Biostatisztika és informatika alapjai

3. előadás:

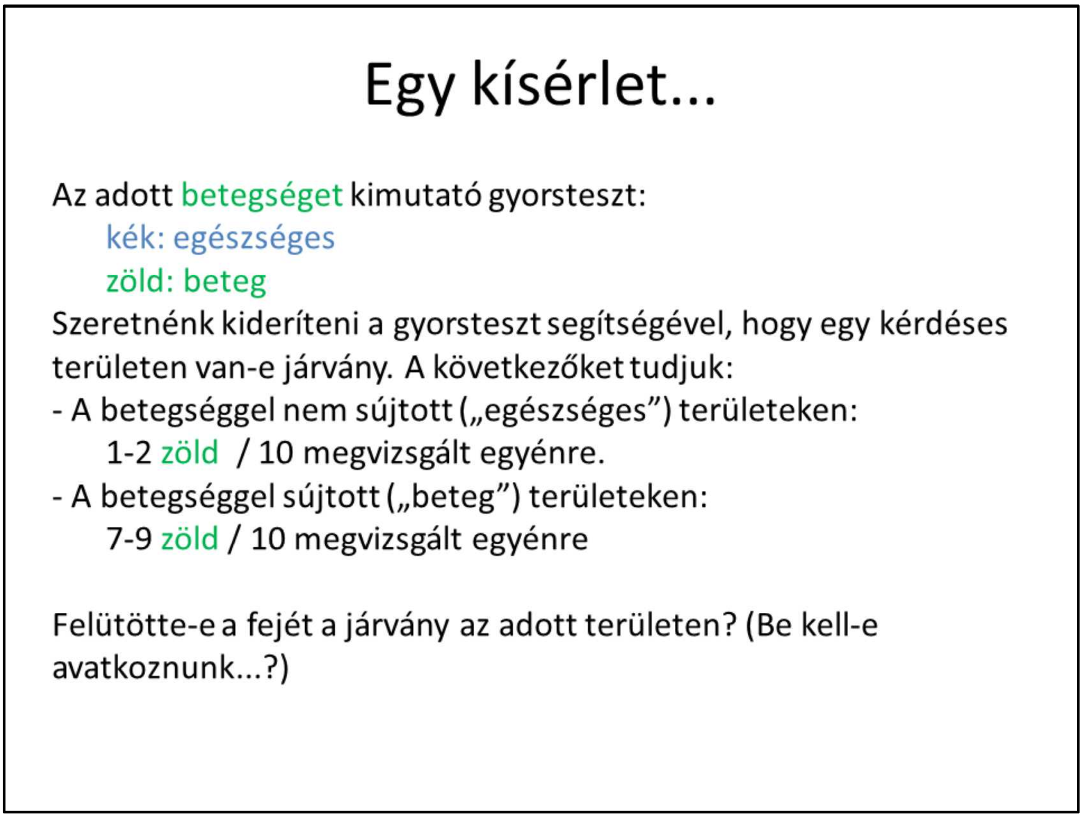
A valószínűségszámítás elemei

2018. szeptember 27.

Veres Dániel

Ebben az előadásban szó lesz valószínűségekről, illetve eloszlásokról.
Az előadás első felében néhány kísérletet mutatok elméletben, illetve gyakorlatban egy kis ízelítőül a valószínűségről, populáció-minta viszonyáról. Ezeken keresztül megpróbálom megmutatni a minta nagyságának szerepét, hogy eljuthassunk a populáció és minta fogalmához és a valószínűség egy definíciójához. Ez a definíció a nagy számok relatív gyakoriságokra vonatkozó törvényén alapszik.
Ezt követően az események valószínűségének alapfogalmaival ismerkedhetünk meg (jelölésrendszer; és/vagy kapcsolat események között; egymást kölcsönösen kizáró események és független események fogalma, használta; Kolmogorov axiómák; feltételes valószínűség és használata).

Az utolsó néhány dián pedig megmutatom, hogy hogyan működik a „mindennapi gondolkodásunk", illetve hogyan kellene működnie, ha a valószínűség-számítás megtanult elemeire is gondolunk...



Ebben a kísérletben egy járványügyi „eljárást” modellezünk.
Az adott betegség kimutatására rendelkezünk egy gyorsteszttel, amelynek az eredménye:

fehér: ha a vizsgált páciens egészséges

zöld: ha az illető beteg

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

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

  1-2 zöld eredményt kapunk 10 megvizsgált egyénre.

- A betegséggel sújtott („beteg”) területeken:

  7-9 zöld eredményt kapunk 10 megvizsgált egyénre

Felütötte-e a fejét a járvány az adott területen?

Sajnálatos módon limitált forrás és idő áll rendelkezésünkre döntésünk meghozatalában, de döntenünk kell.

Egy kísérlet...

Az adott betegséget kimutató gyorsteszt:
   kék: egészséges
   zöld: beteg
Szeretnénk kideríteni a gyorsteszt 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?

Az elvégzett kísérletből azt a tanulságot vonhattuk le, hogy az *elvégzett gyorstesztek (megvizsgált emberek) számának növelésével nő jó döntésünk „biztonsága”*.
Hány mérést kell(ene) elvégeznünk?
Valamekkora *bizonytalansága* mindig lesz a döntésünknek, ha nem mérünk meg *mindenkit*.
Valamekkora *bizonytalansága* mindig lesz a döntésünknek, ha nem mérünk meg *mindenkit*. Mekkora és vajon milyen mennyiség ez a bizonytalanság?

Ahhoz, hogy megválaszolhassuk a fenti kérdéseket először tisztázzuk az *alapsokaság* és a *minta* fogalmát.

Mint említettük, a statisztika végtelen tömegjelenségek leírója. Ez azt jelenti, hogy a jelenségek vizsgálata során *sok*, akár *végtelensok* mérést is végezhetünk. Ezen elméletileg lehetséges összes mérés kimeneteleinek, eredményeinek összefoglaló halmazát nevezzük *alapsokaságnak* (*populációnak*). Elméletben a statisztikai változó teljes megismeréséhez ezt a végtelensok mérést el kellene végeznünk, de erre nyilvánvalóan nincs lehetőségünk.

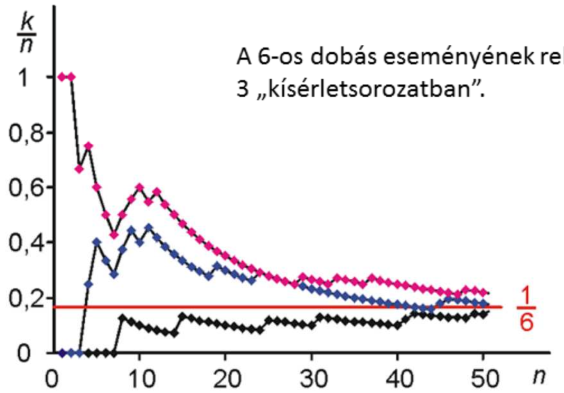
Ezért az alapsokaságnak csak egy részhalmazát vizsgáljuk, amit *mintának* nevezünk. A minta tehát az alapsokaság részhalmaza, amelynek alapja a *véletlen* kiválasztás.

A létrehozott mintán méréseket végzünk, a keletkező mérési eredmények (kimenetelek) halmazát szintén mintának nevezzük. (Magyarán: kevésbé precíz fogalmazással pl. az évfolyam mint alapsokaság egyik csoportjának tagjait tekinthetjük az évfolyamból vett mintának; precízebben fogalmazva az évfolyam hallgatóinak vércsoportadatai jelenthetnek alapsokaságot, míg az egyik csoport tagjainak vércsoportadatai egy lehetséges mintát.)

A mintát jellemezhetjük grafikusan és számszerűen, *ahogyan azt megtanultuk az előző előadáson*. A minta így megállapított tulajdonságait extrapolálhatjuk, azaz kiterjeszthetjük az alapsokaságra. Az előbbi példánál maradva: amilyen arányban a mintában előfordulnak az egyes vércsoportok, kb. olyasmit várunk el az egész alapsokaságtól is. Mivel a minta összeállítása véletlenszerűen történik, nem biztos, hogy tökéletesen reprezentálja az alapsokaságot, a különböző értékek alapsokaságon belüli előfordulási arányát. Így a minta alapján levont következtetésekhez *mindig társul valamekkora bizonytalanság*.

Milyen mennyiség ez a bizonytalanság? Hogyan definiálhatjuk?

Egy másik szemszögből...

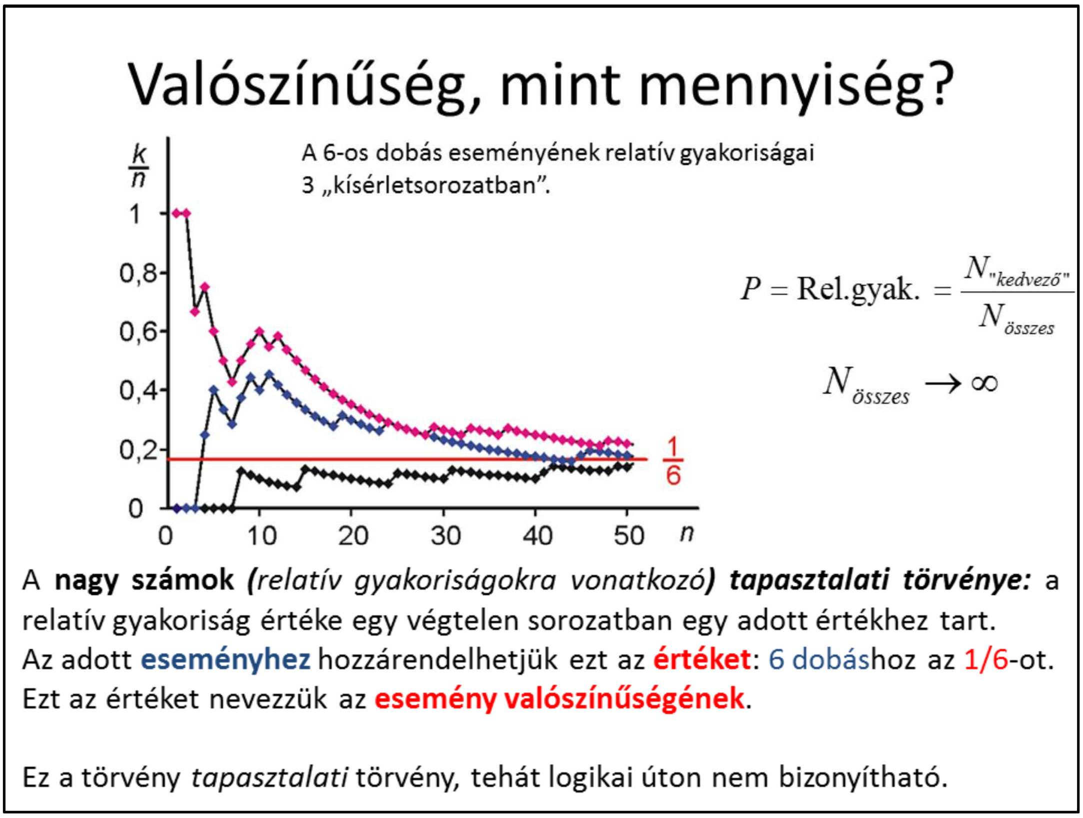


$$\text{Rel.gyak.} = \frac{N_{\text{"kedvező"}}}{N_{\text{összes}}}$$

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.

Ebben a kísérletben egy dobókockát dobunk 50-szer és közben rögzítjük a 6-os dobások relatív gyakoriságát. Ezt a kísérletet 3-szor végezzük el. Ezeknek az eredményei láthatók a dián. A 6-os dobási eredményt (kimenetelt) nevezzek a kedvező kimenetelnek – eseménynek.
Azt tapasztalhatjuk, hogy a *relatív gyakoriság* („kedvező”/összes dobások száma) – bár folyamatosan ingadozva –, de *egy adott értékhez tart*, továbbá ez a stabilizálódó érték az aktuális *„kísérletsorozattól” függetlenül* lényegében ugyanakkora.

Ebben a kísérletben egy dobókockát dobunk 50-szer és közben rögzítjük a 6-os dobások relatív gyakoriságát. Ezt a kísérletet 3-szor végezzük el. Ezeknek az eredményei láthatók a dián.
Azt tapasztalhatjuk, hogy a *relatív gyakoriság* („kedvező”/összes dobások száma) – bár folyamatosan ingadozva –, de *egy adott értékhez tart*, továbbá ez a stabilizálódó érték az aktuális *„kísérletsorozattól” függetlenül* lényegében ugyanakkora.



Ezt a megfigyelést nevezzük a *nagy számok (relatív gyakoriságokra vonatkozó) tapasztalati törvényének: :* a relatív gyakoriság értéke egy végtelen sorozatban egy adott értékhez tart.
Rendeljük hozzá a „kedvező” eseményünkhöz a „stabilizálódó” értéket, mint mennyiséget: jelen esetben a 6-os dobáshoz az 1/6-ot.
Ezt az értéket nevezzük az *esemény valószínűségének*. (Egy adott helyzetben mért változó kimenetelének valószínűsége.)
Tehát azt is mondhatjuk, hogy egy esemény *relatív gyakorisága megegyezik az esemény valószínűségével, ha az ismétlések száma* (kísérletsorozat hossza) *végtelen*.
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)

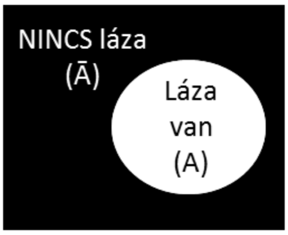
Ellentett esemény: **Ā**

(a páciensnek NINCS láza)

Az A esemény be NEM következésének valószínűsége: **P(Ā) vagy P(nemA)**

(annak a valószínűsége, hogy a páciensnek NINCS láza)

Venn-diagram



A következőkben az események valószínűségeinek alapfogalmait tárgyaljuk.
Először ismerkedjünk meg a jelölésrendszerrel. (*a következőkben zárójelben, dőlt betűkkel találjátok a példákat*)
Az egyes eseményeket nagy betűvel jelöljük – például *A = a betegnek láza van*.
Az *A* esemény bekövetkezésének valószínűségét pedig *P(A)*-val jelöljük. Például *P(A) = annak a valószínűsége, hogy a páciensnek láza van*.

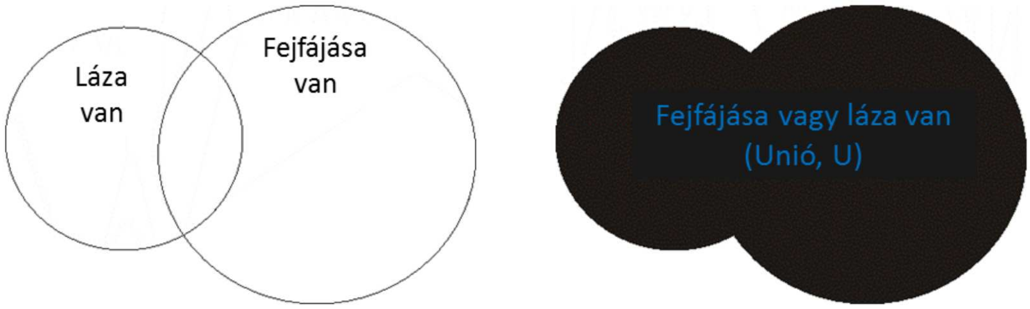
Ellentett eseménynek hívjuk az adott esemény tagadását: az esemény nem következik be. Jelölése: Ā (pl. a páciensnek NINCS láza) Az A esemény be NEM következésének valószínűségének jelölése: P(Ā) vagy P(nemA) (annak a valószínűsége, hogy a páciensnek NINCS láza). Venn-diagramon a dián látható módon ábrázolható.

Események valószínűségei II.

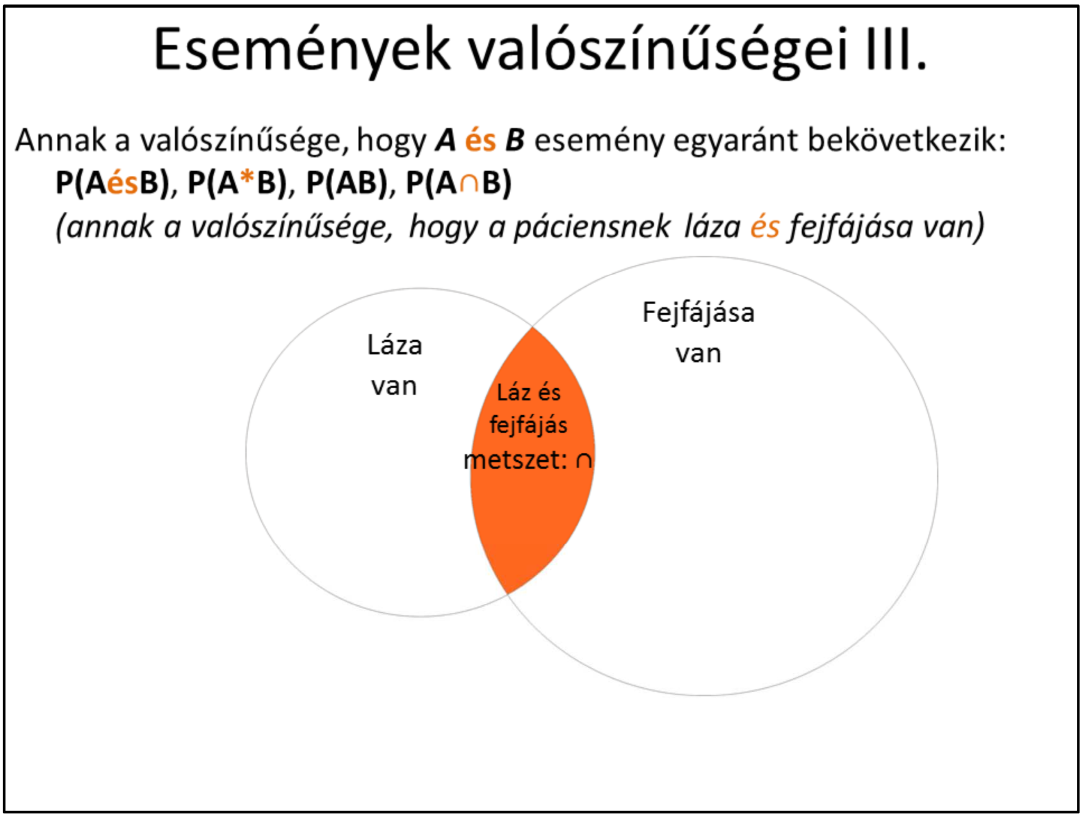
Annak a valószínűsége, hogy *A* **vagy** *B* esemény bekövetkezik:

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

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



Annak a valószínűségét, hogy az *A vagy* a *B* esemény bekövetkezik (*a páciensnek láza vagy fejfájása van*) több féle módon jelölhetjük: *P(AvagyB), P(A+B)*, *P(AUB)*.
Ez utóbbi jelölés a halmazoknál szokásos jelölésnek felel meg: az *U a halmazok unióját* jelenti, azaz legalább az egyik halmazhoz való tartozást. Ennek a Venn-diagramon való ábrázolása látható az ábrán.

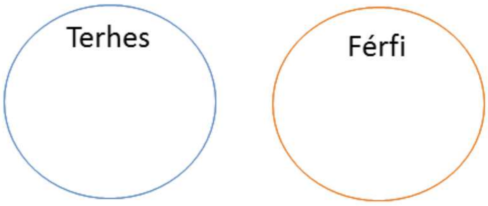


Annak a valószínűségét, hogy *A és B* esemény egyaránt bekövetkezik a következő módokon jelölhetjük: P(AésB), P(A*B), P(AB), P(A∩B). (*Annak a valószínűsége, hogy a páciensnek láza és fejfájása is van.)*
A ∩ jel a halmazok metszetét jelenti. A Venn-diagramon ez a két halmaz közös része.

Események valószínűségei IV.

Egymást (kölcsönösen) kizáró események: *A és B* események együttesen nem következhetnek be.

(a páciens terhes és férfi) *(A∩B)=0*



Független események: *A* esemény bekövetkeztének nincs hatása *B* bekövetkezésére.

(az első páciensünk férfi, a második nő)

Ezen a dián kétféle esemény-kapcsolatra hívnám fel a figyelmet.
Az *egymást kölcsönösen kizáró események* azok, amelyek együttesen nem következhetnek be *(a páciens férfi és terhes)*. Halmazokkal ábrázolva őket látszik, hogy ez azt jelenti, hogy a két halmaz nem metszi egymást (metszetük üres halmaz, azaz (A∩B)=0).

Független eseményekre igaz, hogy az egyik esemény bekövetkezésének nincs hatása a másik esemény bekövetkeztére *(az első páciens férfi, a második pedig nő)*.

Események valószínűségei V.

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)

Mielőtt tovább haladnánk, definiáljuk a *feltételes valószínűséget*.

A „feltételesség” arra vonatkozik, hogy csak egy alcsoporton belül – azaz ha már egy esemény bekövetkezett – számítjuk a valószínűséget. Másként megfogalmazva *A* eseménynek *B*-re vonatkoztatott feltételes valószínűsége az *A* esemény valószínűsége, ha *B* bekövetkezett. Jelölése: P(A|B). Például *annak a valószínűsége, hogy a páciensünknek influenza fertőzése van tudva azt, hogy vírusos eredetű a fertőzése*.

Események valószínűségei VI.

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

1. **0 ≤ P(A) ≤ 1**

2. **P(*biztos*) = 1** *(a páciens előbb vagy utóbb meghal)*
P(*lehetetlen*) = 0 *(a páciens teljesen egészséges*)*

*3. Egymást kölcsönösen kizáró eseményekre: (*P(AésB)=0)
P(AvagyB)=P(A)+P(B)
*(annak a valószínűsége, hogy páciensünk terhes **vagy** férfi)*

Ezekből levezethető:
+4. *Független eseményekre*: **P(AésB)=P(A)*P(B)**
*(annak a valószínűsége, hogy az első páciensünk férfi **és** a második nő)*

Az események valószínűségeinek leírását a Kolmogorov axiómák segítségével tehetjük meg. Az alábbiakban ezeket az axiómákat írom le (kissé egyszerűsített formában).
1. Egy esemény valószínűsége 0 és 1 közötti érték.
2. a) A biztos esemény valószínűsége 1 (*a páciens előbb-utóbb meghal* – mint tudjuk az élet egy szexuális úton fertőző halálos betegség ☺).
2. b) A lehetetlen esemény valószínűsége 0 (*a páciens teljesen egészséges* – mint tudjuk nincs egészséges ember, csak rosszul kivizsgált beteg ☺).
3. Annak a valószínűsége, hogy *A vagy B* esemény bekövetkezik, ha *A* és *B* események *egymást kölcsönösen kizáróak* egyenlő az egyik esemény valószínűségének és a másik esemény valószínűségének az összegével *(annak a valószínűsége, hogy következő páciensünk egy férfi lesz vagy egy terhes = annak a valószínűsége, hogy férfi lesz a következő páciensünk + annak a valószínűsége, hogy egy terhes lesz a következő páciensünk).*

Ezekből az axiómákból számos egyéb összefüggés levezethető, amelyek közül most egyet emelnénk ki.
+4. Annak a valószínűsége, hogy *A és B* esemény is bekövetkezik, ha *A* és *B* események egymástól független események egyenlő az *A* esemény valószínűségének és a *B* esemény valószínűségének szorzatával *(annak a valószínűsége, hogy az első páciensünk férfi és a rákövetkező nő lesz = annak a valószínűsége, hogy férfi lesz a következő páciensünk*annak a valószínűsége, hogy a rákövetkező páciensünk nő lesz)*.

Megjegyzendő, hogy ezek a megállapítások (2-4) „fordítva” is igazak. Például ha P(AésB)=P(A)*P(B), akkor *A* és *B* események egymástól függetlenek.

Események valószínűségei VII.

Feltételes valószínűség számítása

általános forma 2 eseményre: **P(A|B)=P(AésB)/P(B)**

A következőkben a feltételes valószínűséggel kapcsolatos számításokat mutatok be. A fent látható összefüggés az úgynevezett Bayes tétel egyszerűsített, de általános formája 2 esemény valószínűségére vonatkoztatva. Ezt szoktuk a feltételes valószínűségek „szorzási szabályának" is hívni. Először 2 speciális esetben vizsgáljuk meg ezt a 2 összefüggést.

Események valószínűségei VIIa.

Feltételes valószínűség számítása
általános forma 2 eseményre: **P(A|B)=P(AésB)/P(B)**

Különleges esetek:
I. Független eseményekre:
annak a valószínűsége, hogy a második páciensünk férfi,
***HA** az első nő volt*

P(A|B)=P(AésB)/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*

Egyik speciális esete a feltételes valószínűségek szorzási szabályának, ha független eseményekre vonatkoztatjuk, például a kérdésünk az, hogy mekkora a valószínűsége annak, hogy a rendelőnkben a második páciens férfi, ha az első nő volt.
Mint az látjuk ebben az esetben a feltétel valószínűségének nincs hatása az eseményünk valószínűségére.
Ugyanez a helyzet érme feldobásnál, kockadobásnál, stb. – az előző eseményeknek nincs hatása az utána jövő esemény valószínűségére (független tőle)

Események valószínűségei VIIb.

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(AésB)/P(B)

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

Számolási példa:

Annak a valószínsége, 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%.

Másik speciális eset, ha az *A* esemény részhalmaza *B* események.
Példa:
Számolási példa:
Annak a valószínsége, 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%.
Jól látható, hogy a vírusfertőzés megállapításával erősen nőtt az influenzafertőzés valószínűsége.

Kockázat (Rizikó)

		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:

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

A betegség kockázata (valószínűsége), ha NINCS rizikófaktor:

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

Az orvosi gyakorlatban a feltételes valószínségek esetében gyakran használjuk a *kockázat* (rizikó, risk) kifejezést. Ezzel leggyakrabban ún. *rizikófaktor*ok (kockázati tényezők, veszélyeztetettségi csoportok) kutatásánál találkozunk, ahol két nominális változó viszonyát vizsgáljuk: az egyik változó a rizikófaktor (pl. *dohányzás*), a másik a betegség (pl. *tüdőrák*)

Először számítsuk ki a betegség kockázatát (rizikóját, feltételes valószínűségét), ha a kockázati tényező jelen van.
Ez egy feltételes valószínűség, tehát használhatjuk az előzőekben megismert összefüggést. (Itt az és helyett a metszet szimbólumot használtam)
Láthatjuk, hogy ez a valószínűség nem más, mint a betegség relatív gyakorisága a rizikócsoportban.

Hasonló módon számítsuk ki a betegség kockázatát a kockázati tényezővel nem rendelkező csoportban.

Kockázati hányados (Rizikó hányados)

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

Kockázati hányados, rizikóhányados (RR, Risk Ratio, Relative Risk): azt az arányt mutatja, hogy hányszor gyakoribb az esemény kockázata, ha jelen van a rizikófaktor, mint a rizikófaktor hiányában

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

Ezekben a tanulmányokban legtöbbször a kockázatok arányára vagyunk kíváncsiak: hányszor gyakoribb az esemény kockázata, ha jelen van a rizikófaktor, mint a rizikófaktor hiányában. Például: hányszor nagyobb a tüdőrák kockázata a dohányzók körében, mint a nem dohányzóknál.

Ezt az arányt nevezzük *kockázati hányadosnak*, *rizikó hányadosnak,* illetve relatív kockázatnak. Rövidítése: RR (a risk ratio, illetve relative risk szavakból). [Megjegyzés: az amerikai nómenklatúrában a risk ratio a helyes kifejezés, a relative risk alatt néha más hányadosokat is érthetnek.]

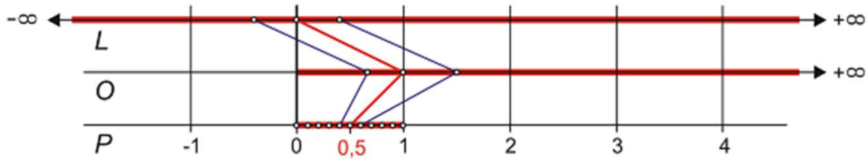
Esély

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

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

Logit (L): esély természetes alapú logaritmusa

Logit – Esély – Valószínűség



Egy további „valószínűség szerű” fogalomra hívnám fel a figyelmet, ez az *esély* (angol nevén odds). Ez az esemény bekövetkezésének valószínűségének és nem bekövetkezésének valószínűségének hányadosa, amely megmutatja, hogy hányszor akkora a valószínűsége annak, hogy az esemény bekövetkezik, mint annak, hogy nem következik be.

A *logit* nem más, mint az esély természetes alapú logaritmusa.
[Megjegyzés: egyes, az orvosi gyakorlatban is gyakran használt haladóbb statisztikai elemzéseknél a logit gyakran előkerül.] A logit-esély-valószínűség értékeinek kapcsolatát mutatja az ábra. Látható, hogy a valószínűség értéke 0 és 1 között, az esélyé 0 és végtelen között, amíg a logit értékei mínusz végtelen és plusz végtelen között vannak. Leolvashatjuk, hogy ha a valószínűség kisebb, mint 0,5, akkor az esély 1-nél kisebb, a logit pedig negatív. Természetesen további összefüggések is láthatóak...
Mellékelt excel fájl.

Esélyhányados

		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 esélye, ha van rizikófaktor:

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

A betegség esélye, ha NINCS rizikófaktor

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

A korábban említett kockázati tényező vizsgálatokban az relatív kockázat mellett az *esélyhányadost* is gyakran számítjuk.
Ehhez számítsuk ki a betegség esélyét, ha van, illetve ha nincs rizikótényező.

Esélyhányados

		Betegség		
		Igen	Nem	Összesen
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

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

Esélyhányados (esélyarány), rövidítése OR (Odds Ratio-ból): hányszor nagyobb a betegség kialakulásának esélye a rizikófaktor meglétes estében, mint annak hiányakor. Például: hányszor nagyobb a tüdőrák esélye a dohányzók körében, mint a nem dohányzóknál.

Rizikóhányados és Esélyhányados

		Betegség		
		Igen	Nem	Összesen
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

$$a \ll b \qquad c \ll d \qquad OR \Rightarrow RR$$

Hasonlítsuk össze az esélyhányadost és a rizikóhányadost. Mint látjuk, OR nem egyenlő RR-rel, de vegyük észre, ha a<<b and c<<d – tehát a betegség ritka, akkor az egyszerűbben számítható esélyhányados értéke a rizikóhányados értékéhez tart. (Egyéb érdekességek a könyvben olvashatók.)

Rizikóhányados és Esélyhányados

		Tüdődaganat		
		Van daganat	Nincs daganat	Összesen
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 faktorral

R<1: kisebb kockázat/esély a faktorral

Lehet, de NEM BIZTOS

Mintavételi bizonytalanság!

Csináljunk egy számítást a példa kedvéért. Az eredmény látható az ábrán. Mit jelent ez a szám? Általában: R=1 – nincs hatása a kockázati tényezőnek, R>1 – a kockázat, esély nagyobb a rizikócsoportban, R<1 – a kockázat kisebb a rizikócsoportban.
De vajon jelen esetben ez tényleg azt jelenti, hogy 2,2-szeres az esély a tüdődaganatra dohányzás esetében? Ez lehet, vagy MÉGSEM - mintavételi bizonytalanság (lásd később a khi-négyzet teszteknél)

Valószínűségszámítás……

Permutációk
Variációk
Kombinációk

A statisztika alapját a kombinatorika, azaz a permutációk, variációk és kombinációk adják.
A tantárgy során azonban nem mélyedünk el ezekben jobban.

Na mire is lehet jó....

Influenzaszezont megelőzően a rendelőnkben 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?

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

Mire is lehetnek ezek jók nekünk? A következőkben olyan példát mutatok, ahol az ismétlés nélküli kombináció könnyen tetten érhető.
Influenzaszezont megelőzően a rendelőnkben 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?

A kérdés megválaszolásához a Bernoulli eloszlást (lásd később) használjuk fel (vegyük észre, hogy a feladat során a 25 emberből választunk ki 4-et jelenti az ismétlés nélküli kombinációt).
Számos ehhez hasonló kérdést vetettem fel a dián. A kérdés az, hogy ezek megválaszolásához mit tegyünk, milyen „egyenletet” válasszunk, mikor melyiket?

Na mire is lehet jó….

Mekkora a valószínűsége annak, hogy páciensünk 3.45 mmol/l-es (normál tartományon kívüli) K^+ 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?

Számos, az előzőhöz hasonló hasonló kérdést vetettem fel a dián. A kérdés az, hogy ezek megválaszolásához mit tegyünk, milyen „egyenletet”, táblázatot, excel függvényt... válasszunk, mikor melyiket?

Az emberi gondolkodás...

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

a) Tomi könyvtáros
b) Tomi kétkezi munkás

A mindennapi gondolkodásunk sok esetben eltér attól, ahogyan annak a valószínűségszámítás ismeretében működnie kellene.
A dián látható két állítás közül sokan közületek az elsőt (Tomi könyvtáros) tartotta valószínűbbnek. Pedig belegondolva abba, hogy mennyire gyakori a könyvtárosi állás, illetve a kétkezi munkás, egyértelmű, hogy annak nagyobb a valószínűsége, hogy 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.

A következő példában két állításra hívnám fel a figyelmet: Linda bankpénztáros, illetve Linda bankpénztáros és feminista. Az előadás során többetek előrébb sorolta azt, hogy Linda bankpénztáros és feminista, mint azt, hogy Linda bankpénztáros. Pedig ismerve a 11. dián is leírt független események valószínűségének együttes előfordulási valószínűségét, beláthatjuk, hogy egy esemény valószínűsége mindig legalább ugyanakkora, mint ugyanezen esemény és egy másik esemény együttes előfordulásának valószínűsége... (De úgy is mondhatjuk, hogy a bankpénztáros és feminista halmaz részhalmaza a bankpénztáros halmaznak)

Valószínűség másképp...

Javaslom mindenkinek, érdekességként olvasson utána a Buffon-féle tűkísérletnek.

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?
- Definiált 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 valószínűségre.
- 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(AvagyB)=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?

A kérdések megválaszolhatók az előadáson elhangzottak, a gyakorlatvezetővel folytatott konzultációk, illetve saját utánaolvasás segítségével. Az ellenőrző kérdések egyben példák arra, hogy milyen tesztkérdések (feleletválasztós formában) fordulhatnak elő.

Ellenőrző kérdések #2

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

Mekkora az esélyhányados és a rizikóhányados?