

# Biostatisztika és informatika alapjai

## 3. előadás:

A valószínűségi számítás elemei

2016. szeptember 22.

Veres Dániel

## Egy kísérlet...

Az adott **betegséget** kimutató gyorssteszt:

kék: egészséges

zöld: beteg

Szeretnénk kideríteni a gyorssteszt segítségével, hogy egy kérdéses területen van-e járvány. A következőket tudjuk:

- A betegséggel nem sújtott („egészséges”) területeken:

1-2 **zöld** / 10 megvizsgált egyénre.

- A betegséggel sújtott („beteg”) területeken:

7-9 **zöld** / 10 megvizsgált egyénre

Felütötte-e a fejét a járvány az adott területen?

**A vizsgálatok számának növelésével nő a „bizonyosság”.**

**Hány mérést kell végeznünk?**

**Valamekkora bizonytalanság mindig lesz ... – De ez mekkora?**

## Alapsokaság és minta

Alapsokaság (populáció)



Az **alapsokaság** rendszerint olyan méretű, hogy az összes eleme nem vizsgálható meg.

Minta



Emiatt az alapsokaságnak csak egy részhalmazát vizsgáljuk, ezt nevezzük mintának.



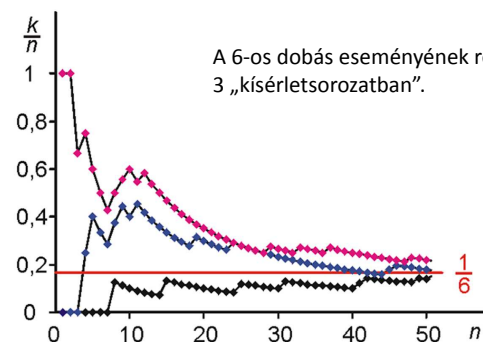
**BIZONYTALANSÁG!**

A minta jellemzői alapján az alapsokaságra vonatkozó következtetést vonhatunk le



A minta elemein méréseket végzünk, majd az így keletkező adathalmazt (amit szintén mintának nevezünk) grafikusan és matematikailag jellemezzük

## Egy másik szemszögből...

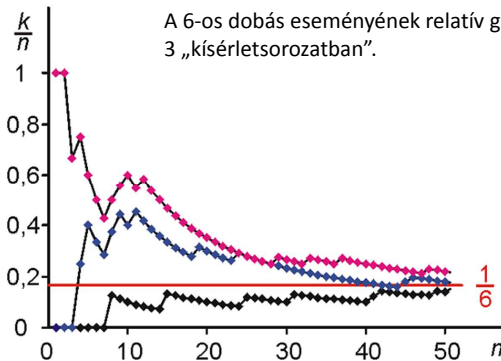


$$\text{Rel.gyak.} = \frac{N_{\text{„kedvező”}}}{N_{\text{összes}}}$$

Azt tapasztaljuk, hogy a **relatív gyakoriságok** ilyen sorozatai – bár ingadozásokat mindig mutatnak – a „kísérletsorozat” hosszának növekedtével egyre inkább **stabilizálódnak valamilyen érték körül**. Továbbá ez az érték az aktuális „kísérletsorozattól” függetlenül lényegében ugyanakkora.

## Valószínűség, mint mennyiség?

A 6-os dobás eseményének relatív gyakoriságai  
3 „kísérletsorozatban”.



$$P = \text{Rel. gyak.} = \frac{N_{\text{„kedvező”}}}{N_{\text{összes}}}$$

$$N_{\text{összes}} \rightarrow \infty$$

A **nagy számok** (relatív gyakoriságokra vonatkozó) **tapasztalati törvénye**: a relatív gyakoriság értéke egy végtelen sorozatban egy adott értékhez tart.  
Az adott **eseményhez** hozzárendelhetjük ezt az **értéket**: 6 dobáshoz az **1/6**-ot.  
Ezt az értéket nevezzük az **esemény valószínűségének**.

Ez a törvény *tapasztalati* törvény, tehát logikai úton nem bizonyítható.

## Események valószínűségei I.

**Jelölések:**

Esemény: **A**

(a páciensnek láza van)

Az A esemény bekövetkezésének valószínűsége: **P(A)**

(annak a valószínűsége, hogy a páciensnek láza van)

Ellentett esemény:  **$\bar{A}$**

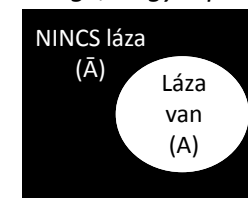
(a páciensnek **NINCS** láza)

Az A esemény be NEM következéseének valószínűsége: **P( $\bar{A}$ )** vagy

**P(nemA)**

(annak a valószínűsége, hogy a páciensnek **NINCS** láza)

Venn-diagram



## Események valószínűségei I.

**Jelölések:**

Esemény: **A**

(a páciensnek láza van)

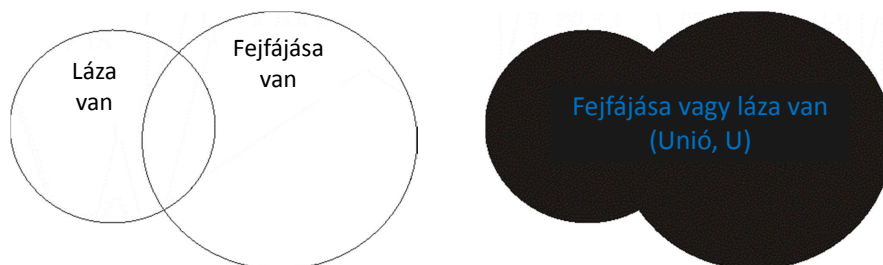
Az A esemény bekövetkezésének valószínűsége: **P(A)**

(annak a valószínűsége, hogy a páciensnek láza van)

Annak a valószínűsége, hogy A **vagy** B esemény bekövetkezik:

**P(AvagyB), P(A+B), P(AUB)**

(annak a valószínűsége, hogy a páciensnek láza **vagy** fejfájása van)



## Események valószínűségei II.

Annak a valószínűsége, hogy **A és B** esemény egyaránt bekövetkezik:

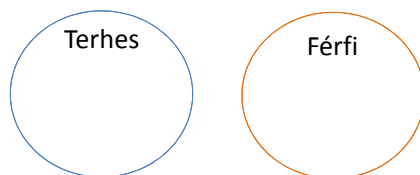
**P(AésB), P(A\*B), P(AB), P(A∩B)**

(annak a valószínűsége, hogy a páciensnek láza **és** fejfájása van)



## Események valószínűségei III.

**Egymást (kölcönösen) kizáró események:** A és B események együttesen nem következhetnek be.  
(a páciens *terhes és férfi*)  $(A \cap B) = 0$



**Független események:** A esemény bekövetkeztének nincs hatása B bekövetkezésére.  
(az első páciensünk férfi, a második nő)

## Események valószínűségei IV.

**Feltételes** valószínűség

A esemény bekövetkezésének valószínűsége tudva, hogy B esemény bekövetkezett:  $P(A|B)$ .  
(annak a valószínűsége, hogy a páciensünknek influenza fertőzése van (nem más fertőzése) tudva azt, hogy vírusos eredetű a fertőzése)

## Események valószínűségei V.

Események valószínűségének alaptörvényei (**Kolmogorov-axiómák**):

1.  $0 \leq P(A) \leq 1$

2.  $P(\text{biztos}) = 1$  (a páciens előbb vagy utóbb meghal)  
 $P(\text{lehetetlen}) = 0$  (a páciens teljesen egészséges\*)

3. **Egymást kölcsönösen kizáró** eseményekre:  $P(A \text{ és } B) = 0$   
 $P(A \text{ vagy } B) = P(A) + P(B)$   
(annak a valószínűsége, hogy páciensünk *terhes vagy férfi*)

Ezekből levezethető:

+4. **Független** eseményekre:  $P(A \text{ és } B) = P(A) * P(B)$   
(annak a valószínűsége, hogy az *első páciensünk férfi és a második nő*)

## Események valószínűségei VI.

**Feltételes** valószínűség számítása

általános forma 2 eseményre:  $P(A|B) = P(A \text{ és } B) / P(B)$

**Különleges esetek:**

I. **Független eseményekre:**

annak a valószínűsége, hogy a *második páciensünk férfi*,  
**HA** az *első nő volt*

$$P(A|B) = P(A \text{ és } B) / P(B)$$

$$P(A|B) = P(A) * P(B) / P(B)$$

$$P(A|B) = P(A)$$

annak a valószínűsége, hogy a *második páciensünk férfi*,  
**HA** az *első nő volt* = annak a valószínűsége, hogy a *második páciensünk férfi*

## Események valószínűségei VI.

### II. A esemény részhalmaza B eseménynek:

annak a valószínűsége, hogy a páciensünknek *influenza fertőzése van*  
HA ismert, hogy *fertőzése vírusos eredetű*

$$P(A|B) = P(A \text{ és } B) / P(B)$$

$$P(A|B) = P(A) / P(B)$$

Számolási példa:

Annak a valószínűsége, hogy páciensünknek vírusos fertőzése van:

$$P(B) = 8\%$$

Annak a valószínűsége, hogy páciensünknek influenza fertőzése van:

$$P(A) = 2\%$$

annak a valószínűsége, hogy a páciensünknek influenza fertőzése van  
HA ismert, hogy fertőzése vírusos eredetű:

$$P(A|B) = 2\% / 8\% = 25\%.$$

## Kockázat

|              |      | Betegség |     |          |
|--------------|------|----------|-----|----------|
|              |      | Igen     | Nem | Összesen |
| Rizikófaktor | Igen | a        | b   | a+b      |
|              | Nem  | c        | d   | c+d      |
| Összesen     |      | a+c      | b+d | a+b+c+d  |

A betegség kockázata (valószínűsége), ha van rizikófaktor:

$$P(Bet_i | Riz_i) = \frac{P(Bet_i \cap Riz_i)}{P(Riz_i)} = \frac{\frac{a}{a+b+c+d}}{\frac{a+b}{a+b+c+d}} = \frac{a}{a+b}$$

A betegség kockázata (valószínűsége), ha NINCS rizikófaktor:

$$P(Bet_i | Riz_n) = \frac{P(Bet_i \cap Riz_n)}{P(Riz_n)} = \frac{\frac{c}{a+b+c+d}}{\frac{c+d}{a+b+c+d}} = \frac{c}{c+d}$$

## Relatív Kockázat

|              |      | Betegség |     |          |
|--------------|------|----------|-----|----------|
|              |      | Igen     | Nem | Összesen |
| Rizikófaktor | Igen | a        | b   | a+b      |
|              | Nem  | c        | d   | c+d      |
| Összesen     |      | a+c      | b+d | a+b+c+d  |

### Relatív kockázat (RR, Relative Risk, Risk Ratio):

azt az arányt mutatja, hogy hányszor gyakoribb az esemény kockázata, ha jelen van a rizikófaktor, mint a rizikófaktor hiányában

$$\frac{P(Bet_i | Riz_i)}{P(Bet_i | Riz_n)} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}} = \frac{a \cdot (c+d)}{c \cdot (a+b)}$$

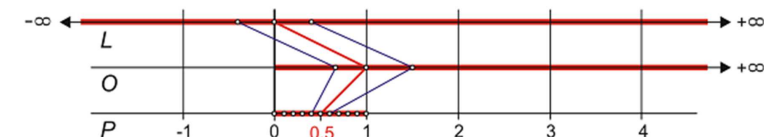
## Esély

**Esély (esélyérték; O - odds):** „hányszor akkora a valószínűsége annak, hogy az esemény bekövetkezik, mint annak, hogy nem következik be”

$$O = \frac{P(A)}{P(A)} = \frac{P(A)}{1 - P(A)}$$

**Logit (L):** esély természetes alapú logaritmusa

Logit – Esély – Valószínűség



## Esélyhányados

|              |      | Betegség |     | Összesen |
|--------------|------|----------|-----|----------|
|              |      | Igen     | Nem |          |
| Rizikófaktor | Igen | a        | b   | a+b      |
|              | Nem  | c        | d   | c+d      |
| Összesen     |      | a+c      | b+d | a+b+c+d  |

A betegség esélye, ha van rizikófaktor:

$$\frac{P(Bet_i | Riz_i)}{P(Bet_n | Riz_i)} = \frac{\frac{P(Bet_i \cap Riz_i)}{P(Riz_i)}}{\frac{P(Bet_n \cap Riz_i)}{P(Riz_i)}} = \frac{P(Bet_i \cap Riz_i)}{P(Bet_n \cap Riz_i)} = \frac{\frac{a}{a+b+c+d}}{\frac{b}{a+b+c+d}} = \frac{a}{b}$$

A betegség esélye, ha NINCS rizikófaktor

$$\frac{P(Bet_i | Riz_n)}{P(Bet_n | Riz_n)} = \frac{c}{d}$$

## Esélyhányados

|              |      | Betegség |     | Összesen |
|--------------|------|----------|-----|----------|
|              |      | Igen     | Nem |          |
| Rizikófaktor | Igen | a        | b   | a+b      |
|              | Nem  | c        | d   | c+d      |
| Összesen     |      | a+c      | b+d | a+b+c+d  |

**Esélyhányados, esélyarány (OR, Odds Ratio):**

hányszor nagyobb a betegség kialakulásának esélye a rizikófaktor meglétes estében, mint annak hiányakor

$$\frac{\left( \frac{P(Bet_i | Riz_i)}{P(Bet_n | Riz_i)} \right)}{\left( \frac{P(Bet_i | Riz_n)}{P(Bet_n | Riz_n)} \right)} = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{a*d}{c*b}$$

## Relatív kockázat és Esélyhányados

|              |      | Betegség |     | Összesen |
|--------------|------|----------|-----|----------|
|              |      | Igen     | Nem |          |
| Rizikófaktor | Igen | a        | b   | a+b      |
|              | Nem  | c        | d   | c+d      |
| Összesen     |      | a+c      | b+d | a+b+c+d  |

**OR**      **RR**

$$\frac{a*d}{c*b} \neq \frac{a*(c+d)}{c*(a+b)}$$

**A betegség ritka**

$$\begin{aligned} a &<< b \\ c &<< d \end{aligned} \quad OR \Rightarrow RR$$

## Relatív kockázat és Esélyhányados

|                   |               | Tüdődaganat |               | Összesen |
|-------------------|---------------|-------------|---------------|----------|
|                   |               | Van daganat | Nincs daganat |          |
| Dohányzási szokás | Dohányzik     | 79          | 71            | 150      |
|                   | Nem dohányzik | 9           | 18            | 27       |
| Összesen          |               | 88          | 89            | 177      |

**OR**

$$\frac{a*d}{c*b}$$

$$\frac{79*18}{9*71} = 2,23$$

**RR**

$$\frac{a*(c+d)}{c*(a+b)}$$

$$\frac{79*27}{9*150} = 1,58$$

**Jelentése? (OR, RR: R)**

R=1: „nincs rizikóhatás”

R>1: nagyobb kockázat/esély a faktorról

R>1: kisebb kockázat/esély a faktorról

**Lehet, de NEM BIZTOS**

**Mintavételi bizonytalanság!**

## Valószínűesszámítás.....

Permutációk  
Variációk  
Kombinációk

## ***Na mire is lehet jó....***

Influenzaszezont megelőzően a rendelőkben az adott napra 4 oltóanyag áll rendelkezésre. Az előző években átlagosan 2989 páciensből 402 személyt kellett beoltanunk. Az előző év alapján mekkora a valószínűsége, hogy a rendelkezésre álló 4 oltóanyag elegendő lesz és el is fogy, ha 25 embert várunk aznapra?

$$P = \binom{n}{k} \cdot (p)^k \cdot (1-p)^{(n-k)} = \binom{25}{4} \cdot \left(\frac{402}{2989}\right)^4 \cdot \left(1 - \frac{402}{2989}\right)^{(25-4)} \approx 0,2$$

## ***Na mire is lehet jó....***

Mekkora a valószínűsége annak, hogy páciensünk 3.45 mmol/l-es (normál tartományon kívüli) K<sup>+</sup> szintje még „egészséges”?

Hány szülés várható az esti ügyeletben, ha az éves statisztika 1000 szülést mutat éjfél és 8:00 között?

Az évfolyamból várhatóan hányan lesznek alkalmasak egy csípőprotézis elvégzésére (tömegük alapján)?

Vajon hat-e az adott gyógyszer?...

Az influenza/AIDS teszt pozitív – milyen valószínűséggel vagyok tényleg beteg?

..... Hogyan számoljunk? Ismerjük a „képletet”? milyen „egyenletet”, táblázatot, excel függvényt... válasszunk, mikor melyiket?

## ***Az emberi gondolkodás...***

Tomi csendes, visszahúzódozó, szerény, szorgalmas fiú, aki másoknak szívesen segít. Melyiket tartod valószínűbbnek:

- a) Tomi könyvtáros
- b) Tomi kétkezi munkás

# Az emberi gondolkodás...

Linda tehetséges, független, filozófia szakot végzett 31 éves nő. Nagyon érzékeny a társadalmi igazságtalanságokra. Diákként részt vett az antinukleáris demonstrációkban. Sorszámozza meg az alábbi állításokat aszerint, hogy mennyire tartja valószínűnek (1-es sorszám a legvalószínűbb):

- a) Linda tanító egy általános iskolában,
- b) Linda könyvesboltban dolgozik, és jóga tanfolyamra jár,
- c) Linda a nőszavazók ligájának tagja,
- d) Linda bankpénztáros,
- e) Linda biztosítási ügynök,
- f) Linda bankpénztáros és feminista.

# Valószínűség másképp...

## Ellenőrző kérdések #1

- Definiáld a valószínűséget a nagy számok törvénye alapján.
- Ismertesd a nagy számok törvényét.
- Hogyan bizonyítható a nagy számok törvénye?
- Hogyan jelölhető az A vagy B esemény bekövetkezésének valószínűsége?
- Hogyan jelölhetjük azt a valószínűséget, hogy A és B esemény egyaránt bekövetkezik?
- Mit jelent két esemény metszete, illetve uniója?
- Definiáld az egymást kölcsönösen kizáró eseményeket.
- Mondj példát az egymást kölcsönösen kizáró eseményekre.
- Mit tudsz az egymást kölcsönösen kizáró események metszetéről?
- Definiáld az egymástól független események fogalmát.
- Mondj példát független eseményekre.
- Mit jelent a feltételes valószínűség?
- Mondj példát a feltételes valószínűségekre.
- Hogyan jelöljük a feltételes valószínűséget?
- Hogyan számítható  $P(A)$  ha  $P(A|B)$  és  $P(B)$  adott?
- Mik a Kolmogorov axiómák?
- Mit tudsz A és B esemény viszonyáról, ha  $P(A \vee B) = P(A) + P(B)$  igaz?
- Mit tudsz A és B esemény viszonyáról, ha  $P(AB) = P(A) * P(B)$  igaz?
- Mekkora a biztos esemény valószínűsége?
- Mekkora a lehetetlen esemény valószínűsége?
- Adj példát biztos és lehetetlen eseményekre.
- Mekkora lehet egy esemény valószínűségének értéke?
- Mit jelent az esély?
- Definiáld a logitot.
- Add meg az esemény logit értékét, ha az esemény valószínűsége 0,12.
- Mekkora az esély értéke, ha a valószínűség 0,4.
- Add meg a valószínűséget, ha az esély 3.
- Add meg a valószínűséget, ha a logit -32.
- Definiáld a populációt és a mintát
- Lehet két esemény együttes bekövetkezésének valószínűsége nagyobb az egyes események bekövetkezésének valószínűségénél?

## Ellenőrző kérdések #2

|                   |               | Tüdődaganat |               | Összesen |
|-------------------|---------------|-------------|---------------|----------|
|                   |               | Van daganat | Nincs daganat |          |
| Dohányzási szokás | Dohányzik     | 79          | 71            | 150      |
|                   | Nem dohányzik | 9           | 18            | 27       |
| Összesen          |               | 88          | 89            | 177      |

Mekkora az esélyhányados és a relatív kockázat?