

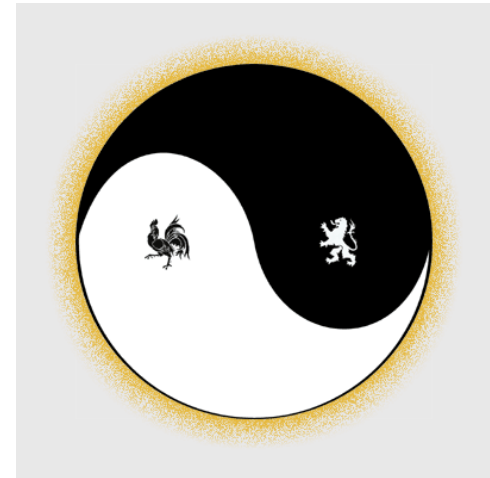
Kapcsolat vizsgálat II: kontingencia táblák jelentősége és használata az epidemiológiában, diagnosztikában: RR, OR. ROC analízis.



Dr. Prohászka Zoltán
Az MTA doktora
Semmelweis Egyetem
III. Sz. Belgyógyászati Klinika
2008-04-17

prohoz@kut.sote.hu

www.kutlab.hu



Tipikus diszkrét eloszlású változók

Sorszám	Életkor (év)	Nem	Diabétesz típusa	Retinopátia	Korábbi miokardiális infarktus	Inzulinkezelés
1	56	Nő	2	Nincs	Nem	Nem
2	59	Nő	2	Nincs	Nem	Nem
3	65	Férfi	2	Egyoldali	Igen	Igen
4	80	Nő	2	Egyoldali	Nem	Nem
5	66	Férfi	2	Egyoldali	Nem	Igen
6	31	Férfi	2	Kétoldali	Nem	Igen
7	25	Férfi	1	Nincs	Nem	Igen
8	55	Férfi	1	Egyoldali	Nem	Igen
9	76	Nő	3	Egyoldali	Nem	Igen
10	67	Nő	2	Nincs	Nem	Igen
11	75	Nő	2	Kétoldali	Nem	Igen
12	48	Férfi	2	Nincs	Nem	Nem
13	81	Férfi	2	Egyoldali	Nem	Igen
14	54	Nő	2	Nincs	Nem	Nem
15	53	Nő	2	Nincs	Nem	Nem

Hogyan összegezzük diszkrét változóinkat?

- Az előfordulási *arány* (proportion) az adott értékű (kódú) megfigyelések száma osztva az összes megfigyelés számával.
 - Az arány *százalékos* értékben is kifejezhető.
- Ezzel szemben a *viszonyszámot* (*ratio*, hányados) is számíthatunk. Esetünkben a férfiak nők hányadosa (viszonyszáma) $7/8=0,875$.

Nem	Betegek száma	Arány az összes beteghez képest
Férfi	7	$7/15=0,46$
Nő	8	$8/15=0,54$
Összesen	15	
Férfi/nő arány	$7/8=0,875$	

- Az *arányok* és *viszonzszámok* igazán akkor válnak értékkessé, ha a nyers számokat több csoportban (dimenzió mentén) tudjuk egymással összehasonlítani
- Ezzel el is jutunk a diszkrét eloszlású változók értékelésének alapvető formájához, a **kontingencia táblázathoz**

Nem	Inzulinkezelés nem	Inzulinkezelés igen	Összesen
Férfi	1	6	7
Nő	5	3	8
Összesen	6	9	15

- A kontingencia táblázat sarokszáma (a jobb alsó cella) megegyezik a minta elemeinek számával (15),
- Egy elem csak egy cellában szerepelhet (ez a diszkrét változók azon jellegzetességéből következik, hogy a változók értékei, kategóriái egymást kizáró természetűek, az inzulinkezelés pl. nem lehet igen is meg nem is).
- A kontingencia táblázatban bemutatott eredmények alapján a populációra kívánunk következtetni, így egy táblázatban **a teljes mintánk** minden eredményének szerepelnie kell, és mindenki csak egy cellában szerepelhet.
- Az oszlopok és sorok sorrendje felcserélhető.

A chi-négyzet próba

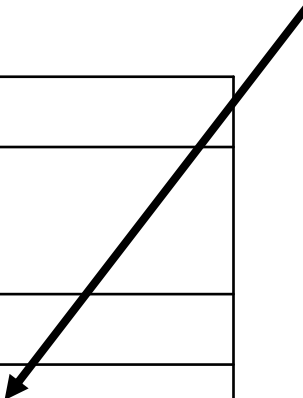
- Tegyük fel, hogy 200 személyt vontunk be vizsgáltunkba, a csoport fele kapott kezelést (pl. vérnyomáscsökkentő gyógyszer), míg másik fele képezte a kontrollt (placebo).
- A 200 személy közül 50 esetében javulást észleltünk (pozitív válasz), míg 150 személynél nem változott a vérnyomás (negatív válasz).

Csoport	Kezelt	Kontroll	
Pozitív válasz			50
Negatív válasz			150
	100	100	200

A chi-négyzet próba

- Tegyük fel, hogy 200 személyt vontunk be vizsgáltunkba, a csoport fele kapott kezelést (pl. vérnyomáscsökkentő gyógyszer), míg másik fele képezte a kontrollt (placebo).
- A 200 személy közül 50 esetében javulást észleltünk (pozitív válasz), míg 150 személynél nem változott a vérnyomás (negatív válasz).
- Ha a kezelés és a kezelésre adott válasz között nem állna fenn kapcsolat (azok *függetlenek* lennének), akkor a kezelték és a kontrollok csoportjában az emberek 25–25%-a esne a pozitív válasz kategóriába (null-hipotézis).
- A várható előfordulási gyakoriságokat a táblázat széli gyakoriságaival (*marginal frequency*, a sorok és oszlopok összesített értékei) számíthatjuk ki, feltételezve, hogy ezek a számok adottak egy vizsgálatban.

<i>Várható (expected) előfordulási gyakoriság</i>			
Csoport	Kezelt	Kontroll	
Pozitív válasz	25	25	50
Negatív válasz	75	75	150
	100	100	200



A megfigyelt gyakoriságok

- A chi-négyzet teszt alapfeltevése az, hogy a várt és a megfigyelt eloszlási gyakoriságok között nincs különbség
- Ha a táblázat oszlopait és sorait meghatározó jellegzetességek egymással kapcsolatot mutatnak, akkor a várt és a megfigyelt eloszlások egymástól lényegesen eltérnek
- Az eltérés *mértékét* fejezi ki a chi-négyzet értéke. Esetünkben a kezelt csoportban jóval nagyobb a pozitív válasz arány, mint a kontroll csoportban, vagyis levonhatjuk azt a következtetést, hogy a kezelés kapcsolatban van a pozitív kimenetellel, a chi-négyzet értéke utal a kapcsolat erősségére.

<i>Megfigyelt (observed) előfordulási gyakoriság</i>			
Csoport	Kezelt	Kontroll	
Pozitív válasz	40	10	50
Negatív válasz	60	90	150
	100	100	200

A példa számaival végzett chi-négyzet próba eredménye

Chi-négyzet:	24,00	
Szabadsági fok:	1	
p-értéke:	<0,0001	
Egy- vagy kétoldalas a p?	Két oldalas	
Relatív kockázat (RR) (95% CI)	2,0 (1,57–2,54)	→ Prospektív kohorsz
Esélyhányados (OR) (95% CI)	6,0 (2,79–12,91)	→ Eset-kontroll

A Physician's Health Study egyes eredményeinek összegzése, követési idő átlagosan 5 év.

- A Physician's Health Study során egészséges amerikai orvosok véletlenszerűen alacsony dózisú aszpirint vagy placebót kaptak kezelésül, majd követték őket átlagosan 5 évig és regisztrálták az új kardiovaszkuláris események számát.
- A megválaszolni kívánt kérdés így hangzik: van-e kapcsolat a gyógyszeresedés és a bekövetkező infarktusok száma között, vagyis csökkenti-e az aszpirin az infarktus-rizikót?

	Infarktus nem alakult ki (n)	Infarktus kialakult (n)	Összesen (n)
Aszpirin	10898	139	11037
Placebo	10795	239	11034
Összesen	21693	378	22071

Az esemény ráta

- Az esemény ráta fogalma: az eseményt elszenvedett személyek száma osztva az összes, rizikónak kitett személy számával, osztva az alappal (5 év) egy adott időszakra vetítve.
- Példánkban az experimental event rate (a gyógyszerrel kezelt csoport esemény rátája): $EER=139/11037=0,0126/5$ év, míg a control event rate (a placebóval kezelt kontroll csoport esemény rátája): $CER=239/11034=0,0217/5$ év.
- A két ráta összehasonlításával nyerhetünk információt a kezelés hatásosságáról.

A relatív kockázat

- A *relatív kockázat*, melynek jelölése általánosan *RR (relative risk)*, mindig egy adott beteg csoport és egy adott kontroll csoport vonatkozásában fejezhető ki, és az adott betegségre vonatkozik.
- Kiszámításakor a rizikóhatásnak kitett (exposed) és a ki nem tett (non-exposed) csoportokban megfigyelt betegség incidenciákat osztjuk egymással.
- A relatív kockázat a két csoport esemény rátáinak hányadosa.
- Példánknál maradva, az infarktus relatív kockázata a vizsgálatban $RR=0,0126/0,0217=0,58$.
- A szám értelmezéséhez gondoljuk végig, hogy a „rizikófaktor” esetünkben valójában egy védőfaktor (hatásosnak gondolt gyógyszer), vagyis a relatív kockázat esetünkben csökkent.

A relatív rizikó értelmezése

- Az aszpirinnal kezelt csoportban kisebb az infarktus kialakulásának kockázata, mint a placebóval kezelt kontroll csoportban. Ha fordítva gondolkodunk, és azt szeretnénk kifejezni, hogy aszpirint nem szedő személyeknek mennyivel nagyobb a kockázata, mint az aszpirint szedőké, vegyük a relatív kockázat reciprokát: $1/0,58=1,72$. Más szavakkal, a placebo csoport tagjainak 1,72-szeres az infarktus kockázata az aszpirinnal kezelt csoport tagjaihoz képest.
- Nagyon fontos szem előtt tartani, hogy relatív kockázatot csak kohorsz vizsgálat vagy klinikai kísérlet eredményeire építve lehet számítani. Először azonosítani kell a rizikónak kitett és ki nem tett csoportokat, majd azokat követve, és az előre definiált eseményeket rögzítve lehet a relatív kockázatot kiszámolni.

Az esélyhányados (odds ratio)

- A esélyhányados a kockázat becslésére szolgál, és leggyakrabban *eset-kontroll* tanulmányok során találkozunk alkalmazásával.
- Az OR annak az esélye, hogy egy eseményt elszenvedett személy a rizikófaktornak *ki volt téve*, osztva annak az esélyével, hogy egy eseményt el nem szenvedett személy nem volt kitéve a rizikófaktornak
- Példánk kohorsz vizsgálat volt, így az esélyhányados értelmezésének nem sok haszna volna...

Egy hipotetikus eset-kontrollos vizsgálat

- Egy eset-kontroll vizsgálatba 900 infarktust elszenvedett, és 1000 kontroll személyt válogatunk be. Felmérjük dohányzási szokásaikat, melyeket össze kívánunk vetni az infarktus kialakulás esélyével, kockázatával.

Csoport	Infarktus igen	Infarktus nem	
Dohányzik/dohányzott	600	220	820
Soha nem dohányzott	300	780	1080
	900	1000	1900

- $(600/900)/(300/900)=2$ ill. $(220/1000)/(780/1000)=0,28$, vagyis $OR=2/0,28=7,14$

A képletek összefoglalása

	Betegek száma	Egészségesek száma	Összesen:
Rizikófaktor igen	A	B	A+B
Rizikófaktor nem	C	D	C+D
Összesen:	A+C	B+D	A+B+C+D

$$\rightarrow EER = A / (A+B)$$

$$\rightarrow CER = C / (C+D)$$

$$\frac{EER}{CER} = RR$$

$$OR = \frac{[A/(A+C)] / [C/(A+C)]}{[B/(B+D)] / [D/(B+D)]} = \frac{A/C}{B/D} = \frac{AD}{BC}$$

$$ARR = | EER - CER | \rightarrow 1/ARR = NNT$$

$$RRR = | EER - CER | / CER = ARR / CER$$

EER: Experimental event rate

CER: Control event rate

ARR: Absolute risk reduction

RRR: Relative risk reduction

NNT: Number needed to treat

RR: Relative risk

OR: Odds ratio

A Fisher's exact teszt

- A chi-négyzet próba a megfigyelt gyakoriságok és a várható gyakoriságok közötti különbségek mértékét ítéli meg.
- A teszt mögött álló képlet ismertetése nélkül jegyezzünk meg annyit, hogy a *várható gyakoriság* szerepel a nevezőben, így ha ennek értéke *túl* kicsi, akkor a chi-négyzet értéke túl nagy lesz, ami hamis következtetések levonásához vezetne.
- De mi az a *túl* kicsi? Erre nézve nincs arany szabály. A gyakorlatban elterjedt szabály, miszerint „ha a kontingencia tábla valamely cellájában 5 ember vagy annál kevesebb szerepel, akkor Fisher's exact tesztet kell végezni” nem teljesen helyes, mert a *megfigyelt* értékekre ad útmutatást, míg a nevezőben a *várható* érték szerepel.
- A gyakorlati szabályt tehát úgy kell megfogalmazni, hogy alternatív tesztet (Fisher's exact tesztet) tanácsos alkalmazni, ha a *várható* előfordulás < 5 a 2×2 kontingencia tábla bármely cellájában. A várható előfordulás könnyen kiszámítható a marginális frekvenciák ismeretében.
- A Fisher's exact teszt nagyon hasonlít a chi-négyzet próbához, a különbség annyi, hogy kevésbé érzékeny extrém eloszlásokra.

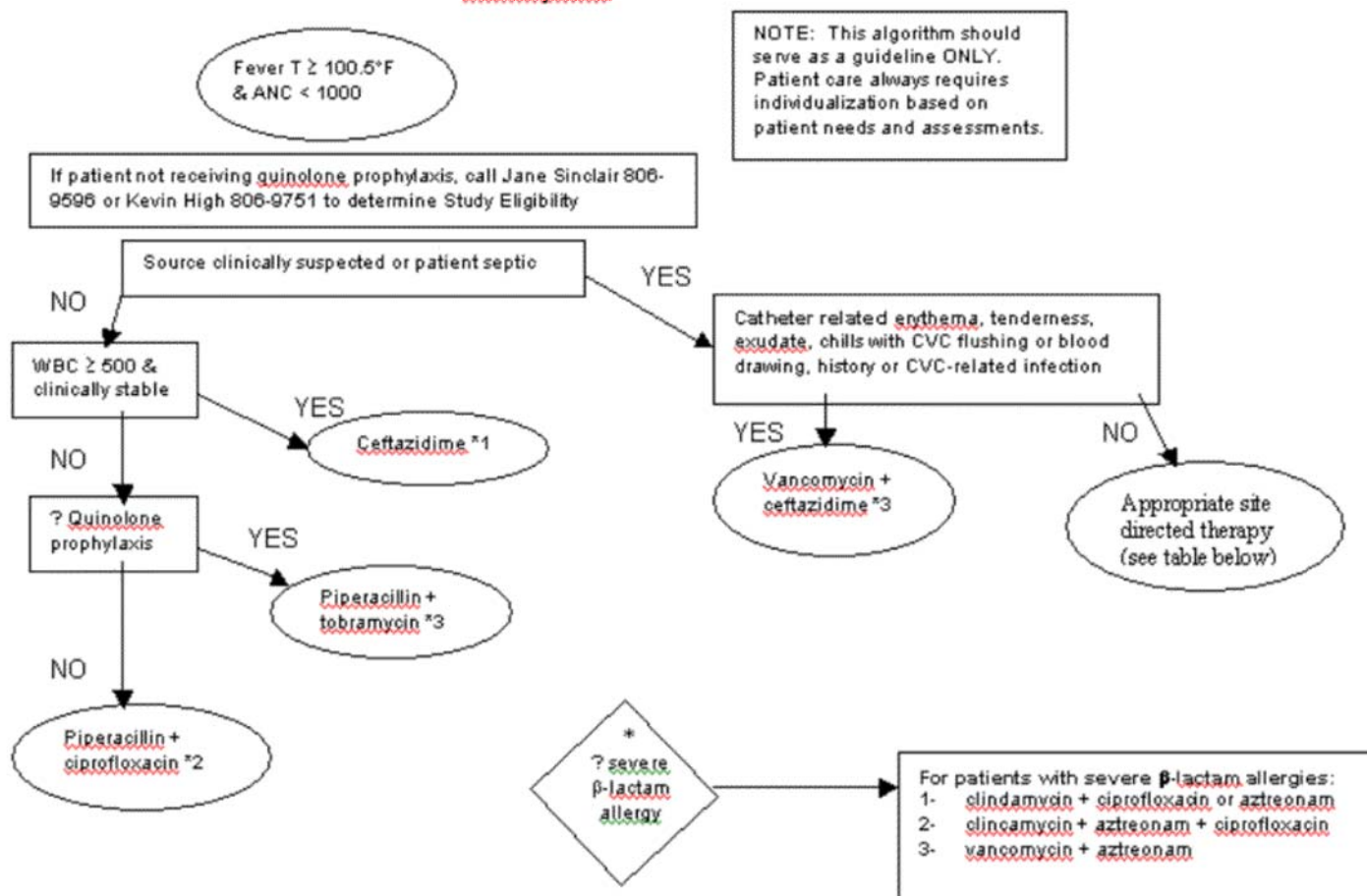
Chi-négyzet for trend teszt

- Az ordinális skálán felvett diszkrét eloszlású változók bemutatásának és kiértékelésének módja.
- A táblázat sorainak sorrendje ebben az esetben lényeges, tükröznie kell a biológiai sorrendet.

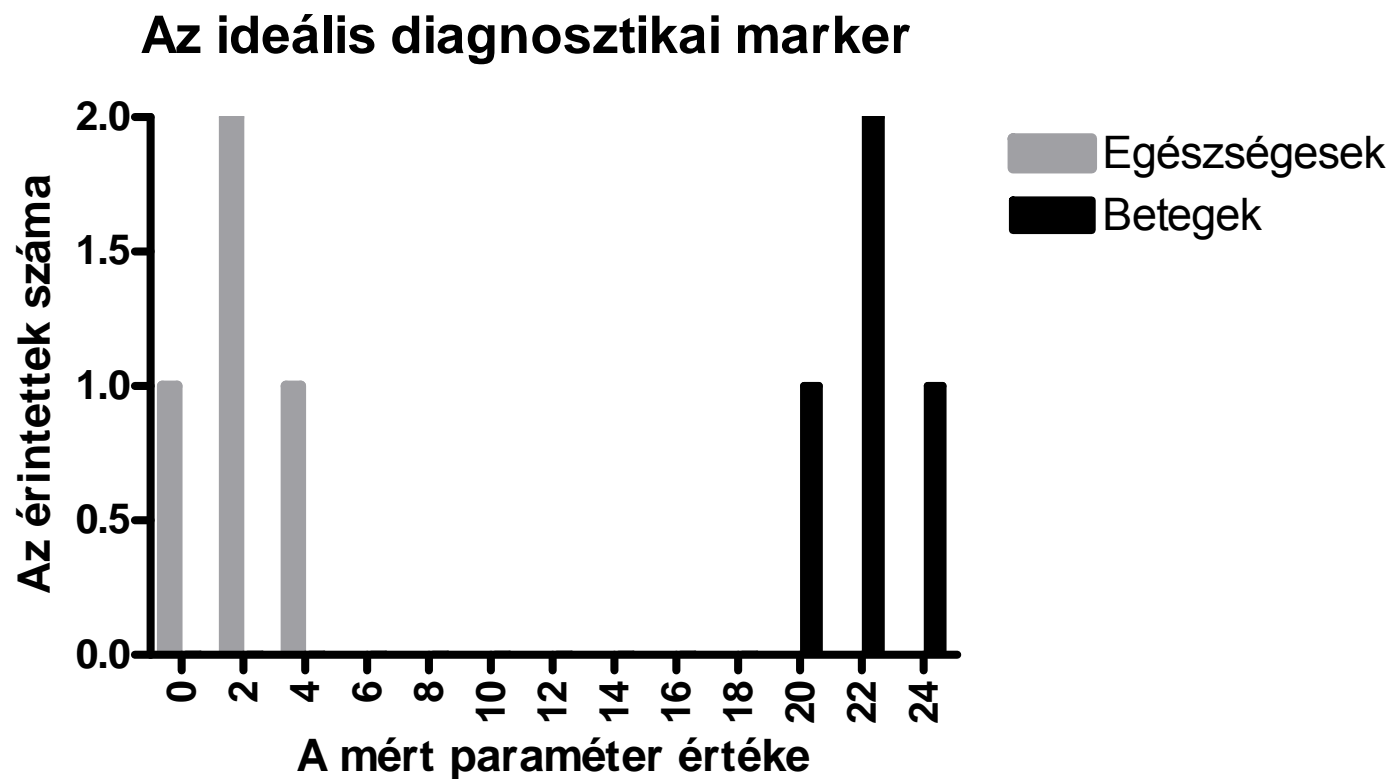
	I-es típus	II-es típus	Összesen:
Retinopátia nincs	16	115	131
Egyoldali	17	29	46
Kétoldali	5	4	9
Összesen:	38	148	186

A nem diszkrét eloszlású változókból alakított csoportok: a dichotomizálás

- Az orvosi döntéshozatalban és a kutatás során szükség lehet a folyamatos változók mentén kialakított kategóriák definiálására
 - Magas/alacsony, fiziológiás/kóros, stb...



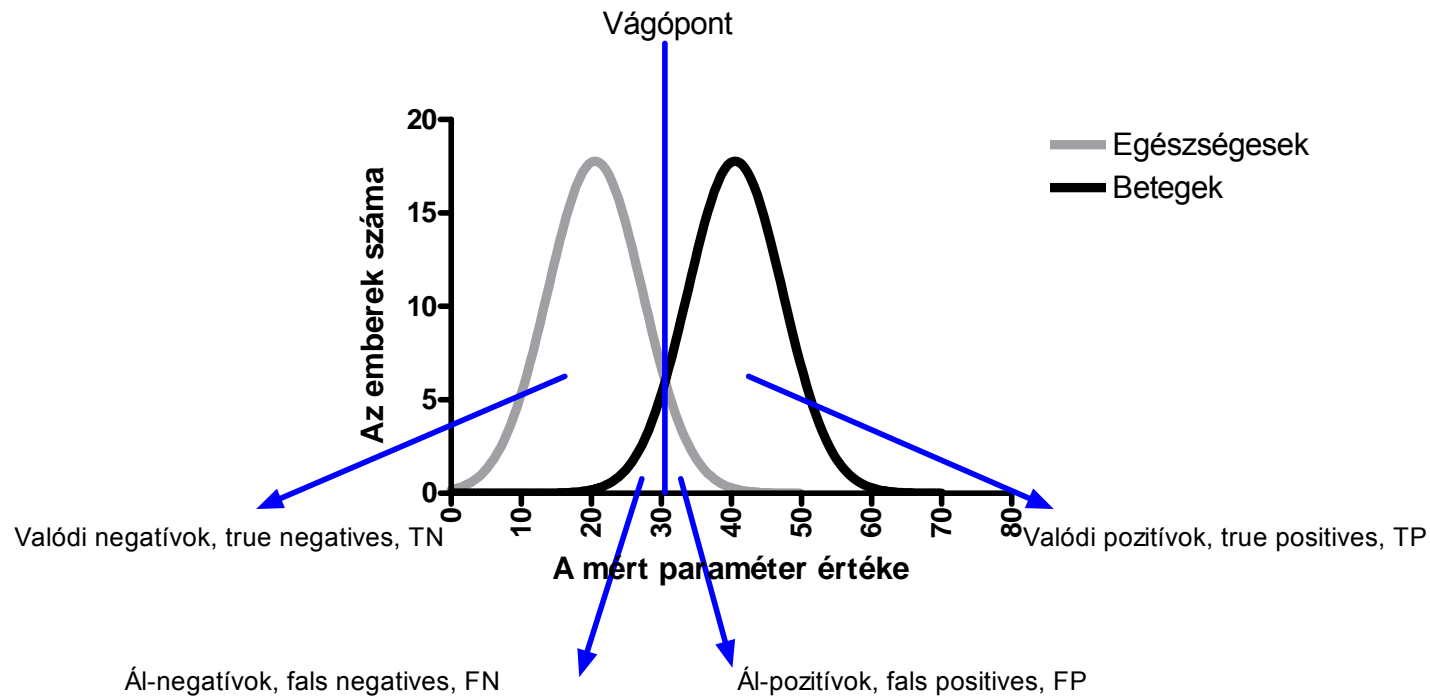
- A csoport-definíció alapvető eleme a vágópont (cut-off point, co)
- De hova tegyük a vágópontot?



A vágópontok megtalálásának gyakorlati kérdései

- **Referens** minta eloszlását figyelembe vevő módszerek
 - Átlag+1SD, vagy 2SD, 3SD, vagy percentilisek
- Az adott megfigyelési csoportok értékeit figyelembe vevő eljárás
 - kontroll csoport, vagy kezelés előtti minta értékei, átlag+SD, percentilisek
- Optimalizálásra törekvő (professzionális) módszer: receiver-operator characteristics plot, vagy ROC analízis

- A vágópont kialakításával kapcsolatos problémák az egyes kategóriák közötti átfedésekből származnak



Az ember \ A teszt	Beteg	Egészséges	Összesen
Pozitív	TP (valódi pozitív)	FP (ál-pozitív)	Összes pozitív lelet
Negatív	FN (ál-negatív)	TN (valódi negatív)	Összes negatív lelet
Összesen	Összes beteg	Összes egészséges	



PPV
 $TP/(TP+FP)$



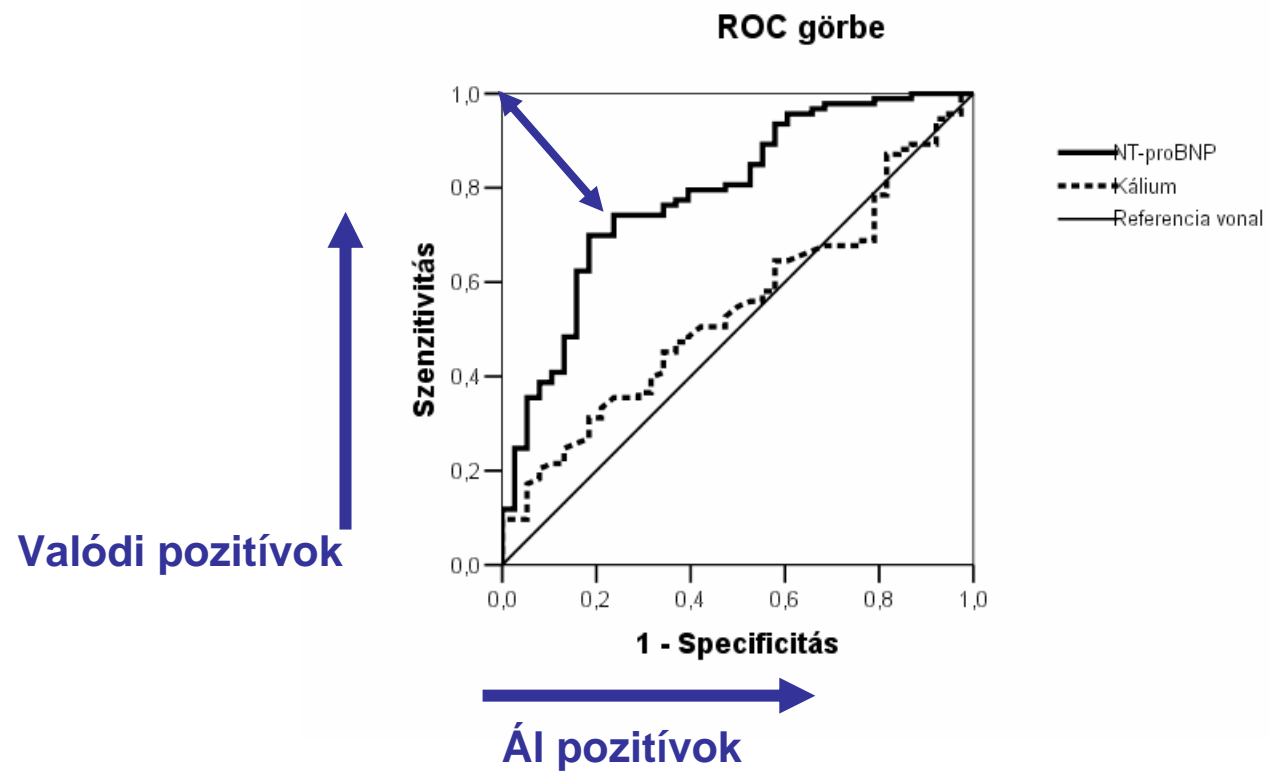
NPV
 $TN/(TN+FN)$



SEN
 $TP/(TP+FN)$



SPEC
 $TN/(TN+FP)$



Változó	Terület (AUC)	Standard hiba*	Szignifikancia**	95%-os konfidencia intervallum	
				Alsó határ	Felső határ
NT-proBNP	0,762	0,047	<0,0001	0,671	0,853
Szérum kálium	0,546	0,054	0,415	0,441	0,651

Összefoglalás


- Diszkrét vagy diszkrétte alakított változóink bemutatásának és értékelésének módja kontingencia táblázat
- Az alkalmazandó teszt a Chi-négyzet próba (vagy Fisher's exact teszt)
- Az egyes értékek előfordulását hányadosokkal, arányszámokkal lehet bemutatni
- A hányadosok és arányszámok összehasonlításában segítenek az esélyhányados és relatív rizikó fogalmak
- Dichotomizálás, ROC analízis
- A témával kapcsolatos további (ma nem említett) területek: adjusztált ráták, incidencia, prevalencia

Biostatistika a klinikumban - Microsoft Internet Explorer

Fájl Szerkesztés Nézet Kedvencek Eszközök Súgó

Cím http://www.semmelweiskiado.hu/kiadvanyok/index.php?item_Id=433

Google biostatistika a klinikumban Go Bookmarks 34 blocked Check AutoLink AutoFill Send to biostatistika a Settings

  **Simmelweis Kiadó és Multimédia Stúdió**

1089 Bp., Nagyvárad tér 4.
Levélcíme: 1445 Bp., Pf. 370
E-mail, Tel.: 210-4403
Fax: 210-1904, 459-1500/56471
Legendus Könyvesbolt:
1089 Bp., Nagyvárad tér 4., Aula
Tel./Fax: 210-4408

Keresés ...

Nyitólap A kiadóról Multimédia Stúdió Szolgáltatásaink Akciók ... És még Vásárlás PoszterBox Letöltések

Kiadványaink

- Újdonságaink
- Szakkönyvek
- Egyetemi, főiskolai tankönyvek
- Egyetemi, főiskolai jegyzetek
- Orvostörténeti kiadványok
- Ismeretterjesztő kiadványok
- Egyéb kiadványaink
- Multimédiás kiadványaink (CD,DVD)

Biostatistika a klinikumban

Prohászka Zoltán, Füst György, Dinya Elek:
Biostatistika a klinikumban

Kiadás éve: 0000
ISBN:
Kötés:
Oldalszám: 0
Kiadó: Semmelweis Kiadó és Multimédia Stúdió

Legendus ár: kalkuláció alatt
Fogyasztói ár: kalkuláció alatt

Bejelentkezés

e-mail címe
.....
OK
[Regisztráció »](#)
Elfelejtette a jelszavát?

Kosár

Az Ön kosara még üres. Tekintse meg [kiadványainkat](#), illetve a [Legendus könyvesbolt bővített kínálatát](#), reméljük talál köztük kedvére valót és vásárol!

Legkeresettebb

Start Beér... Tota... 2 M... Micr... Blost... HU 11:42