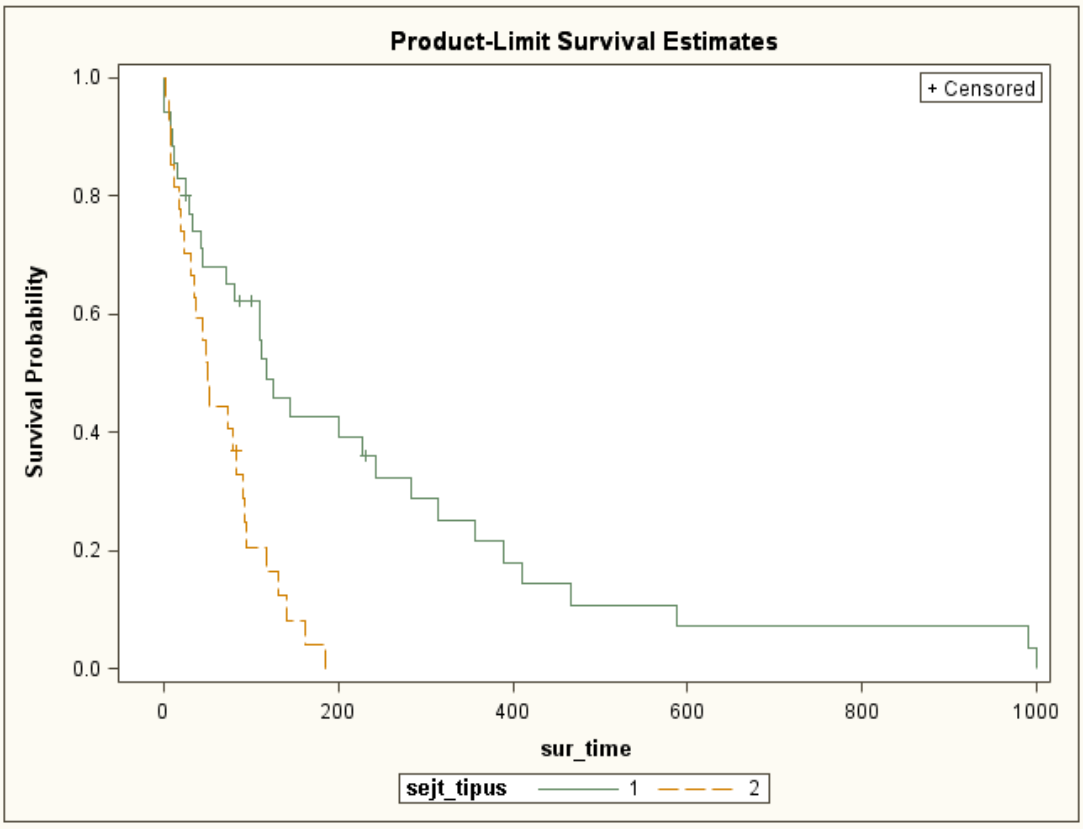
$$e_1 = \frac{r \cdot n_1}{n}$$

e_2: a várható esemény száma a 2. csoportban a t_i időpontban

$$e_2 = \frac{r \cdot n_2}{n}$$

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.



A ***log-rank*** (p = 0.0005) erős *szignifikanciát* jelez a sejt-típusok között. Tehát a két sejt-típus között jelentés eltérés van a túlélési időt illetően.

13.4. Cox-regresszió

A módszer a túlélési idő vizsgálatára alkalmas olyan esetekben, amikor több független változó befolyásoló hatásából a legmarkánsabbakat szeretnénk megismerni. A módszer a logisztikus regresszión alapszik, csak itt a függő változó a túlélési idő. Az eljárást gyakran *Cox proportional hazard* modellnek is hívják (arányos kockázati modell).

Elméletileg a túlélés analízis egyszerű eljárás, ha feltételezzük, hogy a *hazard* időben konstans. A Cox modellben a hazard időben változik, de az esemény kockázati aránya *(ratio of event hazard)* időben két személy között konstans,.Ez az ún. *proportional hazards* feltételezés. Ez azt jelenti, hogyha az életévet vizsgáljuk a modellben mint kovariánst, akkor egy 70 éves és egy 40 éves személy kockázata időben állandó. A hazard függvény:

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

$$h(t) = \lambda(t) \cdot e^{(\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k)}$$

ahol

h(t): a hazard függvény
x_i: a kovariánsok
β_i: a kovariánsok együtthatói (nagyságuk a kovariáns jelentőségét hangsúlyozzák), az

e^{β_i} az 1 egységnyi x_i változásra (a többi kovariáns konstans) adja a kockázat értékét.

$\lambda(t)$: ismeretlen kezdeti hazard függvény.

Az eljárás során tesztelt hipotézisek:

$$\mathbf{H_0}: \beta_i = 0$$

$$\mathbf{H_1}: \beta_i \neq 0$$

A fenti túlélési egyenlet az alábbi formában is írható

$$p(T > t, \underline{x}) = \left[P_0(T > t)\right]^{\exp\left(\sum_{i=1}^{n} \beta_i x_i\right)}$$

Az egyenlet a *t*–nél nagyobb időtartam (T) túlélési valószínűségét adja meg, ahol $P_0(T > t)$ az a túlélési görbe, amikor az összes kovariáns 0. A regressziós vizsgálat legfontosabb szempontja a β_i regressziós együtthatók meghatározása: ha a $\beta_i = 0$ akkor a kovariáns nincs hatással a vizsgált jelenségre. A modell egy lineáris változót vizsgált több kovariáns függvényében, így tehát valóban logisztikus modellnek is tekinthető. A modell igen kellemes tulajdonsága, hogy egyaránt használhatunk diszkrét (dummy) és folytonos kovariánsokat, sőt keverhetjük is ezeket az egyenlet jobb oldalán.

Feladat: vizsgáljuk meg, hogy a kezelés és a sejt-típus hogyan hatnak a túlélési időre.

A kovariánsok fontosságát az alábbi táblából olvashatjuk ki:

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Analysis of Maximum Likelihood Estimates								
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	
kezeles	1	0.25610	0.28410	0.8126	0.3674	1.292	0.740	2.255
sejt_tipus	1	1.04063	0.31143	11.1651	0.0008	2.831	1.538	5.212

Reference Set of Covariates for Plotting	
kezeles	sejt_tipus
0.3870967742	1.435483871

A ***kezelés*** változó nem szignifikáns *(Chi-square p = 0.3674)*, ezért nem fontos a túlélési idő szempontjából, azt nem befolyásolja. A *hazard ratio* (vagy risk ratio, RR) értéke:

$$e^{\beta_i} = e^{0.25610} = 1.292$$

A *95%*-os konfidencia intervalluma *(0.740 – 2.255*) tartalmazza az ***1*** értéket, így valóban nem jelentős változó.

A sejt_típus változó jelentősen befolyásolja a túlélési időt. A *Chi-square p = 0.0008* erősen szignifikáns, a hazard ratio értéke magas érték:

$$e^{\beta_i} = e^{1.04063} = 2.831$$

A Cox–modell alkalmazása során a következőket vegyük figyelembe:

a) Ha túl sok kovariánst veszünk be a modellbe, akkor kiderülhet, hogy a változók között kapcsolat van, ami a modell helyességét befolyásolhatja.

b) A modell feltételezi a kockázat arányának időbeli állandóságát.

c) A mintaszám megválasztásánál alkalmazzunk azt az ökölszabályt, hogy minden kovariánsra legalább 5 esemény (end point) jusson.

d) A Cox–modellt gyakran alkalmazzák az exploratív vizsgálatok során hipotézisek felállítására. Ilyen vizsgálat után a hipotézist csak másik mintán vagy mintákon szabad tesztelni.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

14. Logisztikus regresszió

Gyakoriak az olyan vizsgálatok is, amikor az *y* diszkrét értéket vesz fel: két értékű (binomiális) vagy *többértékű* (polychotomus) lehet az *y* kimenetele. A lényeges különbség az eddigi technikákhoz képest, hogy itt logit transzformált skálát használunk és az odds ratio *(OR)* használatán alapszik. A predictor változók (rizikófaktorok) eloszlása tetszőleges lehet, számukat a kívánalmaknak megfelelően bővíthetjük.
A kapott modell révén a rizikófaktor értékek ismeretében, *egyénre vonatkozóan* megtudjuk határozni a vizsgált esemény bekövetkezési valószínűségét.

Az alábbi hipotéziseket vizsgáljuk:

H₀: nincs kapcsolat az *x* és *y* változók között.
H₁: van kapcsolat az *x* és *y* változók között.

A vizsgált *Y* esemény lehet pl. a *szívinfarktus* (bekövetkezett vagy nem következett be), *transzplantáció eredménye* (a beültetett szerv kilökődött vagy nem lökődött ki) a *tüdőrák megfigyelésének az eredménye* egy prospektív vizsgálat során (kialakult a megfigyelt egyéneknél a tüdőrák vagy sem). Ilyen esetekben - az x_i független változók egyaránt tartalmazhatnak folytonos és nominális adatokat -, az *Y* esemény bekövetkezési valószínűségét logisztikus regresszióval becsüljük.
Mivel *Y* csak két értéket vehet fel, a szokásos lineáris regresszió nem alkalmazható. Ha vesszük a *p/(1-p)* kifejezést, ahol a *p* a vizsgált esemény valószínűsége, akkor ehhez az értékhez a (0, +∞) intervallum tartozik, de az *ln [p/1-p*]- hez viszont már a (–∞, +∞) intervallum. Legyen $u=[x_1, x_2, ...x_N]$ az a vektor, amely a prediktor x_i változókat (rizikófaktorokat) tartalmazza. Vizsgáljuk az *Y=1* esemény bekövetkezését logisztikus regresszióval. A regressziós modell alakja

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu
MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

$$\ln[\frac{P(Y=1|u)}{1-P(Y=1|u)}] = \ln[\frac{P(Y=1|u)}{P(Y=0|u)}] = a + \sum_{i=1}^{N} b_i x_i$$

Az ezzel ekvivalens modell

$$P(Y=1|u) = \frac{\exp\left(a + \sum_{i=1}^{N} b_i x_i\right)}{1 + \exp\left(a + \sum_{i=1}^{N} b_i x_i\right)}$$

vagy

$$\hat{P} = \frac{1}{1 + e^{-(b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_n x_n)}}$$

Ha egy prediktor változóra igaz, hogy $b_i = 0$, akkor az a faktor nincs hatással a vizsgált eseményre. Az eljárás során azt vizsgáljuk, hogy ez a feltevés igaz-e, vagyis teszteljük a H_0: $b_i = 0$ hipotézist a

$$z = \frac{b_i}{b_i \text{ standard errorja}}$$

formulával, ahol b_i a becsült regressziós együttható.

Gyakoriak az olyan vizsgálatok, amikor a prediktor változó hatását csak más zavaró (confounding) változó (pl. az életkor) hatásán keresztül értékelhetjük. A zavaró változóról tudjuk, hogy befolyással van a vizsgált eseményre, ezért figyelembe kell venni az analízis során. Ilyen esetekben a ténylegesen vizsgált rizikófaktorokat korrigáljuk (adjusted) a zavaró változó hatásával, mert csak így kapunk valós eredményt. A logisztikus regresszió alkalmas az ilyen korrekciók elvégzésére. A módszer további előnye, hogy a független változók eloszlására nincs semmi feltétel. A másik előny, hogy a regressziós koefficienseket (b_i) mint relatív kockázati értéket (relativ risk, RR) lehet felhasználni kohort, vagy odds ratio-ként (esély hányadosként, OR) case-control vizsgálatokban. Értelmezésük és számításuk azonos, pl. az odds ratio= $exp(b_i)$ kifejezéssel határozható meg.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

A számítási eljárás bonyolultabb mint a lineáris regressziónál. Általában az iteratív maximum likelihood módszert használják a számítógépes programok. A logisztikus regresszió alkalmazásánál vegyük figyelembe a következőket:

- az egyéneket egymástól függetlenül, random módon válasszuk a mintába
- legalább 5 - 10 esemény jusson mindegyik vizsgált prediktor változóra.

15. Magasabbrendű eljárások

15.1. Általános lineáris modell (GLM)

Általános lineáris modellek (General Linear Models, GLM) a többváltozós lineáris regresszió egyetlen (numerikus) függőváltozóra kiterjesztett módszere: amely számos numerikus és nem-numerikus független változó és egy numerikus függő változó közti összefüggés, kapcsolat minősítésére, számszerűsítésére szolgál, továbbá az összefüggések feltárása után, azok ismeretében történő előrejelzésre szolgál.

A modell alakja

$$\mathrm{E}(\mathbf{Y}) = \boldsymbol{\mu} = g^{-1}(\mathbf{X}\boldsymbol{\beta})$$

ahol E(Y)=μ a függő változó (**Y)** várható értéke, **X*β*** a lineáris predictor (lineáris kombinációja az ismeretlen értékű ***β*** vektornak; *g* egy link függvény.

A variancia egy függvény

$$\mathrm{Var}(\mathbf{Y}) = \mathrm{V}(\boldsymbol{\mu}) = \mathrm{V}(g^{-1}(\mathbf{X}\boldsymbol{\beta})).$$

A ***V*** exponenciális családból származó eloszlás. Az ismeretlen paramétereket tartalmazó ***β*** vektort általában maximum likelihood vagy Bayes becsléssel határozzuk meg.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Modell komponensek

1. Valószínűségi eloszlás az exponenciális családból.
2. Egy lineáris predictor: $\eta = \mathbf{X}\boldsymbol{\beta}$.
3. Egy link függvény: $E(\mathbf{Y}) = \boldsymbol{\mu} = g^{-1}(\eta)$.

Az eljárás alkalmazható bármilyen ANOVA, lineáris, logisztikus és Poisson regressziós vizsgálatokra.

15.2. MIXED modell

A lineáris modell fix, random vagy többszempontos szóráselemzés esetén kevert modell (mixed) lehet. A kevert modellt a szakirodalom általános modellnek is nevezi..A fix modellek főleg minősítő vizsgálatoknál használhatók, ahol adott feltételek mellett vizsgáljuk a hatótényezőket. A fix modellben legtöbbször kvalitatív tényezőket adunk meg. Alkalmazása főleg többszempontos szóráselemzésnél a kevert modellek felépítésénél jelentős.
A random modellnél egyaránt vizsgálhatunk kvantitatív és kvalitatív tényezőket. Ha kvantitatív tényezőket vizsgálunk, elsősorban az összefüggés milyensége (hatásgörbe) érdekel bennünket, és nem a konkrét dózisok közötti különbség. Ebben az estben jó, ha ekvidisztánsan (egyenlő távolság) vagy logaritmikusan nőnek a kezelésfokozatok. Kvantitatív tényező vizsgálata esetén keverhetjük a fix és random hatások elemzését. A random vagy fix modell alkalmazása nem csak elméleti különbség, hanem a variancia-analízis számítása során, a variancia komponensek különbözősége miatt, más számítási metódust is jelent. A hatások felderítésére szolgáló modellek tehát legtöbbször lineáris matematikai modellek. Az alkalmazott matematikai modell nagyban meghatározza a kísérlet elrendezését is, egymástól elválaszthatatlanok. Fordítva is igaz, adott elrendezéshez csak meghatározott matematikai modellek állíthatók fel.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Néhány alkalmazási terület (a teljesség igénye nélkül):

a) Általános lineáris modell illesztés az adatok normalitását feltéve:

- regresszió analízis
- varianciaanalízis (balanced or unbalanced data)
- ismételt méréses ANOVA (pl. hiányzó adatokkal)

b) Illesztett kovariancia struktúra

- variance components
- compound symmety
- factor analytic

c) Becslési eljárások

- Restricted Maximum Likelihood (REML)
- Maximum Likelihood (ML)
- Momentum módszerek (pl. Type III)

A standard lineáris model (GLM, lásd 3.5. pont) az egyik legáltalánosabb statisztikai modell:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

ahol

y: a megfigyelt adataink vektora
β: a fix-hatások ismeretlen vektora
X: ismert design mátrix
ε: residuális hiba, $N(0, \sigma^2)$

A mixed modell általánosítja a standard lineáris modellt:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$$

ahol

γ: a random hatású paraméter ismeretlen vektora
Z: ismert design mátrix

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

16. Idősoranalízis

A klasszikus klinikai biostatisztikusi munkában *idősoranalízis* feladatok ritkán fordulnak elő. Azonban vannak olyan területek pl. bizonyos élettani jelenségek (pl. vérnyomás, hormon szintek) napi (cirkadián) ritmusának elemzése (cosinor elemzés) vagy az epidemiológiában bizonyos betegségek gyakoriságának (vagy azok mortalitásának) szezonális (ciklikus) változásainak elemzése továbbá az analóg elektrofiziológiai jelek (pl. EEG) analízise (auto-, keresztkorreláció, Fourier-analízis, stb) feldolgozása ilyen jellegű feladatok. Következésképpen megismerése és alkalmazása indokolt orvosi területen is. A SAS Guide tartalmaz az idősorelemzésre vonatkozó modult, ezért röviden bemutatom annak használatát, hogy bővíthessük analitikai eszköztárunkat. Hangsúlyozom, hogy a terület nagyon komplex, megismerése mélyebb elméleti hátteret feltételez, amelynek ismertetésére jelen könyv keretei nem adnak lehetőséget.

16.1. Elméleti bevezető

Egy elméleti idősor olyan speciális *sztochasztikus* folyamat (olyan folyamat, ami a valószínűség-számításra épül és egyben a valószínűségi változó fogalom általánosításának is tekinthető, továbbá a különböző gyakorlati problémák megoldásában nagy szerepe van), ahol a valószínűségi változók

$$Y_1, Y_2, Y_3,\ldots,Y_n$$

véges sorozatát vizsgáljuk.

A fenti idősornak tehát minden egyes tagja egy-egy valószínűségi változó, amelyekre vonatkozóan azonban csak egy-egy empírikus (tapasztalati) adat áll rendelkezésünkre (egyetlen realizáció ismeretes)

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

$y_1, y_2, y_3, \ldots, y_n$

A klasszikus idősor elemzés abból a feltételezésből indul ki, hogy az idősort egy tartós, hosszú távú tendencia (trend), szabályos hullámmozgások, periodikus ingadozások (szezonalitás) határozzák meg és ezektől eseti, egyenként nem jelentős eltérítő hatást vált ki a véletlen ingadozás. Az idősor elemzés eszközei:

- *Grafikus ábrázolás:* lehetővé teszi a fő tendenciák vonások felismerését.
- *Bázis ill. láncviszonyszámok:* az idősorok gyors, előzetes elemzésére szolgál.
- *Egyszerűbb eszközök az un. átlagok:*

  Számtani átlag,

  Kronologikus átlag.

  Egy adott időszak jellemzéséhez a vizsgált időszakon kívüli megfigyelés is szükséges, de az első és utolsó megfigyelés csak fél súllyal szerepel. Képletben:

$$\overline{y}_t^k = \frac{\frac{y_1}{2} + y_2 + \ldots + y_{n-1} + \frac{y_n}{2}}{n-1}$$

Idősorok összetevői:

a) ***T****rend vagy alapirányzat:* egy határozottan jelentkező tendencia

b) ***P****eriodikus ingadozás:* rendszeresen ismétlődő hullámzás (pl háziorvosnál a napi betegforgalom)

c) ***V****életlen ingadozás:* szabálytalan mozgás

A periodikus ingadozásnak két típusa van:

- *Additív modell:* az idősor a ***T***rend hatás a ***P***eriodikus hatás és a ***V***életlen ingadozás Összege:

$$y = T + P + V$$

- *Multiplikatív modell:* az idősor érték a három tényező szorzata:

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

$$y = T * P * V$$

Idősor elemzés esetén a feladat, hogy a valódi hatást (trend) megtísztítsuk az egyéb hatásoktól:

$$Y = T + S + C + I$$

***T**rendmozgás:* általános irány egy adott hosszú időszakon belül

***S**zezonális mozgás:* időszakok szerint rendszeres mozgás

***C**iklikus mozgás:* trend körüli hosszú távú mozgások, amelyek nem feltétlenül periódikusak

***I**rreguláris mozgás:* előre nem jósolható véletlen mozgások

A simító módszerek lényege, hogy tényadatok segítségével lépésenként korrigálják a kialakított modell eredményeit: használatuk során a zavaróhatások kiszűrésével lehet valós előrejelzést adni.

16.2. Lineáris és nem lineáris trend modell

A modell használata esetén feltételezzük, hogy az adatsorban nincsenek szezonálisan ismétlődő folyamatok és periódicitás, tehát csak a trendhatás van jelen. Lineáris trend feltételezése esetén a *legkisebb négyzetek* módszer felhasználásával adjuk meg a modellben a legjobban illeszkedő egyenest illetve az ehhez tartozó konstans értéket és a meredekségi együtthatót.

Általános lineáris trendfüggvény:

$$y_t = b_0 + b_1 t$$

és együtthatói:

$$b_1 = \frac{\sum (t - \bar{t})(y_t - \bar{y})}{\sum (t - \bar{t})^2}$$

$$b_0 = \bar{y} - b_1 \bar{t}$$

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Nem lineáris esetben egy rögzített függvény (exponenciális, logaritmikus, négyzetes, stb.) transzformáció segítségével illesztjük a görbét az adatokhoz.

16.3. Exponenciális simítás

Az idősor adataiban nem tételezünk fel sem trend, sem szezonális hatást. Az exponenciális simítás lényege, hogy az előrejelzés során egy adott időponthoz tartozó értéket úgy definiálunk, hogy abban benne vannak a múltbeli értékek is oly módon, hogy időben visszafelé haladva egyre kisebb súllyal szerepelnek A súly értéke *0* és *1* közötti intervallumból származnak. Amennyiben *1*-hez közeli súlyt választunk, akkor az idősor függvénye kis mértékben simítódik ki, azaz nagy súlyt kap az aktuális érték (kis súlyt kapnak a múltbeli értékek). *Nulla* vagy *0*-hoz közeli súly választása esetén az idősoron erős simítást hajtunk végre, kiszűrjük az ingadozásokat, és eredményül hullámzó görbét kapunk. Ebben az esetben kis súlyt kap az aktuális érték, és a múltbeli értékek nagy súlyt kapnak.

Szezonális modell

A szezonális modell olyan idősorokra illeszkedik jól, melyekben nem figyelhető meg trend, azaz a trend egyenes meredeksége *0*: a jövőben nem várható növekedés, egyensúlyi helyzet alakult ki. A szezonális hatás dominánsan és szabályosan jelentkezik és nincs szükség exponenciális simításra.

a) Szezonalitás illesztése additív modellel

Feltételezve additív modell esetén a szezonalitás állandóságát, a szezonalitást kifejező komponens értéke attól függ, hogy melyik szezonban vagyunk, de attól nem, hogy az adott szezon hányadik periódusában. Így a trend határozza meg az idősor fő áramát. A szezonális hatás konstans formában járul hozzá (hozzáadódik vagy kivonódik) az adott időszakhoz tartozó trend értékhez. Ha a szezonalitást a trendtől függetlenül ábrázoljuk, akkor egy periódikusan ismétlődő függvényt kapunk, amelynek hullámhossza és amplitúdója állandó.

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

b) Szezonalitás illesztése multiplikatív modellel

Multiplikatív modell esetén a szezonális hatás nem független az idősor trend függvényének értékétől. Nagyobb trend értékhez nagyobb szezonális érték tartozik, tehát a változás (kilengés) mértéke annál nagyobb, minél nagyobb értékű a trend függvény. A kilengések trendhez viszonyított aránya nagyjából állandónak tekinthető.

Ha a szezonalitást a trend függvény nélkül ábrázoljuk, akkor egy periodikusan ismétlődő függvényt kapunk állandó hullámhosszal és változó (tehát nem állandó) amplitúdóval. Az amplitúdó attól függ, hogy a trend függvénynek az adott pillanatban mekkora az értéke.

Lineáris trend modell Brown-féle vagy Holt-féle simítással

A modell feltételezi, hogy az idősorban. lineáris trend figyelhető meg. Nem bizonyítható szezonális hatás, az ingadozások periodicitása konstans értékhez nem konvergál.

A Brown- és a Holt-féle simítás nagyon hasonlít egymáshoz, csak matematikai formulákban térnek el, amelyek megadásától eltekintünk. Mindkét eljárás az exponenciális simítás egy speciális esetét alkalmazza: egymás után kétszer hajtjuk végre a simítást.

16.4. Winters additív modell

A modell azon idősorok esetén használható jó eredménnyel, ahol lineáris trend figyelhető meg és additív jellegű a szezonális hatás. A szezonalitás additív leírása a trend lineáris voltának megkötésén kívül teljes mértékben helytálló.

Winters multiplikatív modell

A modell azon idősorokhoz illeszkedik a legjobban, ahol lineáris trend figyelhető meg, és multiplikatív a szezonális hatás. A multiplikatív szezonalitás leírása a trend lineáris voltának megkötésén kívül teljes mértékben helytálló.

16.5. Telítődési modell

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

Olyan esetekben használhatjuk akár lineáris, akár nem lineáris módon ahol az idősor egy konstans értékhez, azaz egy vízszintes egyeneshez simul. Például nulla betegforgalomról a görbe hirtelen felugrik egy adott értékig, majd ezen érték körül ingadozik a továbbiakban.

16.6. ARMA

Idősorok elemzésénél pl. jelfeldolgozásban gyakran alkalmaznak korrelációs függvényeket adatsorozatok összehasonlítására. A *keresztkorreláció* segít két adatsor közötti összefüggés megtalálásában. Ha az egyik adatsort eltoljuk (*lag)*, akkor késleltetett hatások is felfedezhetők. Az ilyen adatsorok összehasonlítását az ún. *autokorrelációval* végezzük. Autokorreláció segítségével periódusok mutathatók ki az adatsorban. Ha definiálunk két függvényt (esetünkben diszkrét véges adatsor, vagy idősor) *keresztkorrelációját* és ez alapján *konvolúcióját* (a lineáris művelet két függvényből állít elő egy harmadikat), továbbá egy függvény *autokorrelációját*: a konvolúció lényegében egy időtükrözött függvénnyel vett keresztkorrelációnak felel meg és az *autokorreláció* egy adatsornak saját magával vett keresztkorrelációjával egyenértékű.

ARMA (A**utoRegressive–Moving Average, autoregresszív-mozgó átlagolás): a** statisztikában, de különösen a jelfeldolgozásban nagyon gyakori az ilyen modellek használata, amit ***Box–Jenkins*** modellnek is neveznek. Az autoregresszív modell (lineáris előrejelző függvények) feladata a jövőbeli adatok becslése (forecasting) az előzőleg becsült adatok alapján. A modell általános megadása:

ARMA (p, q)

ahol

p-rendben autoregressziv: az autoregresszió rendj***e,*** AR (p)

q-rendben mozgóátlagú: a mozgóátlag rendje, MA (q)

Integrált Autoregresszív Mozgó Átlagolás (ARIMA, **AutoRegressive Integrated Moving Average**): általánosítása az *ARMA* modellnek. A modell (folyamat) általános megadása:

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

ARIMA (p, d, q)

ahol az egyes paraméterek nem negatív egész értékek

p-rendben autoregressziv: az autoregresszió rendj***e,*** *AR (p)*
q-rendben mozgóátlagú: a mozgóátlag rendje, *MA (q)*
d-rendben integrált: a differenciálás (integrálás) foka, *I (d)*

I (*d*): $\Delta_d Y_t$ stacioner

ahol *d* a differenciázás foka

$$\text{ARIMA}(p,d,q)\_Y_t \sim \text{ARMA}(p,q)\_\Delta_d Y_t$$

Stacionaritási-transzformációk

1. Differencia stacionaritás esetén (DSP):

   a). Elsőrendű differenciázás: $\Delta Y_t = Y_t - Y_{t-1}$
   b) Másodrendű differenciázás: $\Delta\Delta Y_t = \Delta^2 Y_t$
   c) Logaritmálás és differenciázás: $\Delta \log Y_t$

2. Trendstacionaritás esetén (TSP):
   a) Tisztítás a trendtől: Y_t –Trend(t)
3. Szezonalitás esetén:
   a) Szezonális differenciázás: $\Delta_4 Y_t = (Y_t - Y_{t-4})$
   b) Szezonális differenciázás: $\Delta_{12} Y_t = (Y_t - Y_{t-12})$
   c) Szezonális kiigazítás: Y (-/) Szezonhatás

Különböző formában is megadható a modell: amikor az egyik tag *0*, azt a modell nevében is feltüntetjük. Pl. ha a modellünk formája *I (1*), akkor a teljes modell alakja *ARIMA (0,1,0) vagy MA (1)* modell esetén a teljes modell *ARIMA (0,0,1).*

A teljesség igénye nélkül meg kell említeni a nagyon fontos idősor analízis modellt a ***Box-Jenkins*** modellt vagy metódust:

1. Modell-identifikálás: p,d,q meghatározása:

   a) Stacionaritás-vizsgálat:
   b) Transzformáció: d = ?
   c) Stacionaritás? Transzformáció ha szükséges!
   d) ARMA(p,q) rendek behatárolása

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.

2. Paraméterbecslés, Akaike modellszelekció,
3. Diagnosztika: Reziduum Fehér-Zaj-e?
4. Előrejelzés

Semmelweis Egyetem
Cím: 1085. Budapest, Üllői út 26.
Telefon: +36 (1) 459-1500
E-mail: hirek@semmelweis-univ.hu
Honlap: http://semmelweis-egyetem.hu

MAGYARORSZÁG MEGÚJUL
A projektek az Európai Unió támogatásával valósulnak meg.