

**Szent István Egyetem**  
**Állatorvos-tudományi Doktori Iskola**

**Epidemiológiai és ökológiai mutatók  
statisztikai és valószínűségelméleti modellezése,  
térbeli szerkezetének vizsgálata**

PhD értekezés

Lang Zsolt

2014

Témavezető és témabizottsági tagok:

.....  
Prof. Dr. Reiczigel Jenő  
Szent István Egyetem,  
Biomatematikai és Számítástechnikai Tanszék  
témavezető

Prof. Dr. Rózsa Lajos  
MTA-MTM Állatökológiai Kutatócsoport  
témabizottság tagja

Dr. Harnos Andrea  
Szent István Egyetem,  
Biomatematikai és Számítástechnikai Tanszék  
témabizottság tagja

Dr. Franz Rubel  
Vet. Med. Univ. Wien,  
Dept. für Naturwiss.  
témabizottság tagja

Készült 8 példányban. Ez a n. .... sz. példány.

.....  
Lang Zsolt

# Tartalomjegyzék

1	Összefoglalás.....	6
2	Konfidenciaintervallum szerkesztése valódi prevalenciára független mintákból becsült szenzitivitás és specifikusság alapján.....	10
2.1	Bevezetés.....	10
2.2	Módszerek.....	11
2.3	Eredmények.....	15
2.4	Alkalmazási példák.....	19
2.5	Megvitatás.....	21
3	Szeropozitivitás és prevalencia összefüggései hazai BHV-1 mentesítés adatainak járványtani elemzése alapján.....	26
3.1	Bevezetés.....	26
3.2	Anyag és módszer.....	27
3.2.1	Statisztikai módszerek.....	28
3.3	Eredmények.....	30
3.4	Megvitatás.....	34
3.5	Következtetések.....	37
4	Zsúfoltság és diverzitás.....	38
4.1	Kapcsolat a zsúfoltság és a diverzitás fogalma között.....	38
4.2	Két szempontos klasszifikáció, a zsúfoltság additív felbontása.....	46
4.3	A valódi diverzitás és az aggregáltsági index multiplikatív felbontása.....	56
5	Egyedekből és ezek diszjunkt csoportjaiból álló populációk valószínűségelméleti modellezése.....	69
6	A vetési varjún ( <i>Corvus frugilegus</i> ) és a dolmányos varjún ( <i>Corvus cornix</i> ) élő tetűparaziták abundanciája, zsúfoltsága, diverzitása és aggregáltsága.....	82
6.1	Vetési varjú.....	82
6.1.1	Adatok, leíró statisztikák.....	82
6.1.2	Zsúfoltságok additív felbontása.....	84
6.1.3	A valódi diverzitás és az aggregáltság multiplikatív felbontása.....	88
6.1.4	Gazdák és paraziták csoportosítása.....	89
6.1.5	Korrespondenciaanalízis.....	90
6.2	Dolmányos varjú.....	92
6.2.1	Adatok, leíró statisztikák.....	92
6.2.2	Zsúfoltságok additív felbontása.....	93
6.2.3	A valódi diverzitás és az aggregáltság multiplikatív felbontása.....	98
6.2.4	Gazdák és paraziták csoportosítása.....	99
6.2.5	Korrespondenciaanalízis.....	100
6.3	A vetési varjú és a dolmányos varjú parazitáltságának összehasonlítása.....	102

7	A szarvaskullancslégy ( <i>Lipoptena cervi</i> ) ivararányának és a gímszarvas gazda ( <i>Cervus elaphus</i> ) parazitáltságának kapcsolata.....	104
7.1	Adatok .....	104
7.2	Statisztikai módszerek .....	104
7.3	Leíró statisztikák .....	105
7.4	A szarvaskullancslegyek ivararányának és abundanciájának kapcsolata.....	107
7.5	Megvitatás .....	108
8	Új tudományos eredmények .....	110
9	Irodalomjegyzék .....	115
10	A doktori kutatás eredményeinek közlései .....	123
11	A doktori kutatáshoz nem kapcsolódó közlemények.....	124
12	Köszönetnyilvánítás.....	127

## Fontosabb rövidítések jegyzéke

A	aggregáltsági index
$\alpha$	alfa diverzitás
$A_\alpha$	alfa aggregáltsági index
ANOVA	varianciaanalízis
AP	látszólagos prevalencia
$A_\beta$	béta aggregáltsági index
$\beta$	béta diverzitás
$A_\gamma$	gamma aggregáltsági index
C	zsúfoltsági index
CI	konfidenciaintervallum
Cov	kovariancia
D	diverzitási index
E	várható érték
G	tesztcsoport
$\gamma$	gamma diverzitás
p	p-érték
p	részarány
P	valószínűség
P	prevalencia
Se	szenszitivitás
Sp	specifikusság
Var	variancia
$Z_{crit}$	a standard normális eloszlás kritikus értéke
$\omega$	megfigyelő egyed

# 1 Összefoglalás

Betegségek prevalenciáját általában olyan diagnosztikai tesztekkel állapítják meg, amelyek téves eredményt is adhatnak. Rogan és Gladen (1978) módszert közölt a valódi prevalencia becslésére a diagnosztikai eljárás szenzitivitásának és specifikusságának figyelembe vételével. Reiczigel et al. (2010) egzakt konfidenciaintervallumot szerkesztett a valódi prevalenciára a pontos szenzitivitás és specifikusság ismeretét feltételezve. Új módszert ismertetünk a valódi prevalencia közelítő konfidenciaintervallumának megszerkesztésére abban az esetben, amikor a szenzitivitást és a specifikusságot független mintákból becsüljük. A lefedési valószínűség javítása céljából Rogan és Gladen (1978) konfidenciaintervallumának határait Agresti és Coull (1998) eljárásához hasonló módon finomítottuk. A korrekció 95%-os nominális szint mellett azt jelenti, hogy a szenzitivitás és specifikusság mintáiban a helyesen és tévesen diagnosztizált esetek számát 1-gyel, a célpopulációból származó mintában a pozitív és negatív esetek számát 2-vel növeljük. Kiterjedt számítógépes szimulációs kísérletekkel igazoltuk, hogy az új konfidenciaintervallum elég pontosan megtartja az előírt megbízhatósági szintet még kicsi, 30 elemű mintákra is; a lefedés legalább 88%, 93% és 98%, ha az előírt szint rendre 90%, 95% és 99%. Az új eljárást valós megfigyelések szakirodalomban közölt adatainak újbóli feldolgozásával, mint esettanulmányokkal illusztráljuk. Megvizsgáltuk továbbá a torzítás mértékét olyankor, ha a szenzitivitást és specifikusságot ismertnek feltételezzük, pedig valójában mintából becsültük. Kimutattuk, hogy a torzítás csak akkor elhanyagolható, ha a szenzitivitás és specifikusság mintája legalább 5-ször nagyobb a célpopulációból vett minta nagyságánál (Lang és Reiczigel 2014).

A magyarországi szarvasmarha-állományok BHV-1 mentesítését gE-negatív markervakcinákkal végzik 2002 óta. A szeropozitivásban 2006-ig bekövetkezett telepi változások alapján az adatok újrafeldolgozásával, a BHV-1 ellenanyagok kimutatása során alkalmazott diagnosztikai tesztek szenzitivitását és specifikusságát felhasználva, a téves pozitív és téves negatív arányok figyelembe vételével a szeropozitivitásból megbecsültük a BHV-1 fertőzöttség valódi előfordulási arányát, prevalenciáját. A becslést Rogan és Gladen (1978) módszerével végeztük el. A valódi prevalencia ismerete lehetővé tette a mentesítés hatásának pontosabb értékelését, a fertőzés korcsoportok, megyék, megyecsoportok és telepek közötti heterogenitásának és terjedési lehetőségének statisztikai vizsgálatát.

A fertőzés telepek közötti terjedésének jellemzésére az előbbieket mellett egy ökológiai mutató, a zsúfoltsági index alkalmazását is javasoljuk. A bemutatott statisztikai módszerek más fertőző betegségek elleni mentesítések előrehaladásának nyomon követésére is

alkalmasak, és segítségükkel a mentesítési programokban esetlegesen szükséges módosítások idejében meghozhatók. Kimutattuk, hogy a valódi prevalencia mentesítéskor nagyobb, ellenőrzéskor kisebb, mint a szeropozitivitás, így a mentesítés valójában sikeresebb volt, mint amit a szeropozitivitás összehasonlítása jelzett. A mentesítés sikere területenként eltérő, megyéken belül nagyon heterogén, ami az eltérő telep méretekkel, nagy állományi zsúfoltsággal együtt a fertőzés újbóli elterjedésének kockázatát jelentheti. Az állományok létszáma és a fertőzés prevalenciája között mentesítés előtt pozitív, mentesítés után negatív kapcsolatot tártunk fel (Lang et al. 2013).

Egyedek csoportokban élő közösségein, például gazdák parazitái körében a zsúfoltságot, az egyedszemponitú csoportméretet a csoportok monoton növekvő függvényével skálázott egyedszámának egyedekre vett átlagával értelmezzük (Lloyd 1967, Reiczigel et al. 2005, 2008). Szoros fogalmi kapcsolatot, analógiát tártam fel a zsúfoltság és a fajok biológiai diverzitása között. A biológiai diverzitást Patil és Taille (1982) mint a fajok skálázott ritkaságának egyedekre vett átlagát értelmezték. A skálázó függvény megadásával a zsúfoltság és a diverzitás egymással kifejezhetők. Az egymásnak megfelelő legfontosabb zsúfoltsági és diverzitási indexek lineáris skálafüggvényre a Lloyd- és Reiczigel-féle lineáris zsúfoltság és a Simpson-féle diverzitás, logaritmus függvényre a logaritmusos zsúfoltság és a Shannon-féle diverzitási index, hiperbolikus függvényre a hiperbolikus zsúfoltság és a fajgazdagság (a fajok száma mínusz 1).

A diverzitások elméletében értelmezett effektív fajszám megfelelőjeként a zsúfoltság fogalomkörében tetszőleges monoton növekvő skálafüggvény mellett bevezettem az effektív csoportszám fogalmát. Ez olyan egyenlő egyedszámú csoportok számát jelenti, ahol a csoportok egyedeinek zsúfoltsága a vizsgált közösség zsúfoltságával egyezik meg. Erre a fogalomra építve, a tényleges csoportszám és az effektív csoportszám hányadosaként tetszőleges monoton növekvő skálafüggvényre általánosítottam a Bez-féle aggregáltsági indexet (Bez 2000).

Kimutattam, hogy a zsúfoltság és a diverzitás között közvetlen, nem csak a fogalmi analógia szintjén megnyilvánuló kapcsolat is van olyankor, ha a vizsgált közösség csoportjaiban az egyedek több fajhoz tartoznak. A fajok együttesére vett zsúfoltságot felbontottam a fajonkénti átlagos zsúfoltság és az átlagos csoporton belüli, szubjektív diverzitás összegére. Az additív kapcsolaton túl multiplikatív összefüggést is találtam. Ismeretes, hogy több fajból álló csoportok közösségének együttes, ún. gamma valódi diverzitása (azaz effektív fajszáma) a csoportokon belüli átlagos valódi diverzitás (ún. alfa diverzitás) és a csoportok közötti változatosságot kifejező béta diverzitás szorzata (Whittaker 1960, 1972, Hill 1973, Jost 2006, 2007, Tuomisto 2010a). A zsúfoltság–diverzitás fogalmi megfeleltetésénél a valódi

diverzitásnak vagy más néven effektív fajszámnak az effektív csoportszám, ill. az aggregáltsági index felel meg. Nevezzük a teljes közösség aggregáltságát gamma aggregáltságnak, a fajonkénti átlagos aggregáltságot alfa aggregáltságnak. A béta aggregáltság legyen az alfa és gamma aggregáltságok hányadosa. Ez az aggregáltságnak a fajok közötti változatosságát kifejező tényezője.

A zsúfoltságra és a diverzitásra kapott elméleti eredményeket felhasználtam vetési varjak (*Corvus frugilegus*) és dolmányos varjak (*Corvus cornix*) tetűparazitáltságának összehasonlítására. A paraziták zsúfoltsága vetési varjakon nagyobb, emellett a zsúfoltság szubjektív diverzitásos összetevőjének is nagyobb a százalékaránya. A varjófajok közötti eltérést magyarázhatja, hogy a fertőzött dolmányos varjak körében nagyobb arányú az egyetlen tetűfajjal fertőzöttek, így 0 diverzitásúak aránya (50%), mint a fertőzött vetési varjakon (21%).

A tetvek valódi alfa és gamma diverzitásai a vetési varjak esetében valamivel nagyobbak, mint a dolmányos varjaknál. A béta diverzitások a két varjófajnál nagyjából megegyeznek, értékük 1.5-2.5 között van. A tetűfajok átlagos, azaz alfa aggregáltsága a dolmányos varjak körében nagyobb, mint a vetési varjaknál. Ez azt jelenti, hogy a dolmányos varjak tetűfajai átlagosan aggregáltabban oszlanak meg a gazdaállatok között, mint a vetési varjak tetűfajai. A béta aggregáltságok értéke mindkét varjófajnál nagyjából 1 és 2 közötti, ez alapján a vetési varjak génuszait két, gazdákon kevésbé átfedő csoportra, a *Brueelia*, illetve a *Myrsidea* és *Menacanthus* együttesére bontottam blokk klaszterezéssel és korrespondanciaelemzéssel. A dolmányos varjakon a két domináns génusz, a *Myrsidea* és a *Menacanthus* alkot a gazdákon csak kismértékű átfedést mutató csoportot. Érdekes, hogy a vetési varjakon – a dolmányos varjakkal ellentétben – a *Myrsidea* és *Menacanthus* génuszok nem különülnek el, gyakran előfordulnak közös gazdán.

Általános valószínűségelméleti modellt vezettem be csoportokba szerveződő egyedek jellemzőinek egyedszemponitú és csoportszintű eloszlásaira. Képleteket adtam a két szemléletmód közötti alapvető transzformációkra. Több elégséges és szükséges és elégséges feltételt fogalmaztam meg arra, hogy egy általános regressziós modell milyen feltételekkel egyezik meg a két szemléletmód mellett. Ezek egyike az, hogy a csoportok egyedszáma szerepeljen a regressziós modell magyarázó változói között. A többi ennek az egyszerű feltételnek különböző irányú matematikai finomítása.

Az említett valószínűségelméleti modellt és a hozzá kapcsolódó elméleti eredményeket alkalmaztam gímszarvasokat (*Cervus elaphus*) parazitáló szarvaskullancslegyek (*Lipoptena cervi*) ivararánya és a parazitáltság mértéke közötti összefüggés vizsgálatára. Az adatokra



kevert hatásos logisztikus regressziót illesztettem úgy, hogy a gímszarvasokon talált paraziták száma a modell egyik magyarázó változója. Ez biztosította, hogy a regressziós modell ugyanaz legyen a gímszarvasok és a szarvaskullancslegyek szempontjából. Kimutattam, hogy a hím szarvaskullancslegyek részaránya szignifikánsan kisebb az erősebben parazitált gímszarvasokon ( $p = 0.0001$ ). Az eredmény érvényes mind a gímszarvasokra, mint egymástól független entitásokra, megfigyelési egységekre, mind pedig az egyes kullancslegyek saját gazdáján élő mikropopulációkra. A regressziós modell eredményéből levezettem, hogy míg egy tipikus gímszarvason az ivararány nem különbözik szignifikánsan az 1:1 aránytól, egy tipikus kullancslégy gazdakörnyezetén az ivararány a nőstények irányába tolódik el. A paradoxon magyarázata az, hogy egy tipikus kullancslégy az átlagosnál jobban parazitált gímszarvason él.

## **2 Konfidenciaintervallum szerkesztése valódi prevalenciára független mintákból becsült szenzitivitás és specifikusság alapján**

Betegségek prevalenciáját általában olyan diagnosztikai tesztekkel állapítják meg, amelyek téves eredményt is adhatnak. Rogan és Gladen (1978) módszert közölt a valódi prevalencia becslésére a diagnosztikai eljárás szenzitivitásának és specifikusságának figyelembe vételével. Reiczigel et al. (2010) egzakt konfidenciaintervallumot konstruált a valódi prevalenciára a pontos szenzitivitás és specifikusság ismeretét feltételezve. Ebben a fejezetben egy új módszert ismertetünk a valódi prevalencia közelítő konfidenciaintervallumának megszerkesztésére abban az esetben, amikor a szenzitivitást és a specifikusságot független mintákból becsüljük. A lefedés (coverage) javítása céljából Rogan és Gladen (1978) konfidenciaintervallumának határait Agresti és Coull (1998) eljárásához hasonló módon finomítottuk. Kiterjedt számítógépes szimulációs kísérletekkel igazoltuk, hogy az új konfidenciaintervallum elég pontosan megtartja az előírt megbízhatósági szintet még kicsi, 30 elemű mintákra is; a lefedés legalább 88%, 93% és 98%, ha az előírt szint rendre 90%, 95% és 99%. Az új eljárás előnyeit valós megfigyelések adatainak feldolgozásával illusztráljuk.

### **2.1 Bevezetés**

A prevalencia becslése, standard hibájának meghatározása, konfidenciaintervallumának megszerkesztése az epidemiológiai elemzések egyik legalapvetőbb feladata. Felmérésekről közölt elemzésekben sok esetben kimondatlanul, implicit módon feltételezik, hogy a diagnosztikai teszt szenzitivitása és specifikussága 1. Ez a feltételezés azonban gyakran nem megalapozott. Rogan és Gladen (1978) a diagnosztikai teszt szenzitivitásával és specifikusságával korrigálta a prevalencia pontbecslését, megadva a becslés aszimptotikus standard hibáját is. Reiczigel et al. (2010) a prevalenciához egzakt kétoldali konfidenciaintervallumot szerkesztett, a szenzitivitás és specifikusság ismeretét feltételezve. Ha a szenzitivitást és a prevalenciát a priori nem ismerjük, és értéküket mintából becsüljük, akkor ezeknek a becsléseknek a véletlen hibái meg fogják növelni a prevalencia becslésének varianciáját. Ezt figyelembe kell venni a konfidenciaintervallum konstruálása során. Sok közleményben nincs adat a szenzitivitás és specifikusság becslésére szolgáló minták nagyságáról, vagy erre nézve más publikációkra utalnak (Boadella et al., 2012; Pinho et al., 2013; Sarrazin et al., 2013). Gyakran a szenzitivitás és specifikusság becslése 100-nál nem nagyobb elemszámú mintákból történik (irodalmi áttekintést közöl Farnham et al.,

2012). Ha ilyen esetekben nem vesszük tekintetbe a szenzitivitás és specifikusság becsléseinek bizonytalanságát, akkor tévesen kicsi standard hibákat és/vagy túl szűk konfidenciaintervallumokat kapunk eredményül. Több szerző tanulmányozta a prevalencia becslését a szenzitivitásra és a specifikusságra tett különböző feltételezések mellett. Greiner és Gardner (2000a, 2000b) a probléma hatékony megoldására különböző vizsgálati elrendezéseket ismertet. Messam et al. (2008) áttekinti a prevalencia diagnosztikai paraméterekkel korrigált becslésének frekventista és a Bayes-i módszereit. Ezek a megközelítések legtöbbször feltételezik, hogy létezik ún. gold standard teszt, amivel a betegség státusza pontosan megállapítható. Más modellek esetében nincs szükség gold standard tesztre (Hui és Walter, 1980; Enøe, Georgiadis és Johnson, 2000; Toft, Jørgensen és Højsgaard, 2005).

Rogan és Gladen (1978) a prevalencia becslésének varianciájára is közöltek számítási eljárást, azonban a normális eloszlás feltételezésével szerkesztett konfidenciaintervallumuk sokszor rossz hatásfokú (Greiner és Gardner 2000a). A Reiczigel et al. (2010) szerinti egzakt számítási módszer a két kolonc paraméter (közismertebb nevén nuisance paraméter) mellett nem kivitelezhető, ezért Agresti és Coull (1998) „plusz 2 siker, 2 kudarc” nevű eljárására támaszkodtunk, ami hatékonynak bizonyult. Az új konfidenciaintervallum ugyan nem egzakt, de a nominális megbízhatósági szintet elég pontosan megtartja még viszonylag kis mintanagyság mellett is. (A nominális megbízhatósági szint az előírt vagy elvárt valószínűsége annak, hogy a CI lefedje a valódi paraméter értékét).

A továbbiak során bemutatjuk a javasolt eljárást és szimulációs kísérletekkel megállapítjuk a konfidenciaintervallumok tényleges lefedési valószínűségét különböző szenzitivitás, specifikusság és prevalencia értékek mellett. (A tényleges lefedési valószínűség annak valószínűsége, hogy a CI valóban tartalmazza, lefedje a paraméter értékét, ebben az esetben a valódi prevalenciát.) Hogy az új módszert illusztráljuk és demonstráljuk, miért fontos számításba venni a szenzitivitás és a specifikusság becslésének bizonytalanságát, néhány példában eljárásunkat valódi megfigyelésekre, publikációkból vett adatokra alkalmaztuk. A szimulációk részletes eredményei, a javasolt eljárást megvalósító Excel fájl és R programkód letölthető a <http://www.univet.hu/users/jreiczig/prevalence-with-se-sp.html> webcímről.

## 2.2 Módszerek

Feltételezzük, hogy három egymástól független minta áll rendelkezésre:

- az első olyan egyedekből áll, amelyekről ismeretes, hogy a vizsgált betegségben szenvednek (ebből becsüljük meg a diagnosztikai teszt szenzitivitását),

- a másodikba olyan egyedek tartoznak, amelyek a betegségben nem szenvednek (ebből becsüljük a teszt specifikusságát),
- a harmadik egy véletlen minta a célpopulációból, melynek egyedeit a diagnosztikai teszttel vizsgáljuk (az egyedek valódi betegségstátusza nem ismert, csak a teszt eredményét – pozitív vagy negatív – ismerjük).

A prevalencia populációs értékére szeretnénk konfidenciaintervallumot konstruálni. Jelölje az ismeretlen szenzitivitás, specifikusság, látszólagos és valódi prevalencia értékét  $Se$ ,  $Sp$ ,  $AP$  és  $P$ , a megfelelő becsült mennyiségeket  $\widehat{Se}$ ,  $\widehat{Sp}$ ,  $\widehat{AP}$  és  $\widehat{P}$ , a minták nagyságát  $n_{Se}$ ,  $n_{Sp}$ ,  $n_p$ . Megjegyezzük, hogy az  $AP$  látszólagos prevalencia a teszt pozitívok aránya a célpopulációban, ezért a mintában a teszt pozitívok számának eloszlása  $n_p$  és  $AP$  paraméterű binomiális. A valódi prevalencia a

$$P = (AP + Sp - 1)/(Se + Sp - 1) \quad (1)$$

képlettel fejezhető ki, amiből  $P$ -re a

$$\widehat{P} = (\widehat{AP} + \widehat{Sp} - 1)/(\widehat{Se} + \widehat{Sp} - 1). \quad (2)$$

plug-in becslést nyerjük (Rogan és Gladen, 1978).

A  $\widehat{P}$  becsült valódi prevalencia aszimptotikus varianciáját (2) Taylor-sorából kaphatjuk meg (Rogan és Gladen 1978). Ez valójában az ún. delta módszer egy alkalmazása (Agresti 2002). A varianciára kapott becslés

$$var(\widehat{P}) = \left\{ var(\widehat{AP}) + \widehat{P}^2 var(\widehat{Se}) + (1 - \widehat{P})^2 var(\widehat{Sp}) \right\} / (\widehat{Se} + \widehat{Sp} - 1)^2 \quad (3)$$

vagy másképpen, a binomiális varianciákat (3)-ba helyettesítve

$$var(\widehat{P}) = \left\{ \widehat{AP}(1 - \widehat{AP})/n_p + \widehat{P}^2 \widehat{Se}(1 - \widehat{Se})/n_{Se} + (1 - \widehat{P})^2 \widehat{Sp}(1 - \widehat{Sp})/n_{Sp} \right\} / (\widehat{Se} + \widehat{Sp} - 1)^2. \quad (4)$$

Megjegyezzük, hogy  $\widehat{P}$  varianciájának (3) becslése pontosan akkor egyezik meg  $\widehat{AP}$  varianciájának becslésével, ha  $Se$  és  $Sp$  ismert és  $Se = Sp = 1$ . Minden más esetben  $var(\widehat{P}) > var(\widehat{AP})$ .

Ezeknek az eredményeknek alapján a valódi prevalencia legegyszerűbb, közelítőleg normális eloszlást feltételezõ, ún. Wald konfidenciaintervalluma

$$\widehat{P} \pm z_{crit} \cdot var(\widehat{P})^{\frac{1}{2}}. \quad (5)$$

Ha a CI határai 0-nál kisebbek vagy 1-nél nagyobbak, akkor a határokat 0-ra, ill. 1-re kell cserélni. Az (5) formulában  $z_{crit}$  a standard normális eloszlás előírt megbízhatósági szinthez tartozó kritikus értéke. Az aszimptotikus variancia jól közelíti a tényleges varianciát, ha a három mintanagyság elég nagy, és a három binomiális valószínűség elég távol esik a 0 és 1 peremértékektől. Számos, a gyakorlatban előforduló esetben a prevalencia kicsi és a szenzitivitás, specifikusság értéke közel van 1-hez. Ilyenkor a (4) formula 0-hoz közeli varianciát becsül és az (5) CI lefedési valószínűsége jóval a nominális szint alá kerül.

A lefedés javítása céljából hozzáadjuk  $\widehat{AP}$  számlálójához a  $z_{crit}^2/2$ , nevezőjéhez a  $z_{crit}^2$  korrekciókat. Agresti és Coull (1998) megmutatta, hogy ez a módszer jelentősen javítja a binomiális mintából becsült valószínűség konfidenciaintervallumának lefedési valószínűségét. Formálisan, ha

$$n'_p = n_p + z_{crit}^2 \quad (6)$$

$$AP' = (n_p \cdot \widehat{AP} + z_{crit}^2/2)/(n_p + z_{crit}^2), \quad (7)$$

akkor  $AP$  korigált konfidenciaintervalluma

$$AP' \pm z_{crit} \cdot (AP'(1 - AP')/n'_p)^{\frac{1}{2}}. \quad (8)$$

Abban a fontos speciális esetben, amikor az  $Se$  szenzitivitás és az  $Sp$  specifikusság értékét ismertnek tekintjük, az  $AP$  látszólagos prevalenciára szerkesztett Agresti-Coull konfidenciaintervallum az (1) összefüggéssel a  $P$  valódi prevalencia

$$P' \pm z_{crit}/(Se + Sp - 1) \cdot (AP'(1 - AP')/n'_p)^{\frac{1}{2}} \quad (9)$$

konfidenciaintervallumává alakítható, ahol

$$P' = (AP' + Sp - 1)/(Se + Sp - 1). \quad (10)$$

További korrekciókat javasolunk abban az esetben, amikor a szenzitivitás és specifikusság nem ismert és értéküket mintákból becsüljük. Legyen

$$n'_{se} = n_{se} + 2 \quad (11)$$

$$n'_{sp} = n_{sp} + 2 \quad (12)$$

$$Se' = (n_{se} \cdot \widehat{Se} + 1)/(n_{se} + 2) \quad (13)$$

$$Sp' = (n_{sp} \cdot \widehat{Sp} + 1)/(n_{sp} + 2) \quad (14)$$

$$P' = (AP' + Sp' - 1)/(Se' + Sp' - 1). \quad (15)$$

Ezekkel a jelölésekkel a korrigált  $var(P')$  varianciát úgy számítjuk ki, hogy a (4) formulában minden tagot a neki megfelelő, vesszővel jelölt korrigált változatra cserélünk. A  $P$  valódi prevalencia (9) konfidenciaintervalluma a korrekció következtében a

$$P' + dP \pm z_{crit} \cdot var(P')^{\frac{1}{2}} \quad (16)$$

alakot ölti, ahol

$$dP = 2 \cdot z_{crit}^2 \cdot \{P' \cdot Se'(1 - Se')/n'_{se} - (1 - P') \cdot Sp'(1 - Sp')/n'_{sp}\}. \quad (17)$$

A korrekciót megalapozó számítások részletei megtalálhatók Lang és Reiczigel (2014) közleményében.

A kapott konfidenciaintervallumot ugyanúgy 0 és 1 közé korlátozzuk, mint (5) esetében. Fontos megemlíteni, hogy az intervallum  $P' + dP$  középpontja  $P$ -nek a (2) Rogan-Gladen becslésnél jobban torzító becslése, ezért a valódi prevalencia pontbecslésére az utóbbi alkalmasabb. A  $\hat{P}$  Rogan-Gladen becslés a (16) korrigált konfidenciaintervallumba esik, ha  $n_{se} \geq 26, n_{sp} \geq 26, n_p \geq 16$  és  $\widehat{Se} \geq .5, \widehat{Sp} \geq .5$  (Lang és Reiczigel 2014).

Számítógépes szimulációval kiértékeljük a korrigált és az eredeti Wald konfidenciaintervallum lefedési valószínűségeit az alábbi scenáriókra.

- A valódi szenzitivitás és specifikusság vizsgált értékei 1, .99, .95, .90, .70.
- A szenzitivitás, specifikusság és prevalencia becslésére szolgáló minták nagysága 30, 100, 300, 1000, 3000.
- Az előírt, nominális megbízhatósági szint 90%, 95% és 99%.
- A lefedést a valódi prevalencia .005, .01, .02, .03, .05, .10, .20, .30, .50 értékeire állapítottuk meg.

A fenti beállítások minden kombinációjára 20000 véletlen mintát (valójában 20000 mintahármast) generáltunk és meghatároztuk belőlük a lefedési valószínűségeket.

Vegyük észre, hogy az eljárás tulajdonságai a valódi prevalencia .5-nél nagyobb értékeire megkaphatók az eljárás .5-nél kisebb értékekre vonatkozó tulajdonságaiból, ehhez csak az  $(Se, n_{se})$  és  $(Sp, n_{sp})$  párok szerepét kell felcserélni. Például a lefedési valószínűség  $P = .8, Se = .99, Sp = .9, n_{se} = n_{sp} = 100$  mellett ugyanaz, mint  $P = .2, Se = .9, Sp = .99, n_{se} = n_{sp} = 100$  esetében. Ezért a lefedési valószínűségek táblázataiban csak .5-nél nem nagyobb valódi prevalenciák szerepelnek.

Megvizsgáltuk a pontosság romlását mintákból becsült szenzitivitás és specifikusság mellett olyan konfidenciaintervallum szerkesztési eljárásoknál, amelyek ismert szenzitivitást és

specifikusságot feltételeznek. Szimulációs kísérleteket végeztünk Blaker és Clopper-Pearson módszerével szerkesztett konfidenciaintervallumokkal (Reiczig et al. 2010, Blaker 2000, Clopper és Pearson 1934). A szimulációk során a szenzitivitás és specifikusság becslésére szolgáló minták nagysága a prevalencia becsléséhez használt minta nagyságával megegyező, ill. kétszer, háromszor, ötször és tízszer nagyobb volt. A lefedési valószínűségeket 10000 véletlen mintahármas alapján határoztuk meg, különböző mintanagyságok, nominális megbízhatósági szintek, továbbá a valódi szenzitivitás és specifikusság .99-es, .95-ös és .90-es értékei mellett.

## 2.3 Eredmények

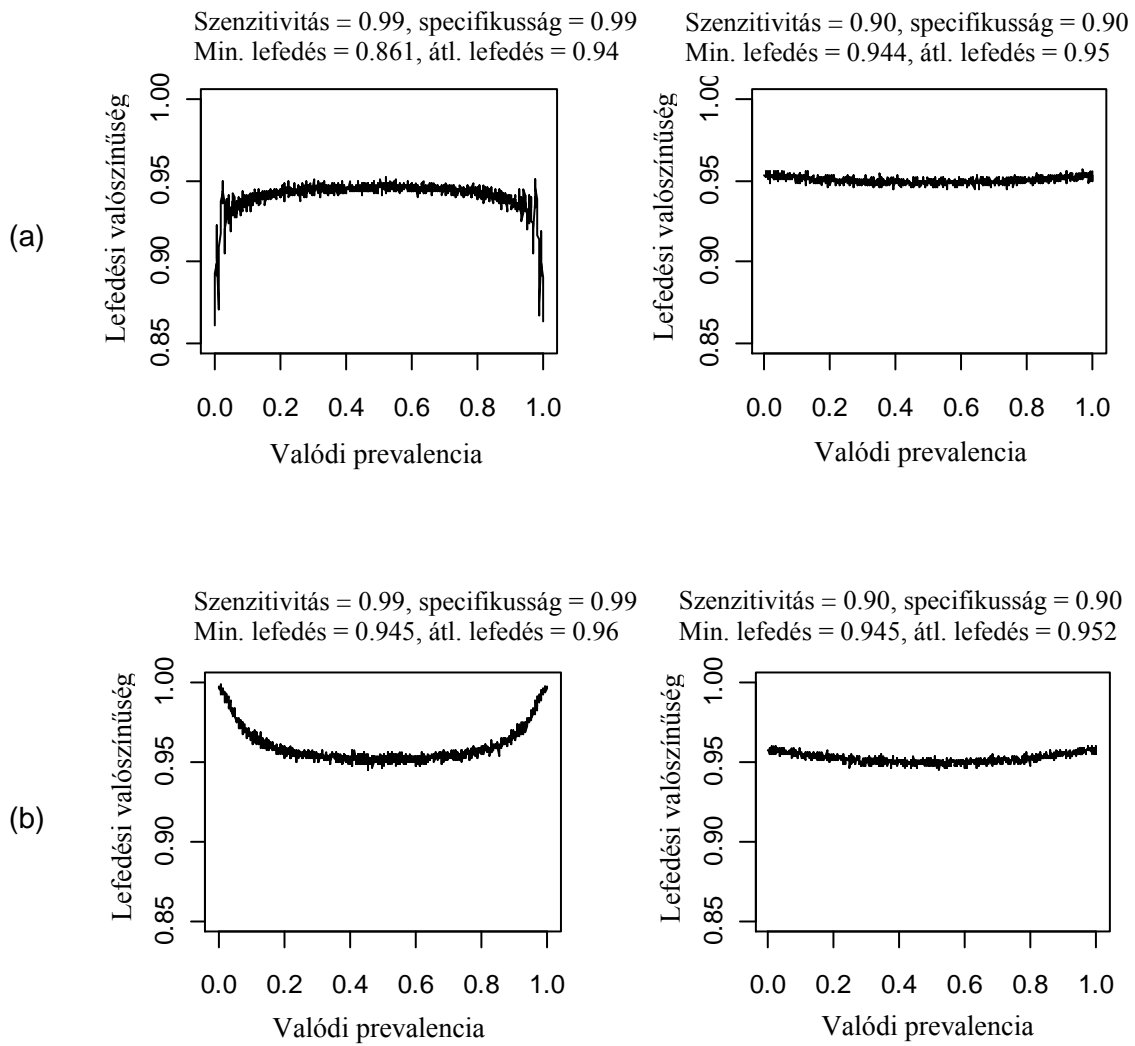
A normális eloszlással történő közelítésre támaszkodó és a korrigálatlan Rogan-Gladen varianciabecslésből számolt (5) konfidenciaintervallum lefedési valószínűsége túlságosan alacsony, számos esetben nominális 95% mellett kevesebb, mint 90%, nominális 90% mellett kevesebb, mint 80%, különösen olyankor, ha a szenzitivitás és a specifikusság értéke nagy, 1 közele. Az általunk javasolt módszer alkalmazása lényegesen javította a lefedési valószínűséget, ami a korrekció után minden vizsgált paraméterkombinációra, nominális megbízhatósági szintre elfogadhatóvá vált. A 95%-os nominális szint mellett a lefedés minimuma ritkán került 94% alá és mindig 93% fölött maradt. A 99%-os nominális szint mellett a lefedési valószínűség soha nem volt kisebb 98%-nál, és 90%-os nominális szint esetében nem lett kisebb 88%-nál. Az alkalmazások szempontjából legfontosabb paraméterek melletti minimális lefedési valószínűségeket mutatja be az 1. táblázat. A részletes eredmények a <http://www.univet.hu/users/jreiczig/prevalence-with-se-sp.html> honlapon tekinthetők meg. A lefedési valószínűség ugyanilyen pontosnak bizonyult az eddig említetteken túl még az olyan paraméterekre is, amelyeknél a Youden index  $Se + Sp - 1$  (Youden 1950) legalább .5, például ha a szenzitivitás .5 és a specifikusság 1 (az ezt alátámasztó szimulációs eredményeket nem közöltük). Az 1. és 2. ábra a valódi prevalencia függvényében mutatja a tényleges lefedési valószínűségeket 95%-os nominális megbízhatósági szint mellett, ha  $n_p = n_{se} = n_{sp} = 100$ ,  $Se = .9$  vagy  $.99$ ,  $Sp = .9$  vagy  $.99$ .

**1. táblázat**

*Szimulált minimális lefedési valószínűség 95% megbízhatósági szintű korrigált CI esetében, ha  $n_p=100$ ,  $n_{Se}=100$ ,  $n_{Sp}=100$  (20000 ismétlés alapján).*

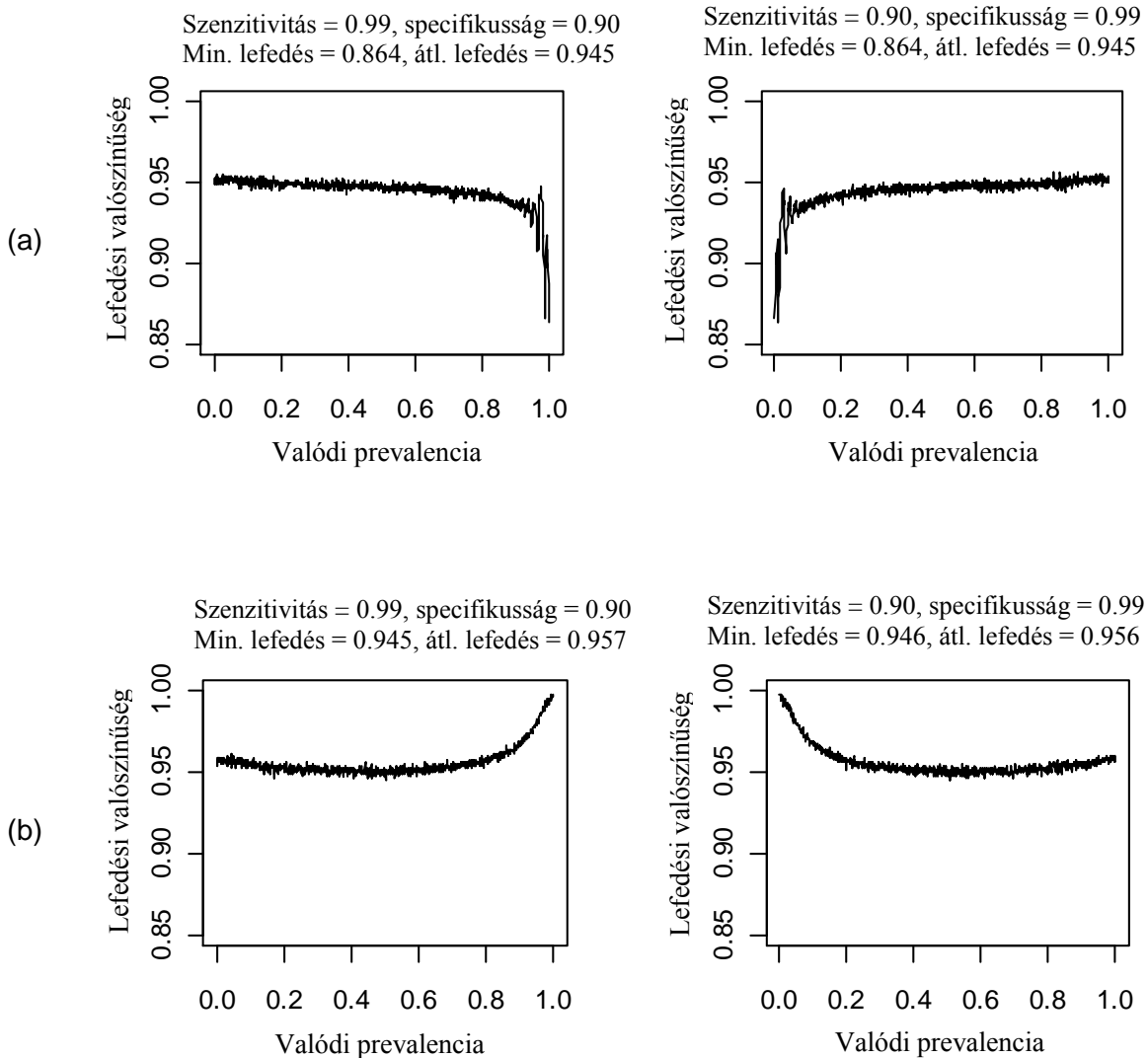
Valódi szenzitivitás	Valódi specifikusság	Valódi prevalencia								
		.005	.01	.02	.03	.05	.10	.20	.30	.50
.70	.70	.957	.961	.960	.959	.958	.960	.961	.958	.959
.70	.90	.960	.958	.958	.958	.956	.957	.954	.952	.950
.70	.95	.969	.967	.968	.967	.963	.963	.958	.953	.953
.70	.99	.996	.996	.995	.991	.983	.972	.963	.955	.952
.90	.70	.955	.953	.955	.953	.953	.953	.951	.950	.949
.90	.90	.957	.957	.958	.958	.956	.956	.954	.949	.952
.90	.95	.968	.969	.968	.966	.964	.963	.954	.952	.950
.90	.99	.993	.995	.990	.990	.980	.964	.959	.954	.948
.95	.70	.951	.952	.954	.954	.952	.952	.952	.954	.952
.95	.90	.961	.959	.956	.957	.959	.954	.954	.950	.950
.95	.95	.966	.969	.966	.966	.967	.960	.955	.951	.950
.95	.99	.993	.995	.991	.992	.980	.965	.959	.952	.956
.99	.70	.952	.953	.952	.951	.953	.951	.950	.950	.953
.99	.90	.958	.958	.956	.955	.957	.957	.955	.951	.948
.99	.95	.963	.968	.966	.967	.966	.962	.951	.951	.952
.99	.99	.994	.994	.994	.992	.984	.968	.957	.954	.950





**1. ábra**

*Szimulált lefedési görbék (20000 ismétlésből) korrekció nélkül (a) és korrekcióval (b), 95%-os nominális megbízhatósági szint mellett. Ha a szenzitivitás és a specifikusság megegyezik, akkor a lefedési görbék szimmetrikusak.*



2. ábra

*Szimulált lefedési görbék (20000 ismétlésből) korrekció nélkül (a) és korrekcióval (b), 95%-os nominális megbízhatósági szint mellett. Ha a szenzitivitás és a specifikusság különbözik, akkor a lefedési görbék nem szimmetrikusak. A szenzitivitás és specifikusság értékét felcserélve, a kapott lefedési görbék egymás tükörképei lesznek.*

A 2. táblázat a pontosság romlását illusztrálja ismert szenzitivitást és specifikusságot feltételező konfidenciaintervallum konstrukciókra abban az esetben, ha a szenzitivitást és a specifikusságot mintákból becsüljük. Azt tapasztaltuk, hogy a pontosság romlása csak akkor válik elhanyagolhatóvá, ha a szenzitivitás és specifikusság becslésére kiválasztott minták nagysága legalább ötször nagyobb, mint a prevalencia becslésére szolgáló minta. A pontosság romlásának mértéke ezen túlmenően függ a prevalencia mintájának nagyságától és a szenzitivitás és specifikusság valódi értékétől is.

## 2. táblázat

Minimális lefedési valószínűségek 95%-os megbízhatóságú Blaker konfidenciaintervallumokra, ha a szenzitivitást és a specifikusságot  $n_{Se}$  és  $n_{Sp}$  nagyságú mintákból becsüljük. (A szimuláció 10000 ismétlésből állt, a valódi szenzitivitás és specifikusság .90 volt).

$n_{prev}$	$n_{Se} = n_{Sp} = k \cdot n_{prev}$			
	$k = 1$	$k = 2$	$k = 3$	$k = 5$
100	.844	.899	.921	.935
300	.844	.907	.921	.933
1000	.846	.903	.916	.933
3000	.853	.906	.921	.937

## 2.4 Alkalmazási példák

### 1. A belgiumi szarvasmarha-állomány 1-es típusú bovin herpesvírus (BHV-1) prevalenciája

Boelaert et al. (2000) az 1-es típusú bovin herpesvírus prevalenciáját vizsgálta a belgiumi szarvasmarha populációban. Adataik szerint a nem vakcinázott állományokban 11284 állat közül 4060 egyed volt fertőzött. A látszólagos prevalencia .360 (95% CI: .351 – .369). Feltételezve, hogy a gB-ELISA teszt szenzitivitása .994 (Kramps et al. 1994), specifikussága .997 (de Wit et al. 1998), a valódi prevalencia becslése .362 (95% CI: .353 – .371).

Egy lépéssel továbbmenve, Kramps et al. (1994) közleményében olvashatjuk, hogy a szenzitivitást 179 elemű mintából becsülték, a vizsgálat ebből 178-at helyesen azonosított, így nyerték a szenzitivitásra a .994 becslést. A specifikusság értékét de Wit et al. (1998) 359 elemű mintából kapta, itt 358 esetben volt helyes az eredmény, amiből a specifikusság becslése .997. Mindezt figyelembe véve, a valódi prevalencia 95%-os megbízhatóságú konfidenciaintervalluma .349 – .372. A valódi prevalencia konfidenciaintervalluma kissé szélesebb lett, de a különbség nem jelentős ebben az esetben. Ennek oka az, hogy mind a szenzitivitás, mind a specifikusság .99 fölötti, mindkettőt aránylag nagy mintaszámmal becsülték, és a valódi prevalencia messze van 0-tól és 1-től.

## 2. *Trichomonas gallinae* prevalenciája házi pirókban (*Carpodacus mexicanus*)

Anderson et al. (2009) trichomonas protozoonok előfordulását tanulmányozták vadon élő énekesmadarak körében. Példaként vizsgáljuk meg a *Trichomonas gallinae* prevalenciáját házi pirókban (*Carpodacus mexicanus*). A mintába került 2971 madáregyed közül 51 esetben mutatták ki a parazitát, ezért a látszólagos prevalencia .017. A diagnosztikai eljárás szenzitivitását és specifikusságát a paraziták kitenyésztésével határozták meg. A szenzitivitásra .97 (32/33), a specifikusságra 1 (20/20) adódott eredményül. Ezekkel számolva, a valódi prevalencia pontbecslése .018.

Elfogadva, hogy .97 és 1 a szenzitivitás és specifikusság ismert pontos értékei, a Reiczigel et al. (2010) módszerével számított Blaker konfidenciaintervallum .013 – .023. Figyelembe véve azt is, hogy a szenzitivitást és specifikusságot nagyon kis, 33 és 20 elemszámú mintából becsülték, a valódi prevalencia 95%-os konfidenciaintervalluma 0 – .053. Ez a CI sokkal szélesebb, mint az ismert szenzitivitás és specifikusság feltételezése mellett kapott konfidenciaintervallum.

## 3. *Tuberculosis* prevalenciája ugandai tehenek körében

Faye et al. (2005) tuberculosis (TB) és brucellosis prevalenciáját becsülte meg ugandai tehenekben. Összesen 11862 állatot vizsgáltak TB teszttel és a látszólagos prevalenciára .06 értéket kaptak. Quirin et al. (2001) szerint a teszt szenzitivitása .80 (95% CI: .44 – .98), specifikussága 1 (95% CI: .74 – 1). Erre támaszkodva Faye et al. (2005) arra következtetett, hogy a valódi prevalencia (amit a látszólagos prevalenciából a szenzitivitással és a specifikussággal történő adjusztálással nyerünk) .10 – .23 között lehet, feltéve, hogy a specifikusság 1, vagy .08 – .19 közé esik, feltéve, hogy a specifikusság .98, vagy .05 – .15 között van, ha a specifikusság .95.

Az általunk javasolt módszer lehetővé teszi, hogy a valódi prevalenciára egyetlen érvényes konfidenciaintervallumot szerkesszünk. Ehhez szükségünk van azokra a mintanagyságokra, amikből Quirin et al. (2001) a TB teszt szenzitivitását és specifikusságát megbecsülték.

A minták elemszáma 10 volt a szenzitivitás, 12 a specifikusság becslésekor. Ezek nagyon alacsony értékek, aminek következtében óriási mértékben megnő a valódi prevalencia becslésének bizonytalansága. A 95%-os CI 0 – .147 lett, a pontbecslés .075. A konfidenciaintervallum tartalmazza a 0-t, ezért az sem zárható ki, hogy a TB populációs prevalenciája zéró, azaz a fertőzés nincs is jelen a populációban. Ha a szenzitivitás és a specifikusság valódi értéke .80 és 1 lenne, akkor a 95%-os CI sokkal szűkebb, .070 – .081 lenne Blaker módszerével számolva (Reiczigel et al., 2010), de most nem ez a helyzet. Meg kell még említeni, hogy ilyen alacsony mintanagyság mellett nincs arra garancia, hogy a CI lefedési valószínűsége elérje, közelítse a nominális megbízhatósági szintet.

#### *4. Vizelet inkontinencia prevalenciája kérdőíves felmérés alapján*

Sandvik et al. (1995) nők vizelet inkontinenciáját vizsgálta kérdőíves felméréssel Norvégiában. A projekt első részében a kérdőíveket 236 vizelet inkontinenciával ambuláns rendelőbe utalt betegen validálták. Három inkontinencia típust különböztettek meg, mi a stresszes és a sürgősségi inkontinenciára vonatkozó eredményeket használjuk fel. A vizsgálat validációs része alapján a kérdőívnek, mint a stresszes inkontinencia diagnosztikai eszközének a szenzitivitása  $84/127 = .661$ , a specifikussága  $96/109 = .881$ . A második részben 509 nő töltötte ki a kérdőívet, közülük 259-et diagnosztizáltak inkontinensnek. A teszt pozitívok aránya így  $.51$ . A szenzitivitással és a specifikussággal korrigálva, a valódi prevalencia  $.72$  (a közleményben megadott  $.77$  bizonyára nyomdahiba). A szenzitivitás és a specifikusság becsléséből adódó bizonytalanságot is figyelembe vevő 95%-os CI  $.58 - .87$ . A szenzitivitás és specifikusság valódi értékét ismertnek feltételezve a CI hossza sokkal rövidebb,  $.64 - .80$  lenne. A sürgősségi inkontinencia esetében a szenzitivitás  $23/41 = .561$ , a specifikusság  $187/195 = .959$ . A látszólagos prevalencia  $51/509 = .100$ , a valódi prevalencia  $.114$  (95% CI:  $.037 - .195$ ). A szenzitivitás és specifikusság pontos értékét ismertnek tekintve a CI  $.069 - .170$  lenne.

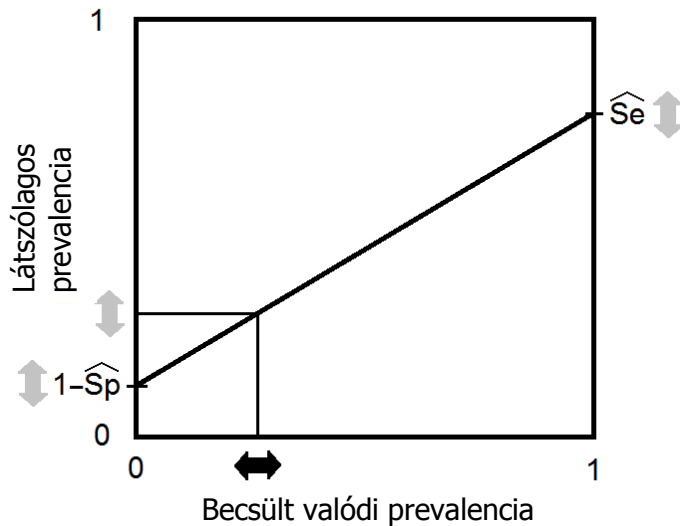
## **2.5 Megvitatás**

Vizsgálataink során azt tapasztaltuk, hogy a szenzitivitás és a specifikusság becsléséből származó bizonytalanság lényeges hatással lehet a prevalencia becslésének pontosságára. Amint azt vártuk, ha a szenzitivitás és a specifikusság becslésére szolgáló minták nagysága sokkal (5-10-szer) nagyobb, mint a prevalencia becslésére vett minta, akkor a CI tényleges lefedése a nominális szintet közelíti ( $>.93$ , ha a nominális  $.95$ ). Ha viszont a szenzitivitás és specifikusság mintáinak nagysága hasonló a prevalencia mintájának nagyságához (legfeljebb 3-szor nagyobb), akkor a CI tényleges lefedési valószínűsége túlzott mértékben lecsökkenhet a nominális szinthez képest ( $<.90$  is lehet nominális  $.95$  mellett). Nem meglepő, hogy a Clopper-Pearson módszer, ami erősen konzervatív ismert szenzitivitás és specifikusság mellett (Reiczigel et al. 2010), valamivel jobban teljesít a szenzitivitás és specifikusság becslése esetén, mint Blaker módszere (a szimulációs eredményeket nem közöljük). Sajnos a gyakorlati kutatások során a szenzitivitást és a specifikusságot rendszerint néhány száz elemű vagy annál is kisebb mintából becsülik, ugyanakkor szűrővizsgálatokban a diagnosztikai eljárást sokkal nagyobb (tipikusan néhány ezres) mintára alkalmazzák. Szimulációs eredményeink azt mutatják, hogy ilyen esetekben a szenzitivitást és specifikusságot ismertnek tekintve a valódi prevalenciára téves intervallumbecslést kapunk.

Az általunk javasolt CI jó lefedési valószínűséget biztosít tág határok között mozgó prevalencia, szenzitivitás és specifikusság értékek mellett, beleértve azt is, amikor a prevalencia közel van 0-hoz vagy 1-hez és a szenzitivitás, specifikusság értéke 1-hez közeli. Az adjusztálás miatt a CI kissé konzervatív, amikor a prevalencia 0 vagy 1 közelében van. A lefedés pontossága akkor is megmarad, ha a három minta nagysága jelentősen különbözik.

A becsült prevalencia variabilitásának összetevőit a 3. ábra illusztrálja. Ha a valódi prevalencia kicsi (0 közeli), akkor a prevalencia adjusztált becslésére a specifikusság becslésének bizonytalansága hat erősebben, míg nagy (1 közeli) prevalencia becslésének variációját inkább a becsült szenzitivitás pontossága befolyásolja. Közepes (.5 körüli) prevalencia variációjára a becsült szenzitivitás és specifikusság szórása azonos mértékben hat. Ezek szerint, ha a valódi prevalencia kicsi (0 közeli), akkor a becsült prevalencia variációját hatékonyan csökkenthetjük a specifikusságot becslő minta nagyságának növelésével, a szenzitivitás mintanagyságának változtatása ebben a vonatkozásban csaknem hatástalan. Ha a valódi prevalencia nagy (1 körüli), akkor a szerepek felcserélődnek, a variancia csökkentéséhez csak a szenzitivitás mintanagyságát érdemes növelni.

Ha a szenzitivitás és a specifikusság értéke ismert, akkor a CI hossza 0-hoz konvergál a mintanagyság növelésekor. Ha a szenzitivitás és a specifikusság nem tekinthető ismertnek és értéküket mintából becsüljük, akkor a CI hossza nem közelít 0-hoz, hanem egy bizonyos küszöbérték felett marad minden mintanagyság mellett. Ez az alsó határ függ a szenzitivitást és specifikusságot becslő minták nagyságától.



3. ábra

A becsült valódi prevalencia variációja (fekete nyíl) 3 független forrásból származik, ezek a becsült látszólagos prevalencia, szenzitivitás és specifikusság variabilitása (szürke nyilak). Az utóbbi 2 komponens súlya függ a valódi prevalenciától (az ábrán bemutatott esetben a specifikusság variabilitásának súlya a nagyobb).

Széles körben elterjedt vélekedés, hogy a populációs részarány  $\bar{p} \pm 1.96 \cdot var(\bar{p})^{\frac{1}{2}}$  formulával értelmezett 95%-os Wald-féle konfidenciaintervalluma elfogadható pontosságú, feltéve, hogy a  $\bar{p} \pm 3 \cdot var(\bar{p})^{\frac{1}{2}}$  próbaintervallum nem tartalmazza sem a 0-t, sem az 1-et, továbbá a becsléshez felhasznált minta nagysága „elég nagy”. Enøe et al. (2000) számos olyan hivatkozást közöl, amelyek megerősítik ezt az állítást. Greiner és Gardner (2000a) javasolják ennek alkalmazását az (5) Wald-Rogan-Gladen konfidenciaintervallumra. Számítógépes szimulációval vizsgáltuk (5) pontosságát, ha a  $\hat{P} \pm 3 \cdot var(\hat{P})^{\frac{1}{2}}$  próbaintervallum nem tartalmazta sem a 0-t, sem az 1-et. A mintanagyságok az  $n_p = 100, 1000, 3000$  és  $n_{se} = n_{sp} = 100, 1000, 3000$  értékek kombinációi voltak. A valódi prevalencia .001 és .999 között változott .001 lépésközzel. A szenzitivitás és specifikusság szintje .99, .95 és .90 volt. A szimulációs kísérlet a módszer két hátrányát tárta fel. A mintanagyság  $n_p = n_{se} = n_{sp} = 100$  beállítása mellett egyrészt, ha a szenzitivitás és a specifikusság .90, a valódi prevalencia  $<.3$  vagy  $>.7$  volt, akkor az esetek kb. 50%-ában a próbaintervallum tartalmazta a 0-t vagy az 1-et, így az eljárás egyáltalán nem szerkesztett konfidenciaintervallumot. Másrészt, a lefedési valószínűség még az eredeti Wald-Rogan-Gladen módszernél is pontatlanabb volt, különösen akkor, ha a valódi prevalencia  $<.2$  vagy  $>.8$ . Ez a következővel magyarázható. Ha a valódi prevalencia kicsi, akkor az olyan esetek aránylag nagy részében, amikor a próbaintervallum nem tartalmazza a 0-t, az (5) CI a véletlen hatására jobbra tolódik és nagy valószínűséggel nem tartalmazza a prevalenciát. Ezek a tulajdonságok akkor sem javulnak

meg, amikor  $n_p = 3000$ ,  $n_{Se}$  és  $n_{Sp}$  értéke továbbra is 100, és csak akkor válnak elfogadhatóvá, ha mindhárom minta nagysága eléri a 3000-et. Ha azonban a valódi prevalencia kisebb, mint .03 vagy nagyobb, mint .97, akkor a lefedési valószínűség .90 alá esik még ilyen nagy mintáknál is. Ezen tapasztalatok birtokában nem javasoljuk a Wald-Rogan-Gladen módszert sem előzetes próbaintervallummal, sem nélküle.

Eljárásunk feltétele, hogy a szenzitivitás és specifikusság becslése egymástól független mintákból történjék. Módszerünk keresztmetszeti, teljesen verifikált mintából becsült szenzitivitás és specifikusság (Greiner és Gardner 2000b) mellett is alkalmazható, ilyenkor ui. a szenzitivitást és specifikusságot a valódi pozitív és valódi negatív esetek számára feltételesen független részmintákból becsülik. Ez esetben a (16) formula alkalmazásához újabb független mintára van szükség a célpopulációból. Módszerünk nem kompatibilis az olyan validációs vizsgálatok becslési eljárásaival, ahol a diagnosztikai  $Se$  és  $Sp$  értékét nem egyszerű részarányként határozzák meg. Ilyen eljárás pl. a keresztmetszeti, részben verifikált mintából történő becslés (Greiner és Gardner 2000b), a részben validált adatsorozatokra épülő becslés (Tang et al. 2012), és a „gold standard” teszt nélküli módszerek (Hui és Walter 1980, Enøe, Georgiadis és Johnson 2000, Toft, Jørgensen és Højsgaard 2005).

Módszerünk megköveteli, hogy az első két minta korrekt módon reprezentálja a diagnosztikai teszt szenzitivitását és specifikusságát a célpopuláción. Az utóbbi gyakran különböző rétegekből áll, pl. kor, ivar, betegség súlyossága szerinti csoportokból, területi, gazdasági egységekből. Ha ismert, hogy a célpopuláció heterogén, akkor rétegspecifikus minták és becslések pontosabb eredményekre vezethetnek (Greiner és Gardner 2000b).

A szenzitivitás és a specifikusság varianciáját binomiális eloszlású mintákból becsültük. Bizonyos körülmények között azonban a tényleges variancia nagyobb lehet a binomiális varianciánál. Túlszóródás pl. akkor fordulhat elő, ha a szenzitivitást és a specifikusságot meta-analízissel határozzuk meg. Harbord et al. (2008) összehasonlító elemzést közölt diagnosztikai tesztek pontosságával kapcsolatos meta-analízisek típusairól. Ha korábbi kísérletek eredményeire támaszkodva megadható a prevalencia, szenzitivitás és specifikusság *a priori* eloszlása, akkor Bayes-modellezéssel leírható a túlszóródás, kiszámítható a prevalencia *a posteriori* várható értéke és kredibilis intervalluma. A bayesiánus modellezés azért lehet attraktív, mert az  $AP, Se, Sp$  valószínűségi változók nemlineáris függvényeinek statisztikai bizonytalansága matematikailag egyszerűbben kezelhető. Lesaffre et al. (2007) Bayes-modell keretében tanulmányozta a valódi prevalencia (2) Rogan-Gladen becslését. Arra a következtetésre jutottak, hogy a Bayes-i megközelítés természetes módon illeszkedik az ilyen típusú adatokhoz, de óvatosság szükséges az



informatív *a priori* eloszlások megadásakor, mert a valódi prevalencia *a posteriori* várható értéke és kredibilis intervalluma ezekre érzékeny. Új fejlesztés, eredmény ebben az irányban az R statisztikai szoftver (R Core Team 2014) *rriskBayes* csomagja (Belgorodski et al. 2012). A Bayes-i megközelítés kedvező tulajdonsága, hogy a valódi prevalenciát olyan ismeretlen paraméternek tekinthetjük, aminek *a priori* és *a posteriori* eloszlása szigorúan a  $[0, 1]$  intervallumra korlátozódik (Greiner et al. 2013). Ha azonban a látszólagos prevalencia túl nagy vagy túl kicsi a szenzitivitást és specifikusságot meghatározó minták adataihoz képest, akkor ez hatással van a bayesiánus módszer eredményére is, mert a valódi prevalencia *a posteriori* eloszlása 0 vagy 1 közelében koncentrálódik. Az ilyen elfajult eloszlás analógiát mutat a frekventista megközelítésben a valódi prevalenciára kapott  $[0, 0]$  vagy  $[1, 1]$  formájú, cenzorálás utáni konfidenciaintervallummal. Frekventista modellben a nulla hosszúságú CI jelzi, hogy a látszólagos prevalencia és a diagnosztikai teszt szenzitivitása, specifikussága egymással nem konzisztens. Messam et al. (2008) összefoglalót közölt a diagnosztikai tesztek szenzitivitásáról, specifikusságáról és valódi prevalenciájáról megjelent frekventista és bayesiánus megközelítésekről.

### **3 Szeropozitivitás és prevalencia összefüggései hazai BHV-1 mentesítés adatainak járványtani elemzése alapján**

A magyarországi szarvasmarha-állományok BHV-1 mentesítését gE-negatív markervakcinákkal végzik 2002 óta. A szeropozitivásban 2006-ig bekövetkezett telepi változások alapján az adatok újrafeldolgozásával, a BHV-1 ellenanyagok kimutatása során alkalmazott diagnosztikai tesztek szenzitivitását és specifikusságát felhasználva, a téves pozitív és téves negatív arányok figyelembe vételével a szeropozitivásból megbecsültük a BHV-1 fertőzöttség valódi előfordulási arányát, prevalenciáját. Ennek ismerete lehetővé tette a mentesítés hatásának pontosabb értékelését, a fertőzés korcsoportok, megyék, megyecsoportok és telepek közötti heterogenitásának és terjedési lehetőségének statisztikai vizsgálatát. A fertőzés telepek közötti terjedésének jellemzésére az előbbiek mellett egy ökológiai mutató, a zsúfoltsági index alkalmazását is javasoljuk. A bemutatott statisztikai módszerek más fertőző betegségek elleni mentesítések előrehaladásának nyomon követésére is alkalmasak, és segítségükkel a mentesítési programokban esetlegesen szükséges módosítások idejében meghozhatók.

#### **3.1 Bevezetés**

Az ellenanyagok kimutatására kifejlesztett, jelenleg használatos diagnosztikai tesztek nem elhanyagolható valószínűséggel adnak téves pozitív vagy téves negatív eredményt. A teszt alapján pozitív (szeropozitív) esetek aránya sok esetben lényegesen eltér a fertőzés valódi előfordulási arányától, prevalenciájától. Rogan és Gladen (1978) eljárást dolgozott ki a valódi prevalencia becslésére a szeropozitívok arányából, figyelembe véve a diagnosztikai eljárás érzékenységét és fajlagosságát. A becsléshez egzakt megbízhatósági intervallumokat szerkesztettek Reiczigel et al. (2010). Érzékenység (szenzitivitás) alatt a valóban fertőzött állatok között a pozitív reakciót adó állatok arányát értjük, a fajlagosság (specifikusság) pedig a nem fertőzöttek körében a negatívak arányát jelenti (1. táblázat).

A prevalencia a fertőzés előfordulási aránya egy adott időpontban. Szoros kapcsolatban áll az incidenciával, ami az időegység alatti új fertőzések számarányát jelenti a kockázatnak kitett egyedek körében (Reiczigel 2005, Thrusfield 2005). A két mutató különböző nézőpontokat tükröz: például mentesítési program elindítása után az incidencia hirtelen csökkenését a prevalencia – a korábban fertőződött állatok kiszűréséig – csak lassan követi. Az állategészségügyi felmérő, ill. mentesítési programok során a gyakorlatban kizárólag a prevalenciával számolunk.

## 1. táblázat

A vizsgálat érzékenysége (Se) és fajlagossága (Sp) közötti kapcsolat

A vizsgálat eredménye	Tényleges állapot	
	Fertőzött	Nem fertőzött
<b>Pozitív</b>	(A) Helyes pozitív eredmény	(B) Téves pozitív eredmény
<b>Negatív</b>	(C) Téves negatív eredmény	(D) Helyes negatív eredmény

Teszt érzékenysége (szenzitivitás; Se %) =  $(A / (A + C)) \times 100$

Teszt fajlagossága (specifikusság; Sp %) =  $D / (B + D) \times 100$

A magyarországi szarvasmarha-állományok BHV-1 mentesítését gE-negatív markervakcinákkal végzik 2002 óta (Soós és Tuboly 2009). Pálfi et al. (2007) 155 magyar szarvasmarhatelepen BHV-1 felmérő vizsgálatokat végeztek mentesítés előtt (2002–2004), majd ellenőrző vizsgálatokat 2–3 évvel (2006) a mentesítés elindítását követően. A felmérés során a BHV-1 ellenanyagok kimutatását, túlnyomórészt egyedi vérmintákból, indirekt ELISA vagy vírusneutralizációs (VN), ill. csekély arányban gB-ELISA-próbákkal végezték. Néhány esetben tejminták vizsgálatára indirekt ELISA-tesztet alkalmaztak. Az ellenőrző vizsgálatok során a DIVA stratégiának megfelelő gE tesztet használták, amelyben csak a BHV-1 vírussal fertőzött szarvasmarhák reagálnak pozitívan, a marker vakcinával oltottak nem (Soós és Tuboly 2009).

Jelen fejezetben a BHV-1 mentesítéshez kapcsolódó felmérő és ellenőrző vizsgálatok eredményeit, adatait újból feldolgoztuk és elemeztük. A szeropozitív esetek arányából Rogan és Gladen (1978) módszerével becsüljük a BHV-1 valódi prevalenciáját. A fertőzöttség mentesítés előtti és azt követően tapasztalt valódi prevalenciája pontosabban mutatja a mentesítés hatékonyságát, a fertőzöttség korcsoportok, megyék, megyecsoportok és telepek közötti különbségeit, mint a téves pozitív és téves negatív eseteket is magában foglaló szeropozitivitás. A prevalencia ismeretében eldönthető, hogy a fertőzöttség függ-e az állománymérettől, és ha igen, mi módon. A fertőzés telepek közötti terjedési feltételeinek vizsgálatához a felsorolt értékek mellett egy ökológiai mutatószám, a zsúfoltsági index alkalmazását is javasoljuk (Lloyd 1967, Reiczigel et al. 2005, 2008).

## 3.2 Anyag és módszer

A 2002–2006 közötti vizsgálatok során a minták összesen 155 telepről származtak, de a felmérő és ellenőrző vizsgálatokban csak részben ugyanazok a telepek szerepeltek (Pálfi et al. 2007). Az adatok megismételt elemzéséhez azokat a teszteredményeket használtuk, ahol

a teszt típusa és a kórisme minden mintára megállapítható. A tejmintákon alapuló tesztek kizártuk, ugyanis ezek szenzitivitására és specifikusságára nem találtunk megbízható, elég nagyszámú mintából nyert, közölt értéket. Kizártuk azokat az ellenőrző vizsgálatokat is, amelyeket nem gE ELISA teszttel végeztek. *A kizárások után összesen 143 telep adatait dolgoztuk fel.* A létszámmal kapcsolatos statisztikai elemzések 115 állományra épülnek, mivel ezekről kaptunk megbízható létszám adatokat. Az adatok telepenként korcsoportos bontásban álltak rendelkezésre, a korcsoportok: 0–6 hónapos borjak, 6–18 hónapos növendék üszők, 18–26 hónapos vemhes üszők és tehenek >24 hónapos korban. A diagnosztikai tesztek szenzitivitása és specifikussága egy nyugat-európai, több nemzetet érintő, átfogó kutatás során került meghatározásra (Kramps et al. 2004).

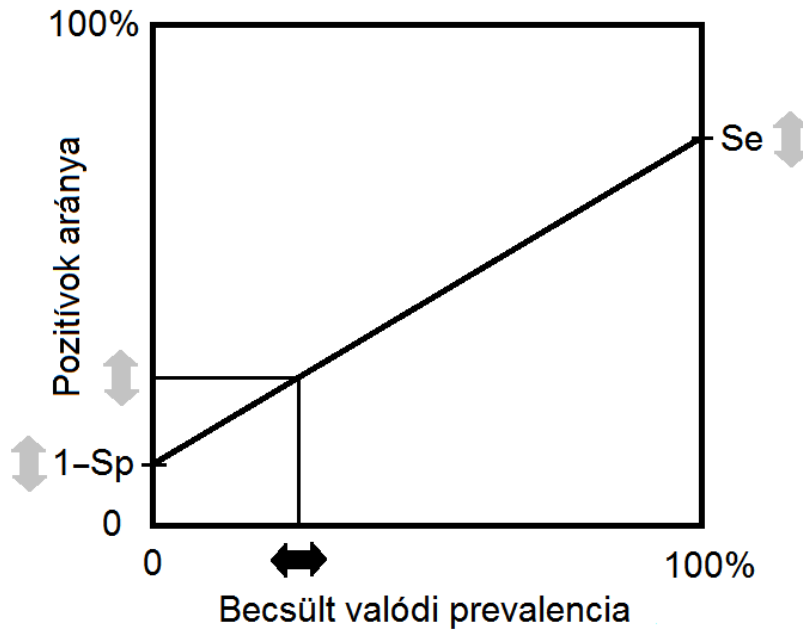
### 3.2.1 Statisztikai módszerek

A pozitív esetek (*poz*) arányából, a diagnosztikai teszt szenzitivitásából (*Se*) és specifikusságából (*Sp*) a valódi prevalenciát (*prev*) a

$$prev = (poz + Sp - 1) / (Se + Sp - 1)$$

képlettel becsültük (Rogan és Gladen 1978). Ha ez negatív számot eredményez, akkor 0, ha 1-nél nagyobb, akkor 1 a prevalencia becslése. A szeropozitivitás és a valódi prevalencia kapcsolatát mutatja az 1. ábra. A valódi prevalencia 95%-os kétoldali megbízhatósági tartományát (konfidencia intervallum – CI) a Blaker-féle egzakt próbából Reiczigel módszerével számítottuk ki (Blaker 2000, Reiczigel et al. 2010). Az egzakt próba alkalmazhatósági feltétele az, hogy az összes pozitív minta száma binomiális eloszlást kövessen. Ez a feltétel teljesül, ha a telepenkénti minták száma polinomiális eloszlású és a pozitív minták száma telepenként binomiális eloszlású (Vágó et al. 2011).

A prevalenciákat korcsoportonként, megyénként, megyecsoportonként és telepenként is meghatároztuk. A mentesítés előtti és utáni prevalenciák szignifikáns eltérését korcsoportonként egyoldali 97,5%-os megbízhatóságú Clopper-Pearson CI-vel teszteltük (Clopper és Pearson 1934, Reiczigel et al. 2010).



1. ábra

*A szeropozitivitás és a valódi prevalencia kapcsolata*

Korcsopontonként megvizsgáltuk, mekkora szórással ingadozik a prevalencia megyék között és megyéken belül. Ehhez a korcsoportos prevalencia teljes variációját megyék közötti és megyén belüli variációk összegére bontottuk, a szórások az átlagos variációk négyzetgyökei (Reiczigel, Harnos, Solymosi 2007). A szórások becslésénél figyelembe kell venni, hogy a telepek prevalenciái a kis mintaszámok miatt pontatlanul becsülhetők. Az ebből a pontatlanságból adódó torzítás azonban korrigálható (Lang et al. 2013, [www.univet.hu/users/zslang/korrekcioik.pdf](http://www.univet.hu/users/zslang/korrekcioik.pdf)).

Meghatároztuk az átlagos prevalenciát és a prevalencia csoporton belüli szórását három földrajzilag összefüggő, állomány nagyság szerint többé-kevésbé homogén megyecsoportra (a Dunántúl Komárom-Esztergom megye nélkül, az Alföld, valamint az északi megyék: Komárom-Esztergom, Pest, Nógrád, Heves, Borsod-Abaúj-Zemplén).

Az ismert létszámú 115 állományban megvizsgáltuk a létszám és a prevalencia kapcsolatát. Nagyobb létszámú telepen több az állatok közötti kontaktus, ezért a fertőzés elvileg gyorsabban terjedhet, és nagyobb szintet érhet el, tehát a prevalencia nagyobb lehet. E hipotézis vizsgálatához a telepeket állatlétszám szerint növekvő sorrendbe rendeztük, majd öt, egyenként 23–23 telepből álló csoportba, kvintilisbe osztottuk (egyedszámok 1–200, 201–300, 301–454, 455–620, 621–3000). Mindegyikhez kiszámítottuk felméréskor és ellenőrzéskor a BHV-1 prevalenciát. A statisztikai összehasonlításához szimultán CI-eket szerkesztettünk a prevalenciákra Bonferroni módszerével: öt kvintiliszhez összesen 10

páronkénti összehasonlítás lehetséges, ezért 99,75%-os nem szimultán CI-re van szükség 5%-os szintű szimultán elsőfajú hiba eléréséhez (Reiczigel, Harnos, Solymosi 2007).

Az állatszám telepek közötti nagy szórása miatt az átlagos létszám nem mutatja elég pontosan az állatok között lehetséges kapcsolatok, érintkezések számát, ami a fertőzés elterjedésének, maximális mértékének, a prevalenciának jellemzője. Egy állat (az eladás ritka esetét leszámítva) kizárólag az azonos telepen élő többi egyeddel érintkezhet, ezek száma a telep létszáma mínusz egy. Ha ezt az egyedekre átlagoljuk, akkor a lehetséges érintkezésszám átlagát kapjuk, amit *zsúfoltsági index*nek nevezünk (Lloyd 1967, Reiczigel et al. 2005, 2008). Ha például három telepen az állatszám 100, 200 és 1000, akkor itt a zsúfoltsági index  $(100 \times 99 + 200 \times 199 + 1000 \times 999)/(100 + 200 + 1000) = 807$ . Ha az állatok zöme néhány nagyobb telepen csoportosul, miközben a kisebb telepekhez ezek számarányához képest kevés egyed tartozik, akkor a zsúfoltsági index jóval meghaladja az átlagos létszámot (a példában az átlagos egyedszám 433).

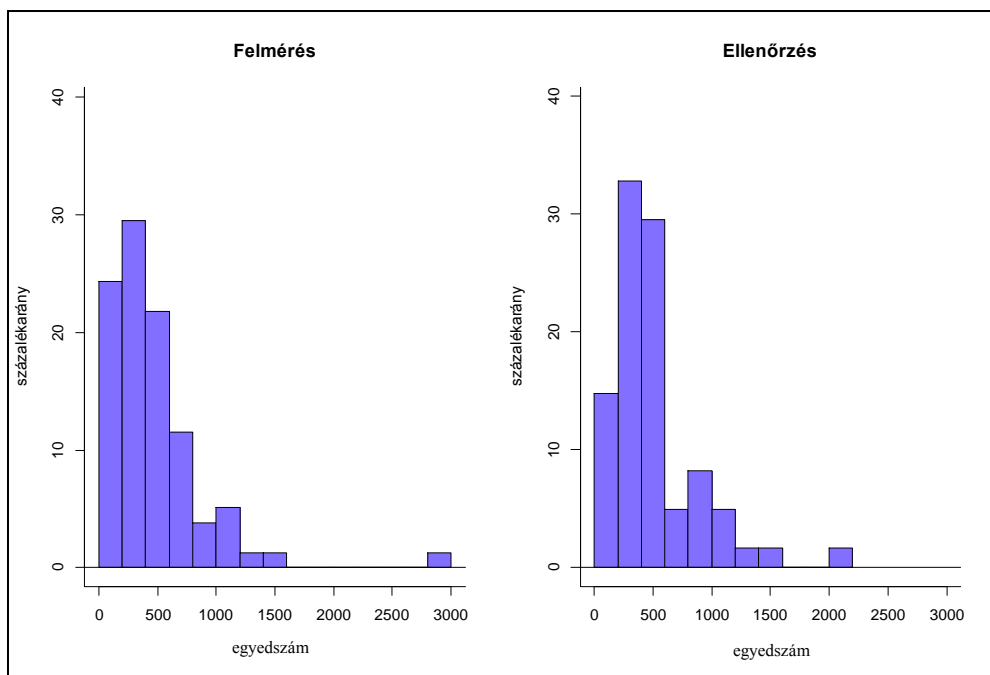
A fertőzés terjedését jellemzi a reprodukciós arányszám ( $R_0$ ), ami egy fertőzőképes (azaz kórokozó ürítő) egyed által megfertőzött állatok számát jelenti. A reprodukciós arányszám többek között függ az állatok érintkezési számától, a fertőzőképes egyedek prevalenciájától, a fertőzőképesség átlagos időtartamától, és az egészséges–fertőző egyedek érintkezése során a fertőzés átadásának valószínűségétől (Heesterbeek 2002). A zsúfoltsági index a reprodukciós arányszám egyik tényezője.

### 3.3 Eredmények

Az elemzésbe vont 143 telep közül 73-ból (51%) csak felmérő, 40-ből (28%) csak ellenőrző, 30-ból (21%) mindkét vizsgálatra küldtek mintákat. Az állatok száma ismert 115 telepre, ezek közül 54-ben volt felmérés (47%), 37-ben ellenőrzés (32%) és 24-ben mindkettő (21%). Felméréskor a létszám átlaga 485, szórása 433. Ellenőrzéskor az átlag 523, a szórás 395. A létszámok eloszlása jobbra ferde, a hisztogramok alakja felméréskor és ellenőrzéskor nagyon hasonló (2. ábra).

#### ***Szeropozitivitás és valódi prevalencia***

A valódi prevalencia minden korcsoportban és összességében is kisebb volt a mentesítés utáni ellenőrzéskor a mentesítést megelőző felméréshez képest. A valódi prevalenciák felméréskor nagyobbak, ellenőrzéskor pedig kisebbek, mint a szeropozitivitások (2. táblázat). Tehát a vizsgált periódusban (2002 és 2006 között) a mentesítés valójában eredményesebb volt, előbbre jutott, mint azt a szeropozitivitási arányok mutatják.



**2. ábra**

*Egyedszámok hisztogramja felméréskor és ellenőrzéskor*

**2. táblázat**

*Szeropozitivitás és valódi prevalencia*

Csoport	Minták száma		Szeropozitivitás %		Valódi prevalencia (Blaker 95% CI)	
	Felmérés	Ellenőrzés	Felmérés	Ellenőrzés	Felmérés	Ellenőrzés
Borjak	1093	142	48	15	54 (51,58)	11 (3,21)
Növendék üszők	1562	778	19	6	20 (18,23)	0 (0,0)
Vemhes üszők	715	549	45	9	53 (48,57)	1 (0,5)
Tehenek	2408	839	79	17	91 (89,93)	15 (11,19)
<b>Összesen</b>	<b>5778</b>	<b>2308</b>	<b>53</b>	<b>11</b>	<b>61 (59,62)</b>	<b>5 (3,7)</b>

A felmérések prevalenciáihoz 97.5%-os szintű egyoldali alsó, az ellenőrzésekhez egyoldali felső Clopper-Pearson-féle megbízhatósági határokat számítottunk. Ha ez utóbbi kisebb, mint az előbbi (más szóval a CI-k nincsenek átfedésben), akkor ez 5%-os hibavalószínűség mellett bizonyítja, hogy ellenőrzéskor a prevalencia kisebb volt, mint a felméréskor. A 3. táblázatból látható, hogy minden csoportban és összesítve is 5%-os szinten szignifikánsan alacsonyabb a prevalencia ellenőrzéskor a felméréshez képest.

**Területi különbségek**

A megyéket külön-külön vizsgálva nagy eltérések tapasztalhatók mind a felméréskori, mind pedig a mentesítés utáni prevalenciák vonatkozásában (4. táblázat).

### 3. táblázat

Prevalenciák eltérése felméréskor és ellenőrzéskor

Csoport	Prevalencia felméréskor (2002)	Clopper- Pearson alsó határ	Prevalencia ellenőrzéskor (2006)	Clopper- Pearson felső határ
Borjak	54	51	11	21
Növendék üszők	20	18	0	0
Vemhes üszők	53	48	1	5
Tehenek	91	89	15	19
Összesen	61	59	5	7

### 4. táblázat

Megyéenkénti prevalenciák

Megye	Minták száma		Valódi prevalencia (Blaker 95% CI)	
	Felmérés	Ellenőrzés	Felmérés	Ellenőrzés
Bács-Kiskun	267	10	65 (57,72)	0 (0,32)
Baranya	444	20	45 (40,50)	0 (0,13)
Borsod-Abaúj-Zemplén	247	236	62 (54,69)	0 (0,0)
Békés	104	79	40 (28,53)	0 (0,4)
Csongrád	70	122	100 (100,100)	13 (4,25)
Fejér	310	374	82 (76,88)	13 (7,19)
Győr	360	268	67 (61,73)	15 (9,23)
Hajdú-Bihar	1061	205	76 (72,79)	0 (0,0)
Heves	20	-	18 (4,43)	-
Jász-Nagykun-Szolnok	631	217	70 (65,74)	0 (0,0)
Komárom-Esztergom	494	62	18 (13,23)	8 (0,25)
Pest	198	174	69 (61,78)	18 (10,28)
Somogy	351	101	59 (53,65)	39 (25,54)
Szabolcs-Szatmár- Bereg	270	46	35 (29,42)	0 (0,0)
Tolna	453	193	54 (48,60)	0 (0,3)
Vas	58	-	97 (81,100)	-
Veszprém	440	201	61 (55,67)	7 (0,15)
Összesen	5778	2308	61 (59,62)	5 (3,7)

A legtöbb megyében jelentős a csökkenés, egyedül Somogy megyében maradt kirívóan nagy a fertőzöttség. Tendenciát nehéz megállapítani, talán annyi igaz, hogy ahol nagyobb a felméréskor a fertőzöttség, ott a mentesítés után is általában nagyobb marad (a Spearman-féle korreláció pozitív, értéke 0.28, de nem szignifikáns,  $p=0.3100$ .)

A prevalenciák megyék közötti és megyéken belüli ingadozását vizsgálva megállapítható, hogy a telepek között, megyétől függetlenül, nagy a heterogenitás mind a felmérő, mind az ellenőrző vizsgálatok során, más szóval a BHV-1 fertőzöttség önmagában is, ill. a mentesítés előrehaladása is elsősorban teleptől és nem megyétől függ.

A földrajzilag összefüggő homogén megyecsoportokba sorolt megyékre a prevalencia átlagát és csoporton belüli szórását az 5. táblázat mutatja. Felméréskor nincs lényeges különbség a



csoportok átlaga és szórása között. Mentés után a Dunántúlon és az északi megyékben valamivel jobban sikerült csökkenteni a fertőzés prevalenciáját, mint az Alföldön. A heterogenitás is csökkent, leginkább az északi megyékben. (Megjegyzendő, hogy ellenőrzéskor a Dunántúlon az átlagos valódi prevalencia nyers torzítatlan becslésére negatív érték adódott, ezt korrigáltuk 0-ra.)

## 5. táblázat

*A valódi prevalencia átlaga és szórása megyecsoportonként*

Megyecsoport	Felmérés (2002)			Ellenőrzés (2006)		
	Telepszám	Átlag	Szórás	Telepszám	Átlag	Szórás
Dunántúl	39	60	42	24	0	29
Alföld	52	58	37	34	8	32
Északi megyék	13	54	39	12	3	15

### ***Az állománylétszám és a prevalencia kapcsolata***

Az állománylétszám és a prevalencia kapcsolatát létszám szerint kvintilisbe sorolt telepescsoportokkal vizsgáltuk meg azon a 115 telepen, ahol ismertük a létszámot. Az eredmények összefoglalása a 6. táblázatban található.

Azt tapasztaljuk, hogy felméréskor a nagyobb létszámú telepek csoportjához növekvő tendenciájú BHV-1 prevalencia tartozik. A szimultán konfidencia intervallumok az első három és az utolsó két létszámcsoport között nem nyúlnak át egymásba, ezért 5%-os szimultán, az összes kvintilisre egyszerre vonatkozó tévedési valószínűség mellett a kb. 450 egyedet meghaladó méretű telepeken nagyobb a prevalencia, mint a kisebb egyedszámú telepeken.

A mentés utáni ellenőrzésnél fordított a trend, a nagy létszámú telepeken kisebb a prevalencia. Az első négy kvintilis között a konfidencia intervallumok átfedők, a csökkenés nem szignifikáns. Csak az ötödik, legnagyobb létszámú kvintilis prevalenciája különbözik szignifikánsan a többitől. A csökkenő trend bizonyítékát gyengíti, hogy az első, legkisebb állatszámú kvintilisben az átlagos 38%-nál jóval nagyobb, 60%-os a felnőtt tehenekből vett minták aránya, a negyedik kvintilisben viszont ez kisebb, 24%-os. Az ötödik kvintilisben azonban nincs lényeges különbség a többihez képest a korcsoportonkénti mintaszámok arányában. Csak ennél a kvintilisenél mutattunk ki szignifikáns csökkenést, ezt a megállapítást nem kérdőjelezi meg a mintaszám arányok eltérése.

## 6. táblázat

*A telep létszáma és a prevalencia kapcsolata*

Létszám- kvintilis	Minták száma		Valódi prevalencia (Blaker-féle 95%- os szimultán konfidencia intervallum)	
	Felmérés	Ellenőrzés	Felmérés	Ellenőrzés
200-ig	735	269	46 (39,52)	17 (7,30)
201–300	955	234	34 (28,40)	18 (6,31)
301–454	618	396	54 (47,61)	10 (3,20)
455–620	1185	671	77 (72,82)	7 (1,13)
621–3000	958	487	70 (64,75)	0 (0,0)

Korcsoportonként vizsgálva, felméréskor a borjak és tehének esetében figyelhető meg a telep állatszám és a prevalencia pozitív kapcsolata. A növendék és vemhes üszőknél nem vehető észre kapcsolat (7. táblázat). Ellenőrzéskor nincs korcsoportonként kimutatható kapcsolat a telepek állatszám és a prevalencia között.

## 7. táblázat

*Felméréskor a telepi állomány létszáma és a prevalencia kapcsolata korcsoportonként*

Létszám- csoport	Borjak	Növendék üszők	Vemhes üszők	Tehenek
200-ig	35	12	67	81
201–300	45	19	55	66
301–454	37	12	31	97
455–620	74	49	42	92
621–3000	76	16	49	94

## Zsúfoltság

A zsúfoltsági index felméréskor 867, ellenőrzéskor 817. Az átlagosan lehetséges kontaktusszám 50–80%-kal nagyobb, mint a telepek átlagos létszáma, a nagyobb létszámú telepeken található ugyanis az egyedek nagyobb hányada.

## 3.4 Megvitatás

### *Szeropozitivitás és valódi prevalencia*

A valódi prevalenciára felméréskor nagyobb, mentesítéskor kisebb értéket kaptunk, mint a szeropozitivitás aránya. Ennek a paradox jelenségnek elsősorban az a magyarázata, hogy a mentesítés jelentősen csökkenti a szeropozitivitást, és a Rogan-Gladen formulából kiolvasható, hogy nagy szeropozitivitásnál a valódi prevalencia még nagyobb, kis szeropozitivitásnál még kisebb lesz (1. ábra). A prevalenciák és szeropozitivitások különbségét tovább növeli, hogy mentesítés előtt és után eltérő hatékonyságú diagnosztikai tesztekkel használtak. A felméréskor használt indirekt ELISA- és VN-próbák specifikussága

97–98%-os, ezért a téves pozitívok legfeljebb 2–3%-kal emelik a szeropozitivitást. A szenzitivitás kisebb, 82–86%-os, ami a téves negatívok nagy arányával jár együtt. A számos téves negatív eset miatt a valódi prevalencia jóval nagyobb a szeropozitivitáshoz képest. Az ellenőrzéskor használt gE-ELISA teszt specifikussága kicsi, 92%-os, ami viszonylag több téves pozitív eredményt ad. A teszt szenzitivitása nagyon kicsi, 72%-os, de ez a kisebb szeropozitivitás miatt nem tudja ellensúlyozni a téves pozitívok hatását.

Az ellenőrzésben nagyobb súllyal voltak jelen a vemhes üszők és növendékek, valamint a teheneknél többségében egyszer ellett, azaz fiatal állatokból vették a mintát. A korcsoportonként kapott eltérések azonban teljes összhangban állnak az összesítések eltéréseivel, így a korcsoportos minták felméréskor és ellenőrzéskor különböző aránya nem torzítja el az összképet.

### ***Milyen diagnosztikai tesztet érdemes választani?***

Mentesítés előtt, amikor a fertőzés aránya feltehetően nagy, elsősorban a téves negatív minták ronthatják el a valódi prevalencia becslésének pontosságát (ha a szenzitivitás 100%, akkor téves negatív eredmény nem fordulhat elő). Ezért mentesítés előtt fontos, hogy a választott teszt szenzitivitása minél jobb legyen. Nagy fertőzöttségi arány esetén a specifikusság kevésbé befolyásolja a prevalencia becslésének pontosságát, továbbá a Rogan-Gladen formula a kisebb specifikusság felfelé torzító hatását megszünteti. Sikeres mentesítés után, kis prevalencia mellett a téves pozitív minták növelhetik a prevalencia becslésének standard hibáját (ha a teszt specifikussága 100%, akkor téves pozitív eredmény nem lehetséges). Mentés után ezért minél jobb specifikusságú tesztet érdemes választani, a szenzitivitásnak ekkor kisebb a jelentősége. A Rogan-Gladen korrekció miatt a kisebb szenzitivitás nem torzítja lefelé a prevalencia becslését. Ez a megállapítás látszólag illogikus, hiszen a mentesítés befejezése után az ellenőrző vizsgálatokkal az a cél, hogy az esetleges néhány fertőzött egyedet minél hamarabb és minél nagyobb hatékonysággal találjuk meg, ezért a teszt szenzitivitását tekintenénk elsődlegesnek. Az elmondottakból azonban nyilvánvaló, hogy egy gyengébb specifikusságú teszttel megtalált néhány pozitív állat nagyobb eséllyel lehet téves pozitív, mint valóban fertőzött. A teszt típusának kiválasztásakor ügyelni kell arra is, hogy a szenzitivitás és specifikusság becslése minél nagyobb, lehetőleg legalább néhány száz (200–400) mintából történjen, a minták minél heterogénebb állatpopulációból származzanak és több laboratórium azonos mintákon végzett vizsgálatán („körvizsgálat”) alapuljanak. Ekkor ugyanis kizárható a laboratóriumok közötti eltérésből, ill. az állatfajon belüli eltérő csoportokból (pl. fajta, hibrid) adódó torzítás. A

tejvizsgálatok eredményét éppen azért zártuk ki az elemzésből, mert kis mintaszámmal becsülték meg a szenzitivitásukat és a specifikusságukat (Kramps et al. 2004).

Az eredményeket annak feltételezésével kaptuk, hogy a diagnosztikai vizsgálatok szenzitivitása és specifikussága pontosan ismert és a magyar szarvasmarha-populációra is alkalmazható. A szenzitivitás és specifikusság azonban tartalmazhat kismértékű véletlenszerű hibát, ugyanis ezek közepes, tesztípusonként más és más, 100–400-as mintaszámból becsült értékek (Kramps et al. 2004). A szenzitivitás és specifikusság lehetséges véletlen ingadozásának nagyságrendjét és hatását a (Lang és Reiczigel 2014) közlemény módszerével megbecsülve érzékenységvizsgálatot végeztünk és azt tapasztaltuk, hogy a véletlen okozta eltérések nem vezetnek a következtetéseket, eredményeket illetően lényeges változáshoz.

### ***Területi és telepek közötti különbségek***

A mentesítés hatására a BHV-1 vírus prevalenciája a telepek átlagát tekintve kismértékűre csökkent, nagyok maradtak ugyanakkor a telepek közötti különbségek a vizsgált időszakban. A hatékonyan mentesített telepek szomszédságában gyakran sikertelenül vagy egyáltalán nem mentesített telepek vannak, ami nyilvánvalóan megkönnyítheti a fertőzés terjedését.

A területi és telepek közötti prevalencia különbségeinek mérése sokat segíthet a mentesítéshez szükséges idő előrejelzésében. A heterogenitást számszerűsítő szórásokkal vizsgálhatjuk meg, hogy mindenhol következetesen végrehajtsák-e a mentesítést.

### ***Zsúfoltság***

A magas zsúfoltsági index, azaz nagy átlagos kontaktusszám az átlagos létszám nagyságán kívül az állományok méretének változatosságából is adódik. Értéke kizárólag a telepek állatszámaitól, a létszámmintázattól függ, időben csak lassan változik és a fertőzöttség nem befolyásolja. A nagy zsúfoltsági index a nem egyformán hatékony mentesítés miatt heterogén prevalenciával együtt visszafertőzést okozhat. Magyarországon a vizsgált időszakban a zsúfoltsági index nagy, emellett a mentesítés sikere megyéken, megyecsoportokon belül nagyon heterogén (4. táblázat), főleg a kisebb létszámú telepeken gyengébb (6. táblázat). Ezért számottevő annak a kockázata, hogy egy-egy nagyobb telep mentesített állományát a közelében lévő kisebb telepek mentesítetlen állománya visszafertőzi.

Mentesítés előtt a nagyobb létszámú telepeken nagyobb a BHV-1 prevalencia (6. táblázat), ezért a nagy zsúfoltsági index, kontaktusszám a BHV-1 fertőzés prevalenciáját mentesítés nélkül növelheti.

### 3.5 Következtetések

*A valódi prevalencia a fertőzés tényleges arányát mutatja.* A diagnosztikai vizsgáló módszer szenzitivitása és specifikussága ismeretében mindig megbecsülhető a szeropozitivitás arányából. A szeropozitív minták léte önmagában nem bizonyítja, hogy az állományban jelen van a fertőzés, hiszen a tesztek nem 100%-osak, ezért egyes minták lehetnek téves pozitívak. A fertőzés tényét úgy lehet kimutatni, hogy kiszámítjuk a valódi prevalencia 95%-os megbízhatóságú egyoldali alsó konfidenciahatárát (a 90%-os megbízhatóságú kétoldali Clopper-Pearson konfidencia intervallum alsó határa). Ha ez pozitív, akkor 5%-os tévedési valószínűség mellett megállapítható, hogy a valódi prevalencia pozitív, azaz a fertőzés jelen van a vizsgált állatpopulációban. Megfelelő mintaszám mellett a prevalencia előre adott pontossággal megbecsülhető. A valódi prevalencia gyakorlati alkalmazása így mindenképpen ajánlott, ha a vizsgáló módszer szenzitivitására és specifikusságára megbízható, közölt értékek ismeretesek.

Ha a diagnosztikai módszer szenzitivitása és specifikussága nem ismert vagy csak kis mintaszámmal becsülték, akkor inkább válasszunk másik módszert. *A szeropozitivitás önmagában nem ad teljes körű és megbízható tájékoztatást a fertőzés arányáról.*

A prevalencia elfogadható pontosságú becsléséhez viszonylag nagy, legalább 100-as mintaszámra van szükség. Telepenként átlagosan 46 mintát vizsgáltak meg, ekkora mintaszámmal egy-egy telepen csak a fertőzés létének vagy adott szint alatt maradásának kimutatására van esély, a konfidencia intervallum túl tág a prevalencia nagyságának pontos megállapításához. (A jelenleginél pontosabb diagnosztikai módszerek kifejlesztésével a helyzet a jövőben változhat.) A mentesítés előrehaladását telepenként úgy ellenőrizhetjük, hogy időről időre megnézzük, a prevalencia 95%-os megbízhatóságú egyoldali felső konfidenciahatára adott szint alá süllyedt-e.

*A bemutatott módszereket a leghatékonyabban országos vagy legalábbis több telepet felölelő vizsgálati program keretében lehet alkalmazni.* Ilyenkor lehetőség van pontos becslésekre, korcsoportonkénti, területi, időbeli eltérések összehasonlítására. A valódi prevalencia Rogan-Gladen becslését és konfidencia intervallumát kiszámító program, valamint on-line kalkulátor elérhető a <http://www.univet.hu/users/jreiczig/prevalence-with-se-sp.html> weboldalon.

## 4 Zsúfoltság és diverzitás

Gazda-parazita közösségben a zsúfoltságot a gazdákon élő paraziták száma valamilyen monoton növekvő függvényének parazitákra vett átlagával értelmezzük. Lloyd (1967) meghatározásában ez a függvény a paraziták száma, eggyel csökkentve. Reiczigel et al. (2005) a paraziták számát átlagolják a parazitákra az általuk bevezetett lineáris zsúfoltság esetében, így Lloyd zsúfoltság-fogalmához képest 1-gyel nagyobb értékhez jutnak. A lineáris zsúfoltság mellett Reiczigel et al. (2005) értelmezték a logaritmikus zsúfoltságot, ez a gazdákon élő paraziták számának természetes alapú logaritmus, a parazitákra átlagolva (a parazitákra vett átlag miatt automatikusan kimaradnak azok a gazdák, amelyekben nincs parazita, így a logaritmust mindig csak pozitív számokra kell kiszámítani). A zsúfoltságot általánosan, tetszőleges monoton növekvő valós értékű függvényre Reiczigel et al. (2005) vezették be. A zsúfoltság fogalmát Jarman (1974) és Reiczigel et al. (2005, 2008) általánosította csoportokban élő populációk szubjektív, a csoporttagok, egyedek szemszögéből vett csoportméretének meghatározására, az egyedeken átlagolva a csoportok létszámának valamely monoton növekvő függvényét.

Ha a csoportok fajokat jelentenek egy életközösségen belül, akkor az átlagos zsúfoltság vagy csoportméret helyett a fajok átlagos ritkaságával értelmezett diverzitási index használata terjedt el. Patil és Taillie (1982) egy faj ritkaságát olyan  $R(p)$  függvénnyel méri, ami a faj közösségen belüli  $p$  részarányának monoton csökkenő függvénye és  $R(1) = 0$ . A közösség  $D$  fajdiverzitását az említett szerzők az  $R$  ritkaságnak az egyedszámokkal súlyozott átlagaként definiálják. Vegyük észre, hogy ez nem más, mint a ritkaság egyedek szerinti egyszerű aritmetikai átlaga. A Patil-Taillie diverzitási indexek családjába az  $R$  ritkasági függvény alkalmas megválasztásával speciális esetként sok, a gyakorlatban elterjedt diverzitási indexet foglal magában, pl. a fajsámot, a Shannon indexet, a Simpson indexet, a Berger-Parker indexet, továbbá számos más, a Rényi-féle általánosított entrópia fogalmára épülő indexet (Shannon és Weaver 1963, Simpson 1949, Berger és Parker 1970, Rényi 1961, Patil és Taillie 1982, Tóthmérész 1995, 1997).

### 4.1 Kapcsolat a zsúfoltság és a diverzitás fogalma között

A diverzitás, mint az egyedek átlagos ritkasága nagyon hasonlít a zsúfoltság értelmezésére. Valóban, bármely rögzített  $c$  konstans mellett a  $c - R(p)$  függvény a faj egyedszámának monoton növekvő függvénye és az ezzel kiszámolt zsúfoltság éppen  $c - D$ .

A továbbiakban csoportok valamely közösségét vizsgáljuk. Pontosan definiáljuk a csoportok zsúfoltságát vagy más néven szubjektív méretét és diverzitását a közösségben. Az  $i$ -ik

csoport tagjainak (egyedeinek) száma legyen  $X_i, i = 1, \dots, N$ , ahol  $N$  a csoportok száma. Legyen  $f$  monoton növekvő valószínűségi függvény, ami a nemnegatív valószínűségeken van értelmezve. Zsúfoltságon (vagy szubjektív csoportméreten) a

$$C_f = \frac{\sum_{i=1}^N X_i f(X_i)}{X}$$

átlagot értjük, ahol  $X = \sum_{i=1}^N X_i$ . Az egyedekhez társuló vagy társítható szubjektív, egyedi zsúfoltságot számszerűsíti az  $f(x)$  függvény, ahol  $x$  az egyed csoportjának egyedszáma. A szubjektív zsúfoltságon egyedekre vett átlaga adja a közösség átlagos zsúfoltságát. Ennek képletét kicsit más formában felírva,  $C_f = \sum_{i=1}^N p_i \cdot f(X \cdot p_i)$ , ahol  $p_i = X_i/X$  az egyedek részaránya az  $i$ -ik csoportban. Ebből a felírásból látszik, hogy a zsúfoltság denzitásfüggő, hiszen függ a közösség teljes  $X$  létszámától. Másrészt, ha az  $X$  teljes létszámot rögzítjük, akkor a zsúfoltság csupán a  $p_i$  csoport-részarányoktól függ. Ez lehetőséget ad arra, hogy rögzített  $X$  összlétszám mellett a zsúfoltsági indexhez diverzitási indexet társítsunk. Ehhez vezessük be az

$$R(p) = f(X) - f(p \cdot X)$$

függvényt. Ez rögzített  $X$  összlétszám esetén ritkasági függvény lesz a Patil-Taillie (1982) által adott definíció értelmében. Az  $R$  függvény ugyanis monoton csökkenő, nemnegatív és  $R(1) = 0$ . Az  $R$  függvényhez tartozó diverzitási index az átlagos ritkaság,

$$D_R = \sum_{i=1}^N p_i \cdot R(p_i) = f(X) - C_f.$$

Összességében azt kaptuk eredményül, hogy a teljes egyedszám rögzítése mellett a csoportok diverzitási és zsúfoltsági indexe egyszerű formulákkal egymásba átírható, információtartalmuk megegyezik.

A diverzitási indexek elméletéről és alkalmazási lehetőségeiről terjedelmes szakirodalom áll rendelkezésre, az ezekben szereplő ismeretek, fogalmak mind átfogalmazhatók zsúfoltsági indexekre. Néhány összefoglaló mű, messze a teljesség igénye nélkül: Jost 2006, 2007, Tuomisto 2010a, 2010b, Veech 2002, Lande 1996, Tóthmérész 1997. A zsúfoltsági indexről kevesebb publikáció és közölt alkalmazás érhető el (pl. Lloyd 1967, Jarman 1974, Reiczigel et al. 2005, 2008, Neuhauser et al. 2010), ennek ellenére a zsúfoltság fogalmának diverzitási indexszel való rokonítása gyümölcsöző lehet a diverzitási indexek elmélete szempontjából is.

Egy fajközösség effektív fajszáma (numbers equivalent) azt a hipotetikus fajszámot jelenti, amihez fajoként egyenlő egyedszámú közösségben ugyanakkora diverzitási index tartozik, mint a vizsgált fajközösségben (MacArthur 1965, Hill 1973, Lande 1996, Tóthmérész 1997, Jost 2006, 2007, Tuomisto 2010a). Ennek analógiájaként gazda-parazita kapcsolatban a zsúfoltsághoz tartozó effektív gazdaszám legyen a gazdák száma egy olyan elképzelt

gazda-parazita közösségben, ahol a paraziták egyenletesen oszlanak meg a gazdák között, továbbá a zsúfoltság és az összesített abundancia megegyezik a vizsgált gazda-parazita közösség zsúfoltságával és összes parazitáinak számával. Az effektív gazdaszám fogalmának bevezetésekor, definiálásakor szükség van az összes parazitaegyed számának rögzítésére. Ennek oka az, hogy a zsúfoltság (átlagos csoportméret) denzitásfüggő, a diverzitás pedig denzitásfüggetlen.

Általánosan, valamely csoportokból álló közösség effektív csoportszámát értelmezzük a csoportok számaként egy olyan elképzelt közösségben, ahol minden csoportnak ugyanannyi tagja van, emellett a tagok összesített száma és a csoportméret megegyezik a vizsgált közösség tagjainak számával, ill. átlagos csoportméretével.

A definíció alapján az  $M_f$  effektív gazdaszámra/csoportméretre az alábbi egyenlet írható fel:

$$C_f = \frac{\sum_{i=1}^N X_i \cdot f(X_i)}{X} = \frac{\sum_{i=1}^{M_f} X./M_f \cdot f(X./M_f)}{X} = f\left(\frac{X.}{M_f}\right) = f\left(\frac{N}{M_f} \cdot \bar{X}\right),$$

ahol  $\bar{X}$  a csoportonkénti átlagos egyedszám vagy gazdánként a paraziták átlagos abundanciája. Az egyenletből az effektív gazdaszám/ csoportszám az

$$M_f = \frac{\bar{X}}{f^{-1}(C_f)} \cdot N$$

képlettel fejezhető ki. Eszerint az effektív csoportszám a csoportonkénti átlagos létszám és a létszámra visszaskálázott csoportméret hányadosa szorozva a csoportok számával. Gazda-parazita közösségben az effektív gazdaszám a gazdánkénti átlagos parazitagyakoriság és a parazitaszámra skálázott zsúfoltság hányadosa szorozva a gazdák számával. Egyszerű számítás mutatja, hogy ez megegyezik az  $R(p) = f(X) - f(p \cdot X)$  ritkasági függvény diverzitási indexéhez tartozó effektív fajszám (itt: csoportszám, gazdaszám) értékével. Figyeljük meg, hogy – a diverzitási indexek effektív fajszámához hasonlóan – az effektív gazdaszám/csoportméret törtszám is lehet.

Az  $M_f$  effektív csoportszám adott átlagos létszám és átlagos csoportméret mellett arányos a vizsgált közösség  $N$  tényleges csoportszámával. Hasonlítsuk össze  $M_f$  és  $N$  értékét.

**1. Állítás:** Tegyük fel, hogy  $f$  szigorúan monoton növekvő függvény. Ha a  $g(x) = f\left(\frac{1}{x}\right)$  függvény konvex, akkor  $C_f \geq f(\bar{X})$  és  $0 < M_f \leq N$ .

**Bizonyítás:** A Jensen-egyenlőtlenségből  $C_f = \sum_{i=1}^N \frac{X_i}{X} g\left(\frac{1}{X_i}\right) \geq g\left(\frac{N}{X}\right) = f(\bar{X})$ . Az  $f$  függvény szigorúan monoton növekvő, így  $f^{-1}(C_f) \geq \bar{X}$ , amiből  $0 < M_f \leq N$  adódik.



A továbbiakban végig feltesszük, hogy az  $f\left(\frac{1}{x}\right)$  függvény konvex. Az 1. állítás szerint ekkor az  $\frac{M_f}{N} = \frac{\bar{x}}{f^{-1}(C_f)}$  arány azt méri, hogy az egyedek mennyire egyenletesen oszlanak el a csoportok között (a paraziták mennyire egyenletesen oszlanak meg a gazdák között). Ha az arány kicsi, nulla körüli, akkor az eloszlás aggregált, foltos, ha 1 közelébe esik, akkor az eloszlás közelíti az egyenletest. Ennek megfelelően nevezhetnénk az  $\frac{\bar{x}}{f^{-1}(C_f)}$  arányt egyenletességi indexnek. Az egyenletesség fogalma a diverzitási indexek elméletében jelentős szerepet játszik (Hill 1973, Olszewski 2004). A zsúfoltság irodalmában az eloszlás egyenletessége helyett inkább az eloszlás foltosságát, aggregáltságát mérő, ezt kifejező indexek terjedtek el. Ilyen mérőszámhoz az egyenletességi index reciprokát véve jutunk: nevezzük  $f$ -aggregáltsági indexnek az  $A_f = \frac{f^{-1}(C_f)}{\bar{x}} = \frac{N}{M_f}$  mennyiséget. Ennek legkisebb értéke az 1. állítás szerint 1, ami az egyedek egyenletes eloszlásának felel meg. Aggregált eloszlásokra az index értéke 1-nél nagyobb és felülről nem korlátos. Az  $id(x) = x$  speciális esetben az aggregáltsági index a Lloyd-féle „mean demand” (Lloyd 1967) vagy az ezzel megegyező, Reiczigel et al. (2005) által bevezetett,  $id(x) = x$ -hez tartozó  $C_{id} = \frac{\sum_{i=1}^N X_i^2}{\bar{x}}$  lineáris zsúfoltság és az  $\bar{x}$  átlagos egyedszám hányadosa, képlettel  $A_{id} = \frac{\frac{1}{N} \sum_{i=1}^N X_i^2}{\bar{x}^2}$ . Az így értelmezett lineáris aggregáltsági index speciális esete a Bez-féle aggregáltsági indexnek (Bez, 2000). Az aggregáltság indexét meg kell különböztetni a Lloyd-féle foltossági indextől. Az  $f(x) = x - 1$  jelöléssel a Lloyd-féle foltossági index (patchiness index)  $P_{Lloyd} = \frac{C_f}{\bar{x}} = \frac{\frac{1}{N} \sum_{i=1}^N X_i \cdot (X_i - 1)}{\bar{x}^2}$  (Lloyd, 1967), az ennek az  $f$  függvénynek megfelelő aggregáltsági index viszont éppen az előbbi  $A_{id}$ . A kettő hasonlít egymásra, de nem egyeznek meg, egyik a másikból azonban egyszerűen kiszámítható a  $P_{Lloyd} = A_{id} - \frac{1}{\bar{x}}$  képlettel. Az összefüggésből látszik, hogy a foltossági index értéke 1-nél kisebb is lehet, ha pl. minden csoportban ugyanannyi egyed van, akkor értéke  $1 - \frac{1}{\bar{x}}$ . Lloyd úgy tervezte meg a foltossági indexet, hogy annak átlagos értéke 1 legyen a „teljesen véletlenszerű” csoportok közösségére, azaz független és ugyanolyan Poisson eloszlású egyedszámok mellett. A pontosan ugyanakkora méretű csoportok együttesének foltossága 1-nél kisebb, a Poisson eloszlásnál aggregáltabb egyedszám-eloszlások foltossága 1-nél nagyobb. Azonban egymástól független Poisson eloszlású csoportlétszámok mellett is megfigyelhető foltosság, ami biológiailag jelentős mértékű lehet különösen akkor, ha az egyedek csoportonkénti átlagos száma alacsony. Ezen kívül az egyedszámok más, Poissontól különböző, aggregált eloszlása mellett is előfordulhat, hogy a foltossági index 1. Ezért véleményem szerint helyesebb az aggregáltságot, foltosságot az előbb bevezetett aggregáltsági indexek valamelyikével mérni,

mert ezek értéke csak a teljesen egyenletes eloszlás mellett 1 és minden más esetben 1-nél nagyobb.

A lineáris aggregátsági index további előnye, hogy denzitásfüggetlen, azaz értéke nem változik, ha az abundanciák arányosan nőnek vagy csökkennek. A Lloyd-féle foltossági index denzitásfüggő, érzékeny a kis abundanciaértékekre, különösen a nulla gyakoriságokra (Bez 2000).

Csoportokból álló közösségek modellezése során általában nem kerül sor az összes csoport egyedeinek leszámolására. Ehelyett a csoportokból  $N$  elemű véletlen mintát veszünk és csak a mintába került csoportok egyedeit számoljuk meg. Ha a csoportok egyedszámai egymástól független és ugyanolyan eloszlású valószínűségi változók, akkor a nagy számok erős törvénye miatt az  $N$  mintanagyság növelésével a zsúfoltsági index (átlagos csoportméret) 1 valószínűséggel az  $\frac{EXf(X)}{EX}$  populációs értékhez konvergál, ahol  $X$  egy véletlenszerűen választott csoportban lévő egyedek száma. Az állítás érvényességéhez még feltesszük, hogy  $X$  és  $Xf(X)$  második momentuma létezik és véges. Az  $N$  mintanagyság növelésekor az  $M_f$  effektív csoportszám is korlátlanul nő, azonban hányadosuk, az  $A_f = \frac{f^{-1}(C_f)}{\bar{X}} = \frac{N}{M_f}$  aggregátsági index konvergál az  $\frac{f^{-1}\left(\frac{EXf(X)}{EX}\right)}{EX}$  populációs értékhez, ha  $f$  folytonos és szigorúan monoton növekvő.

A fejezet hátralevő részében az általános  $f$  függvény mellett értelmezett zsúfoltsági és aggregátsági indexeket speciális, a gyakorlati alkalmazások szempontjából fontos függvények esetében külön is megvizsgáljuk. Tekintsük az

$$f_q(u) = \begin{cases} \frac{u^{q-1} - 1}{q - 1}, & \text{ha } u > 0 \text{ és } q \geq 0, q \neq 1 \\ \ln(u), & \text{ha } u > 0 \end{cases}$$

alakú függvények családját. Elemi számítás mutatja, hogy  $\lim_{q \rightarrow 1} f_q(u) = \ln(u)$ , így mondhatjuk, hogy a természetes logaritmus függvényt  $q = 1$  mellett tartalmazza a függvényhalmaz.

Az  $f_q$  függvényhez  $q \neq 1$  esetében tartozó zsúfoltság

$$C_q = \frac{X^{q-1} \cdot \sum_{i=1}^N p_i^q - 1}{q - 1},$$

a megfelelő diverzitási index

$$D_q = f_q(X) - C_q = X^{q-1} \frac{1 - \sum_{i=1}^N p_i^q}{q - 1} = X^{q-1} \cdot d_q,$$

ahol  $p_i = X_i/X$ , az egyedek részaránya az  $i$ -ik csoportban. A  $D_q$  és  $d_q$  indexeket a denzitásfüggő  $X^{q-1}$  tényező kapcsolja össze. Patil és Taillie (1982) a  $-f_q$  függvényeket választva ritkasági függvénynek, részletesen tanulmányozták az ezekhez tartozó  $d_q = \frac{1 - \sum_{i=1}^N p_i^q}{q-1}$  diverzitási indexeket.

Az  $f_q$  függvény melletti effektív gazdaszám az

$$M_q = \left( \sum_{i=1}^N p_i^q \right)^{\frac{1}{1-q}},$$

az aggregáltsági index pedig az

$$A_q = N \left( \sum_{i=1}^N p_i^q \right)^{\frac{1}{q-1}}$$

képlettel számítható ki. A formula alakjából leolvasható, hogy az  $f_q$  függvényekhez tartozó effektív gazdaszám és aggregáltsági index denzitásfüggetlen.

Ha a csoportok/gazdák  $N$  száma tart a végtelenhez, akkor a  $C_q$  mintából becsült átlagos csoportméret/zsúfoltság tart az

$$\frac{EX f_q(X)}{EX} = \frac{1}{q-1} \left( \frac{EX^q}{EX} - 1 \right)$$

populációs értékhez, a mintából becsült aggregáltsági index pedig az

$$\frac{f_q^{-1} \left( \frac{EX f_q(X)}{EX} \right)}{EX} = \left( \frac{EX^q}{(EX)^q} \right)^{\frac{1}{q-1}}$$

populációs aggregáltsági indexhez konvergál. (Ennek következtében természetesen a populációs aggregáltsági index is denzitásfüggetlen minden  $q \geq 0$  esetén.)

A  $q = 2$  speciális esetben  $f_2(u) = u - 1$ . A csoportméret/zsúfoltság a

$$C_2 = \frac{\sum_{i=1}^N X_i(X_i - 1)}{X}$$

Lloyd-féle zsúfoltsági indexszel egyezik meg (Lloyd 1967). Az ehhez tartozó, ennek megfelelő diverzitási index

$$D_2 = X \cdot \left( 1 - \sum_{i=1}^N p_i^2 \right)$$

éppen a Simpson diverzitási index, szorozva az  $X$  denzitási tényezővel (Simpson 1949). Az aggregáltsági index a már korábban tárgyalt Lloyd-féle „mean demand” és az  $\bar{X}$  átlagos abundancia hányadosa, azaz

$$A_2 = A_{id} = \frac{1}{N} \frac{\sum_{i=1}^N X_i^2}{\bar{X}^2}$$

(vö. Lloyd 1967 és Reiczigel et al. 2005, 2008). Az  $N$  csoportszám növelésével határértékben a populációs zsúfoltság  $\frac{EX^2}{EX} - 1$ , a populációs aggregáltsági index pedig  $\frac{EX^2}{(EX)^2}$ .

Az  $f(u) = u$  függvényhez a Reiczigel et al. (2005) által bevezetett

$$\frac{\sum_{i=1}^N X_i^2}{X}$$

lineáris zsúfoltság tartozik (ez megegyezik a Lloyd-féle „mean demand”-dal, vö. Lloyd 1967). Az ennek megfelelő diverzitási függvény ugyanaz, mint az előző  $D_2$  diverzitási index, az aggregáltsági index pedig az  $A_2$  indexszel egyezik meg.

A  $q = 1$  esetben  $f_1(u) = \ln(u)$ , a zsúfoltság a  $C_1 = \frac{\sum_{i=1}^N X_i \cdot \ln(X_i)}{X}$  logaritmosos zsúfoltság (Reiczigel et al. 2005). A hozzá társítható diverzitási index

$$D_1 = \ln(X) - C_1 = - \sum_{i=1}^N p_i \cdot \ln(p_i)$$

éppen a Shannon diverzitási index (Shannon és Weaver 1963). Vegyük észre, hogy  $D_1$ -ben nincs denzitásfüggő tényező, ui.  $X_i^0 = 1$ . Érdemes megemlíteni, hogy a logaritmosos zsúfoltság

$$C_1 = \ln(X) + \sum_{i=1}^N p_i \cdot \ln(p_i)$$

alakban is felírható. Ebből látható, hogy a logaritmosos zsúfoltság a csupán az átlagos abundanciától és a csoportszámtól függő  $\ln(X) = \ln(N) + \ln(\bar{X})$  tag és a denzitástól nem függő Shannon diverzitási index különbsége.

Az effektív gazdaszám most

$$M_1 = \exp\left(- \sum_{i=1}^N p_i \cdot \ln(p_i)\right),$$

az aggregáltsági index pedig

$$A_1 = \frac{\exp(C_1)}{\bar{X}} = \frac{\exp\left(\frac{\sum_{i=1}^N X_i \cdot \ln(X_i)}{X}\right)}{\bar{X}} = N \cdot \exp\left(\sum_{i=1}^N p_i \cdot \ln(p_i)\right).$$

Ha  $N \rightarrow \infty$ , akkor a logaritmikus zsúfoltság populációs értéke  $\frac{EX \cdot \ln(X)}{EX}$ , az aggregáltsági indexé pedig  $\frac{\exp\left(\frac{EX \cdot \ln(X)}{EX}\right)}{EX}$ .

Ha  $q = 0$ , akkor a zsúfoltság skálázó függvénye  $f_0(u) = 1 - \frac{1}{u}$ , ezért az ehhez tartozó zsúfoltságot nevezzük el hiperbolikus zsúfoltságnak. Ha egy gazdán  $u > 0$  parazitaegyed él, akkor egy ott kiválasztott parazita szempontjából  $1 - \frac{1}{u}$  a parazitatársak részarányát jelenti. Ennek a parazitákra vett átlagát a

$$C_0 = \frac{\sum_{i=1}^N X_i \cdot \left(1 - \frac{1}{X_i}\right)}{\bar{X}} = 1 - \frac{N^+}{\bar{X}} = 1 - \frac{1}{\bar{X}} \cdot \frac{N^+}{N} = 1 - \frac{\hat{P}}{\bar{X}}$$

hiperbolikus zsúfoltság adja meg, ahol  $N^+$  azoknak a gazdáknak a száma, amelyeken parazita él és  $\hat{P} = \frac{N^+}{N}$  a gazdák parazitafertőzöttségének becsült prevalenciája. Ehhez az eredményhez a  $0 \cdot \frac{1}{0} = 0$  ideiglenes, csak ebben a képletben alkalmazott szabállyal jutunk. Azért célszerű ez a definíció, mert így  $C_0 = \lim_{q \downarrow 0} C_q$ , vagyis a zsúfoltság  $q = 0$ -ban jobbról folytonos lesz.

A  $C_0$  hiperbolikus zsúfoltság csupán az  $\bar{X}$  átlagos abundanciától és a  $\hat{P}$  becsült prevalenciától függ, értékét az átlagos abundancián túl csak csekély mértékben befolyásolja, hogy az egyedek hogyan oszlanak meg a gazdák/csoportok között, csupán a paraziták jelenlétét vagy hiányát veszi számításba. Az ehhez tartozó diverzitási index

$$D_0 = 1 - \frac{1}{\bar{X}} - \left(1 - \frac{N^+}{\bar{X}}\right) = \frac{1}{\bar{X}}(N^+ - 1),$$

itt a denzitásfüggetlen tényező éppen az  $N^+ - 1$  fajgazdagság (Tóthmérész 1997). A hiperbolikus aggregáltsági index  $A_0 = \frac{N}{N^+} = \frac{1}{\hat{P}}$ . A hiperbolikus zsúfoltság populációs értéke  $1 - \frac{P}{EX}$ , ahol  $P$  a gazdák parazitafertőzöttségének populációs prevalenciája. A hiperbolikus aggregáltság populációs értéke  $\frac{1}{P}$ . A  $q = 0$  eset tanulmányozásából érthető meg, hogy a zsúfoltsági és aggregáltsági indexek egyre kevésbé függenek az egyedek számának csoportok közötti eloszlásától, aggregáltságától, a csoportméretek különbözőségétől, ha  $q$  értéke 0-hoz közelít.

A  $C_q$ ,  $D_q$  és  $A_q$  indexek képletéből kiolvasható, hogy  $q > 1$  esetén a nagy létszámú, domináns csoportok a létszám szerinti részarányuknál nagyobb súllyal szerepelnek az indexekben,  $0 \leq q < 1$  esetén pedig a kis esetszámú csoportok súlya nagyobb a részarányuknál. A  $q = 1$  eset különleges, az ennek megfelelő logaritmikus zsúfoltsági és aggregáltsági indexben, valamint a Shannon diverzitási indexben a csoportok/gazdák súlya éppen a

létszámuk/parazitáltságuk arányának felel meg. A most említett, a részarányok súlyára vonatkozó tulajdonságok diverzitási indexekre régóta ismertek (Hill 1973, Patil és Taillie 1982, Jost 2006, 2007, Tuomisto 2010a). A Shannon diverzitási indexet e kedvező tulajdonságára hivatkozással, mint a leginkább kiegyensúlyozott indexet szokták ajánlani (Patil és Taillie 1982, Jost 2006, 2007).

A megfelelő zsúfoltsági és aggregáltsági index kiválasztásakor érdemes mérlegelni, hogy inkább a domináns, nagy létszámú csoportok súlya legyen kifejezettebb ( $q > 1$ ), vagy arányuknál jobban hangsúlyozza, emelje ki az index a kisebb létszámú csoportok hatását ( $q < 1$ ). A logaritmikus átlagos zsúfoltság/csoportméret ill. aggregáltsági index mellett szól, hogy a létszámarányok szempontjából ez a legkiegyensúlyozottabb. Különösen a logaritmusalapú  $A_1$  aggregáltsági index javasolható, ez a kiegyensúlyozottság mellett könnyen értelmezhető, lévén az egyedszám-skálára az exponenciális függvénnyel visszaskálázott zsúfoltság és az átlagos abundancia hányadosa. A logaritmikus indexek multiplikatív modellel leírható, arányaiban, százalékosan változó, biológiailag elsősorban arányai által meghatározott közösségekre természetes módon, könnyen értelmezhetőek (Reiczigel et al. 2005, 2008). A megfelelő index kiválasztásakor vegyük azonban azt is figyelembe, hogy a  $q = 2$ -höz tartozó Lloyd-index értelmezése a lehető legkézenfekvőbb, ui. ez az egy egyedre jutó csoporttársak átlagos száma. Hasonlóan egyszerű értelmezést tesz lehetővé a Reiczigel et al. (2005) által javasolt lineáris zsúfoltsági index, ez az egyedek saját csoportjainak/gazdáinak egyedekre átlagolt létszámát jelenti. További előnyük, hogy a zsúfoltsági skálázó függvény linearitása miatt itt nincs szükség skálatranszformációra. Hátrányuk lehet ugyanakkor, hogy a nagyobb csoportokat részarányukhoz képest nagyobb súllyal veszik figyelembe.

## 4.2 Két szempontos klasszifikáció, a zsúfoltság additív felbontása

A zsúfoltság és a diverzitás kapcsolata különösen jól szemléltethető, ha több parazitafaj is jelen van a gazdákon. Van-e összefüggés a parazitafajok diverzitása és zsúfoltsága között? Mit fejez ki a fajonkénti és a fajok együttesének zsúfoltsága? Hogyan viszonyulnak ezek a parazitafajok gazdánkénti és a gazdák együttesére számolt diverzitásához?

Tételezzük fel, hogy a gazdák végtelen populációján  $K$  különböző parazitafaj fordulhat elő. (Vagy mondhatnánk általánosabban, hogy a csoportok végtelen populációjában  $K$  különböző típusú egyed fordulhat elő. A továbbiak során a szemléletesebb tárgyalásmód érdekében minden a gazda–parazitafaj–parazitaegyed fogalomkör keretében kerül bemutatásra, miközben a tulajdonságok érvényesek lesznek az általánosabb csoport–egyed típus–egyed modellkeretben is.) Egymástól függetlenül kiválasztunk  $N$  gazdát, ezen a mintán a paraziták

gyakoriságait (abundanciáit) az 1. táblázat szerinti elrendezésben adjuk meg. A sorok egy-egy gazdán mutatják a parazitafajok egyedszámait, az oszlopokban egy-egy parazitafaj gazdák közötti egyedszám-megoszlása szerepel.

### 1. táblázat

*Parazitafajok gazdánkénti és együttes abundanciái*

		Fajok				Gazdák súlya
		$S_1$	$S_2$	...	$S_K$	
	$H_1$	$X_{11}$	$X_{12}$		$X_{1K}$	$p_{1.} = \frac{X_{1.}}{X_{..}}$
Gazdák	...					...
	$H_N$	$X_{N1}$	$X_{N2}$	...	$X_{NK}$	$p_{N.} = \frac{X_{N.}}{X_{..}}$
Fajok súlya		$p_{.1} = \frac{X_{.1}}{X_{..}}$	$p_{.2} = \frac{X_{.2}}{X_{..}}$		$p_{.K} = \frac{X_{.K}}{X_{..}}$	

Legyen  $X_{ij}$  az  $i$ -ik gazdán a  $j$ -ik parazitafaj gyakorisága,  $X_{i.}$  az  $i$ -ik gazdán az összes parazita száma,  $X_{.j}$  a  $j$ -ik parazitafaj összes egyedének száma minden gazdát számításba véve,  $X_{..}$  a paraziták száma fajra és gazdára összesítve. A relatív gyakoriságok  $p_{ij} = \frac{X_{ij}}{X_{..}}$ ,  $p_{i.} = \frac{X_{i.}}{X_{..}}$ ,  $p_{.j} = \frac{X_{.j}}{X_{..}}$ ,  $i = 1, \dots, N, j = 1, \dots, K$ .

Az 1. táblázat a gazdák és a parazitafajok szempontjából osztályozza a parazitaegyedek előfordulási gyakoriságait. Két,  $A$  és  $B$  szempont szerinti klasszifikáció teljes diverzitását Patil és Taillie (1982, 8.1-es formula) bontotta fel a szempontoknak megfelelő komponensekre. Ha  $R$  tetszőleges dichotom ritkasági függvény, akkor a teljes  $\Delta(A \times B)$  diverzitás felbontható az  $A$  szempont szerinti diverzitás és a  $B$  szempont  $A$  kategóriái melletti feltételes diverzitásának összegére (Patil és Taillie 1982, 8.1-8.3 formulák). Az említett szerzők által közölt legáltalánosabb 8.1 felbontást alkalmazzuk a parazita-egyedszámok gazdák ( $H$ ) és fajok ( $S$ ) szerinti klasszifikációjára az

$$R(p_{ij}) = f(X_{..}) - f(X_{..} \cdot p_{ij})$$

dichotom ritkasági függvény mellett. Itt  $f$  tetszőleges zsúfoltsági skálafüggvény, azaz monoton növekvő, a nemnegatív valós számokon értelmezett valós értékű függvény, melyről még feltesszük, hogy  $f\left(\frac{1}{x}\right)$  konvex.

Ha

$$\Delta(H \times S) = \sum_{i=1}^N \sum_{j=1}^K p_{ij} \cdot R(p_{ij})$$

és

$$\Delta(H) = \sum_{i=1}^N p_{i.} \cdot R(p_{i.}),$$

akkor a Patil-Taillie felbontás szerint

$$\Delta(H \times S) = \Delta(H) + \sum_{i=1}^N p_{i.} \sum_{j=1}^K \frac{p_{ij}}{p_{i.}} (R(p_{ij}) - R(p_{i.})).$$

Visszaírva ebbe a formulába a ritkasági függvény képletét, átrendezés után kapjuk, hogy

$$\sum_{j=1}^K \frac{X_{.j}}{X_{..}} \cdot \frac{\sum_{i=1}^N X_{ij} f(X_{ij})}{X_{.j}} + \sum_{i=1}^N \frac{X_{i.}}{X_{..}} \left[ f(X_{i.}) - \frac{\sum_{j=1}^K X_{ij} f(X_{ij})}{X_{i.}} \right] = \frac{\sum_{i=1}^N X_{i.} f(X_{i.})}{X_{..}}. \quad (1)$$

(Ez az összefüggés közvetlenül, egyszerű számolással is ellenőrizhető.)

A jobboldalon a parazitafajok együttes zsúfoltsága áll, az a zsúfoltság, amiben fajtól függetlenül minden gazdán minden parazitaegyed szerepe ugyanaz. A baloldal első tagja a parazitafajonkénti zsúfoltságoknak a fajok egyedszámával súlyozott átlaga. A fajonkénti zsúfoltságban csak az adott parazitafaj kap szerepet, a többi parazita jelenléte ezt nem befolyásolja. A paraziták szintjén a baloldal első tagja a saját faj zsúfoltságának aritmetikai átlaga. A baloldal második tagjában a szögletes zárójelben lévő kifejezés éppen a parazitafajok diverzitása az  $i$ -edik gazdán. A második tag ezeknek a diverzitásoknak a paraziták szemszögéből vett aritmetikai átlaga. Ez pontosabban azt jelenti, hogy minden parazitához kiszámítjuk a saját gazdáján ezt a diverzitást, majd e diverzitásokat átlagoljuk a parazitákra. Ezzel a paraziták saját környezetében (gazdaállatán) lévő, a paraziták életkörülményeivel, lehetőségeivel kapcsolatba hozható lokális vagy szubjektív diverzitás fogalmához jutunk. A szubjektív diverzitást ebben az értelemben, de egy speciális diverzitási indexre, a Brillouin indexre Lloyd (1967) vezette be. Összefoglalva, a kapott (1) egyenlőség a

$$\text{fajonkénti átlagos zsúfoltság} + \text{szubjektív diverzitás} = \text{fajok együttes zsúfoltsága}$$

összefüggést jelenti.

Egy gazdaállaton lévő parazitafaj-diverzitás a fajok lokális, a kérdéses gazdán lévő niche-inek átfedését méri, számszerűsíti. Nevezhetjük ezért a szubjektív diverzitást szubjektív niche átfedésnek (subjective niche overlap) is, vö. Patil-Taillie (1982).



Az (1) felbontásban a szubjektív diverzitásnak megfelelő tagot az együttes zsúfoltság és az egységre normált lokális, gazdánkénti diverzitások súlyozott átlagának szorzataként is kifejezhetjük:

$$\sum_{j=1}^K \frac{X_j}{X_{..}} \cdot \frac{\sum_{i=1}^N X_{ij} f(X_{ij})}{X_j} + \frac{\sum_{i=1}^N X_i f(X_i)}{X_{..}} \cdot \frac{\sum_{i=1}^N X_i f(X_i) \left[ 1 - \frac{\sum_{j=1}^K X_{ij} f(X_{ij})}{X_i f(X_i)} \right]}{\sum_{i=1}^N X_i f(X_i)} = \frac{\sum_{i=1}^N X_i f(X_i)}{X_{..}} \quad (2)$$

A (2) felbontás a szubjektív diverzitást az együttes zsúfoltság részarányaként, annak százalékában fejezi ki. Ez a forma bizonyos esetekben megkönnyítheti a biológiai értelmezést.

Az (1) felbontás alkalmazásakor érdemes a fajonkénti zsúfoltságok paraziták szintjén értelmezett hisztogramját ábrázolni. (Minden parazitához kiszámítjuk a saját fájának zsúfoltságát, majd megrajzoljuk ezeknek a zsúfoltságoknak a hisztogramját.) A kapott zsúfoltságok átlaga (1) baloldalának első tagja. A hisztogram megmutatja a fajonkénti zsúfoltságok átlag körüli eloszlását. Az (1) felbontás gazdánkénti szubjektív diverzitásainak paraziták szintjén vett hisztogramja az (1) baloldalának második tagja körüli eloszlást adja meg. A (2) felbontás baloldalának második tagjának második tényezője az egységre normált szubjektív diverzitások súlyozott átlaga. Ehhez kapcsolható a normált diverzitások súlyozott sűrűségfüggvénye, ami pl. az R 3.0.2 statisztikai szoftver *stats* csomagjának *density* eljárásával hozható létre (R Core Team 2014). A sűrűségfüggvény az egységre normált szubjektív diverzitások eloszlását mutatja (2) baloldalának második tagjának második tényezője, mint súlyozott átlag körül. Az (1) és (2) felbontás tagjaihoz kapcsolható hisztogramokra példa a 6. fejezet 2-5. és 9-12. ábrásorozata.

A parazitaegyedek gazdánkénti  $X_i$  abundanciáját rögzítve, a szubjektív diverzitás akkor a legnagyobb, ha minden gazdán minden faj egyenletesen oszlik el, azaz  $X_{ij} = \frac{X_i}{K}$ ,  $i = 1, \dots, N, j = 1, \dots, K$ . Ez a megállapítás a 4.1. fejezet 1. állításának következménye. A szubjektív diverzitás legkisebb értéke 0, ezt akkor éri el, ha minden gazdán pontosan egy parazitafaj fordul elő,  $X_i$  egyedszámmal. Rögzített  $X_i$  értékek mellett a fajok együttes zsúfoltsága is rögzített, így a fajonkénti átlagos zsúfoltság akkor minimális, ha a szubjektív diverzitás maximális, vagyis ha  $X_{ij} = \frac{X_i}{K}$ ,  $i = 1, \dots, N, j = 1, \dots, K$ . Ugyanígy, a fajonkénti átlagos zsúfoltság maximális, ha minden gazdán pontosan egy parazitafaj fordul elő,  $X_i$  egyedszámmal.

Ha csupán az  $X_{..}$  összes parazita-egyedszámot tekintjük rögzítettnek, akkor az együttes zsúfoltság minimális, ha minden gazdán ugyanannyi parazita él, azaz  $X_i = \frac{X}{N}$  (vö. 4.1. fejezet 1. állításával). Az együttes zsúfoltság maximális, ha minden parazita ugyanazon a gazdán él. Ugyanígy, a fajonkénti átlagos zsúfoltság minimális, ha minden gazda minden

parazitafajának ugyanannyi,  $X_{ij} = \frac{X_i}{K \cdot N}$  egyede van. A fajonkénti átlagos zsúfoltság maximális, ha csak egyetlen parazitafaj fordul elő egyetlen gazdán. Ekkor a fajonkénti átlagos zsúfoltság nyilvánvalóan megegyezik az együttes zsúfoltsággal.

A szubjektív diverzitás legkisebb értéke 0, ezt akkor éri el, ha minden gazdán pontosan egy parazitafaj fordul elő. (Ha egy faj se fordulna elő valamelyik gazdán, akkor az a gazda nem lenne parazitált és így kizárhatnánk anélkül, hogy a zsúfoltság- és diverzitási értékek megváltoznának.)

Ha  $X_i$  rögzített, akkor az  $i$ -ik gazdán a parazita fajdiverzitás akkor a legnagyobb, ha  $X_{ij} = \frac{X_i}{K}, j = 1, \dots, K$ . Ilyenkor a diverzitás értéke  $f(X_i) - f\left(\frac{X_i}{K}\right)$ , ezt (1)-be helyettesítve kapjuk a szubjektív diverzitás maximumát. A  $h(x) = f(x) - f\left(\frac{x}{K}\right)$  függvény monoton nő, ui. az  $f\left(\frac{1}{x}\right)$  függvényről feltettük, hogy konvex. Ezért  $X_i$  rögzítése mellett a szubjektív diverzitás maximális, ha minden parazitafajnak ugyanannyi egyede van és az összes parazita egyetlen gazdához tartozik. Ez a gazdák szempontjából a leginkább aggregált, a parazitafajok szempontjából pedig a legegyszerűsebb, legdiverzebb abundancia-eloszlásnak felel meg. Az (1) és (2) összefüggés bármilyen gazda–parazitafaj–parazita abundanciákra érvényes, nem veszi figyelembe, hogy az adatoknak milyen feltételeknek kell eleget tenniük. Tegyük fel, hogy a gazdákat egymástól függetlenül mintavételezzük és minden gazdán ugyanolyan valószínűségeloszlású a parazitafajok gyakoriságaiból összeállított  $\underline{X}_j = (X_{j1}, X_{j2}, \dots, X_{jK})^T$  valószínűségi vektorváltozó. A továbbiak során feltesszük, hogy az abundanciák gazdánkénti, fajonkénti összegeinek, ezek  $f$ -transzformáltjainak, mindezek négyzeteinek és páronkénti szorzatainak létezik és véges a várható értéke.

Ha a gazdák  $N$  száma végtelenbe tart, akkor a nagy számok erős törvénye miatt az (1) és (2) egyenlőségekben szereplő tagok az alábbi egyenlőségek megfelelő tagjaihoz konvergálnak 1 valószínűséggel. Az együttes zsúfoltság (1)-nek megfelelő populációs felbontása az

$$\sum_{j=1}^K \frac{EX_{1j}}{EX_1} \cdot \frac{EX_{1j}f(X_{1j})}{EX_{1j}} + \frac{1}{EX_1} \cdot E \left\{ X_1 \cdot \left[ f(X_1) - \frac{\sum_{j=1}^K X_{1j}f(X_{1j})}{X_1} \right] \right\} = \frac{EX_1 f(X_1)}{EX_1} \quad (3)$$

egyenlőségre vezet, (2)-nek pedig a

$$\sum_{j=1}^K \frac{EX_{1j}}{EX_1} \cdot \frac{EX_{1j}f(X_{1j})}{EX_{1j}} + \frac{EX_1 f(X_1)}{EX_1} \cdot \frac{E \left\{ X_1 f(X_1) \left[ 1 - \frac{\sum_{j=1}^K X_{1j}f(X_{1j})}{X_1 f(X_1)} \right] \right\}}{EX_1 f(X_1)} = \frac{EX_1 f(X_1)}{EX_1} \quad (4)$$

populációs felbontás felel meg.

A most következő részben alkalmazzuk az együttes zsúfoltság (1), (2), (3) és (4) felbontását a már vizsgált és a gyakorlati felhasználás szempontjából legfontosabb

$$f_q(u) = \begin{cases} \frac{u^{q-1} - 1}{q-1}, & \text{ha } u > 0 \text{ és } q \geq 0, q \neq 1 \\ \ln(u), & \text{ha } u > 0 \end{cases}$$

függvények családjára. A  $q \neq 1$  feltétel mellett az (1) felbontás

$$\sum_{j=1}^K p_j \frac{X_j^{q-1} \sum_{i=1}^N p_{ij}^q - 1}{q-1} + \sum_{i=1}^N p_i \left[ X_i^{q-1} \frac{1 - \sum_{j=1}^K p_{ji}^q}{q-1} \right] = \frac{X_{..}^{q-1} \sum_{i=1}^N p_i^q - 1}{q-1}, \quad (5)$$

a (2) felbontás pedig

$$\sum_{j=1}^K p_j \frac{X_j^{q-1} \sum_{i=1}^N p_{ij}^q - 1}{q-1} + \frac{X_{..}^{q-1} \sum_{i=1}^N p_i^q - 1}{q-1} \cdot \frac{\sum_{i=1}^N X_i (X_i^{q-1} - 1) \left[ 1 - \frac{\sum_{j=1}^K X_{ij} (X_{ij}^{q-1} - 1)}{X_i (X_i^{q-1} - 1)} \right]}{\sum_{i=1}^N X_i (X_i^{q-1} - 1)} = \frac{X_{..}^{q-1} \sum_{i=1}^N p_i^q - 1}{q-1} \quad (6)$$

alakot ölt, ahol  $p_{i|j} = \frac{X_{ij}}{X_j}$ ,  $p_{j|i} = \frac{X_{ij}}{X_i}$ ,  $i = 1, \dots, N$ ,  $j = 1, \dots, K$ .

Az (5) felbontás populációs megfelelője

$$\sum_{j=1}^K \frac{EX_{1j}}{EX_1} \cdot \frac{1}{q-1} \left( \frac{EX_{1j}^q}{EX_1^q} - 1 \right) + \frac{EX_{1.}^q}{EX_1} \cdot \frac{1 - \sum_{j=1}^K \frac{EX_{1j}^q}{EX_1^q}}{q-1} = \frac{1}{q-1} \left( \frac{EX_1^q}{EX_1} - 1 \right), \quad (7)$$

a (6) felbontásé

$$\sum_{j=1}^K \frac{EX_{1j}}{EX_1} \cdot \frac{1}{q-1} \left( \frac{EX_{1j}^q}{EX_1^q} - 1 \right) + \frac{1}{q-1} \left( \frac{EX_1^q}{EX_1} - 1 \right) \cdot \frac{1 - \sum_{j=1}^K \frac{EX_{1j}^q}{EX_1^q}}{1 - \frac{EX_1^q}{EX_1}} = \frac{1}{q-1} \left( \frac{EX_1^q}{EX_1} - 1 \right). \quad (8)$$

Az (5) felbontás  $q = 2$  esetén, az  $f_2(u) = u - 1$  zsúfoltsági skálafüggvénnyel a Lloyd-féle zsúfoltsági indexre (Lloyd, 1967) a

$$\sum_{j=1}^K p_j (X_j \sum_{i=1}^N p_{ij}^2 - 1) + \sum_{i=1}^N p_i X_i (1 - \sum_{j=1}^K p_{ji}^2) = X_{..} \sum_{i=1}^N p_i^2 - 1 \quad (9)$$

képletet eredményezi. Ennek populációs változata

$$\sum_{j=1}^K \frac{EX_{1j}}{EX_1} \cdot \left( \frac{EX_{1j}^2}{EX_1^2} - 1 \right) + \frac{EX_1^2}{EX_1} \cdot \left( 1 - \sum_{j=1}^K \frac{EX_{1j}^2}{EX_1^2} \right) = \frac{EX_1^2}{EX_1} - 1. \quad (10)$$

A (6) felbontás a Lloyd-féle zsúfoltságra

$$\sum_{j=1}^K p_j (X_j \sum_{i=1}^N p_{ij}^2 - 1) + (X_{..} \sum_{i=1}^N p_i^2 - 1) \cdot \frac{\sum_{i=1}^N X_i (X_i - 1) \left[ 1 - \frac{\sum_{j=1}^K X_{ij} (X_{ij} - 1)}{X_i (X_i - 1)} \right]}{\sum_{i=1}^N X_i (X_i - 1)} = X_{..} \sum_{i=1}^N p_i^2 - 1. \quad (11)$$

Megjegyzendő, hogy (11)-ben a gazdánkénti normált szubjektív, lokális diverzitást kifejező  $1 - \frac{\sum_{j=1}^K X_{ij}(X_{ij}-1)}{X_i(X_i-1)}$  formula annak valószínűségét adja, hogy az  $i$ -ik gazdán két véletlenszerűen kiválasztott parazita nem tartozik ugyanahhoz a fajhoz. Ez az érték tekinthető a Simpson diverzitási index véges egyedszám melletti változatának (Simpson 1949). A megfelelő populációs felbontás

$$\sum_{j=1}^K \frac{EX_{1j}}{EX_1} \cdot \left( \frac{EX_{1j}^2}{EX_{1j}} - 1 \right) + \left( \frac{EX_1^2}{EX_1} - 1 \right) \cdot \frac{1 - \sum_{j=1}^K \frac{EX_{1j}^2}{EX_1^2}}{1 - \frac{EX_1}{EX_1^2}} = \frac{EX_1^q}{EX_1} - 1. \quad (12)$$

A Lloyd-féle zsúfoltság (9) felbontása a következőképpen értelmezhető:

<i>Saját fajhoz tartozó, saját gazdán élő parazitátársak számának átlaga</i>	+	<i>Saját gazdán élő paraziták átlagos abundanciával súlyozott Simpson fajdiverzitásainak átlaga</i>	=	<i>Összes, saját gazdán élő parazitátárs számának átlaga</i>
--	---	---	---	--

(Ebben a kontextusban az átlag a parazitákra, mint szubjektumokra vett átlagot jelenti. )

A Lloyd-féle zsúfoltság (11) felbontásának szöveges értelmezése:

<i>Saját fajhoz tartozó, saját gazdán élő parazitátársak számának átlaga</i>	+	<i>Az összes, saját gazdán élő parazitátárs számának átlagából a gazdánkénti Simpson fajdiverzitások súlyozott átlagának megfelelő részarány</i>	=	<i>Összes, saját gazdán élő parazitátárs számának átlaga</i>
--	---	--	---	--

A Reiczigel et al. (2005) által definiált,  $f(u) = u$  skálafüggvényhez tartozó lineáris zsúfoltság esetén az (1) felbontás

$$\sum_{j=1}^K p_j X_{.j} \sum_{i=1}^N p_{i|j}^2 + \sum_{i=1}^N p_i X_i (1 - \sum_{j=1}^K p_{j|i}^2) = X_{..} \sum_{i=1}^N p_i^2 \quad (13)$$

alakot ölt. A populációs felbontás a

$$\sum_{j=1}^K \frac{EX_{1j}}{EX_1} \cdot \frac{EX_{1j}^2}{EX_{1j}} + \frac{EX_1^2}{EX_1} \cdot \left[ 1 - \sum_{j=1}^K \frac{EX_{1j}^2}{EX_1^2} \right] = \frac{EX_1^2}{EX_1} \quad (14)$$

képlettel számítható ki. A (2) felbontás most

$$\sum_{j=1}^K p_j X_{.j} \sum_{i=1}^N p_{i|j}^2 + X_{..} \sum_{i=1}^N p_i^2 \cdot \frac{\sum_{i=1}^N p_i^2 [1 - \sum_{j=1}^K p_{j|i}^2]}{\sum_{i=1}^N p_i^2} = X_{..} \sum_{i=1}^N p_i^2 \quad (15)$$

alakú. Vegyük észre, hogy (15) baloldalán a szubjektív, gazdánkénti Simpson diverzitások súlyozott átlaga denzitásfüggetlen. A (15) formula populációs megfelelője (14).

A lineáris zsúfoltság (13) felbontása szavakkal kifejezve, képletek nélkül:

$$\begin{array}{l} \text{Saját fajhoz tartozó, saját} \\ \text{gazdán élő paraziták} \\ \text{számának átlaga} \end{array} + \begin{array}{l} \text{Saját gazdán élő paraziták} \\ \text{átlagos abundanciával} \\ \text{súlyozott Simpson} \\ \text{fajdiverzitásainak átlaga} \end{array} = \begin{array}{l} \text{Összes, saját gazdán} \\ \text{élő parazita számának} \\ \text{átlaga} \end{array}$$

A lineáris zsúfoltság (15) felbontásának értelmezése:

$$\begin{array}{l} \text{Saját fajhoz tartozó, saját} \\ \text{gazdán élő paraziták} \\ \text{számának átlaga} \end{array} + \begin{array}{l} \text{Az összes, saját gazdán} \\ \text{élő parazita számának} \\ \text{átlagából a gazdánkénti} \\ \text{Simpson fajdiverzitások} \\ \text{súlyozott átlagának} \\ \text{megfelelő részarány} \end{array} = \begin{array}{l} \text{Összes, saját gazdán} \\ \text{élő parazita számának} \\ \text{átlaga} \end{array}$$

Számoljuk ki most a felbontást a  $q = 1$  esetnek megfelelő logaritmusos zsúfoltságra! A zsúfoltság skálafüggvénye  $f_1(u) = \ln(u)$ . Az (1) képletbe behelyettesítve kapjuk, hogy

$$\begin{aligned} & \sum_{j=1}^K p_j (\ln X_j + \sum_{i=1}^N p_{i|j} \ln p_{i|j}) + \sum_{i=1}^N p_i (-\sum_{j=1}^K p_{j|i} \ln p_{j|i}) = \\ & \ln X_{..} + \sum_{i=1}^N p_i \ln p_i. \end{aligned} \quad (16)$$

A felbontás első tagja, a fajonkénti átlagos logaritmusos zsúfoltság átírható  $\ln X_{..} + \sum_{i=1}^N p_{ij} \ln p_{ij}$  alakúra. Ebből látszik, hogy a fajonkénti átlagos zsúfoltságban és a jobboldalon lévő együttes zsúfoltságban a denzitásfüggő komponens ugyanakkora és megegyezik  $\ln X_{..}$ -tal. A logaritmusos zsúfoltság felbontásában a szubjektív diverzitásnak nincs denzitásfüggő komponense. A szubjektív diverzitás egyszerűen a paraziták gazdáin a Shannon diverzitások parazitánként vett aritmetikai átlaga. A teljes gazdapopuláción a logaritmusos zsúfoltság (16)-nak megfelelő additív felbontása

$$\frac{1}{EX_1} \cdot \sum_{j=1}^K EX_{1j} \ln X_{1j} + \frac{1}{EX_1} \cdot \left\{ -\sum_{j=1}^K EX_{1j} \ln \frac{X_{1j}}{X_1} \right\} = \frac{EX_1 \ln X_1}{EX_1}. \quad (17)$$

A szubjektív diverzitást a (2) felbontással az együttes zsúfoltság részarányaként kifejezve, a

$$\begin{aligned} & \sum_{j=1}^K p_j (\ln X_j + \sum_{i=1}^N p_{i|j} \ln p_{i|j}) + \frac{\sum_{i=1}^N X_i \ln X_i}{X_{..}} \cdot \frac{\sum_{i=1}^N X_i \ln X_i \left( \frac{-\sum_{j=1}^K p_{j|i} \ln p_{j|i}}{\ln X_i} \right)}{\sum_{i=1}^N X_i \ln X_i} = \\ & \frac{\sum_{i=1}^N X_i \ln X_i}{X_{..}} \end{aligned} \quad (18)$$

formulát nyerjük. A baloldalon lévő szubjektív diverzitás második tényezője a gazdánkénti Shannon egyenletességi indexek súlyozott átlaga (Magurran 2004, Pielou 1967). A populációs változat

$$\frac{1}{EX_1} \cdot \sum_{j=1}^K EX_{1j} \ln X_{1j} + \frac{EX_1 \ln X_1}{EX_1} \cdot \frac{-\sum_{j=1}^K EX_{1j} \ln \frac{X_{1j}}{X_1}}{EX_1 \ln X_1} = \frac{EX_1 \ln X_1}{EX_1}. \quad (19)$$

A logaritmusos zsúfoltság (16) felbontásának jelentése, értelmezése a következő:

$$\begin{array}{ccc} \text{Saját fajhoz tartozó, saját} & & \text{Saját gazdán élő paraziták} \\ \text{gazdán élő paraziták} & & \text{Shannon} \\ \text{számának logaritmus} & + & \text{fajdiverzitásainak átlaga} \\ \text{skálán vett átlaga} & & \\ & & = \\ & & \text{Saját gazdán élő} \\ & & \text{összes parazita} \\ & & \text{számának logaritmus} \\ & & \text{skálán vett átlaga} \end{array}$$

(Az átlag a parazitaegyedekre vett aritmetikai átlagot jelenti most is, ugyanúgy, mint a Lloyd-féle és a lineáris zsúfoltság esetében, illetve az általános (1) felbontásnál.)

A zsúfoltság logaritmusos skálán történő kiszámítása multiplikatív modellt eredményez, ami olyan biológiai rendszerekre alkalmazható, ahol a hatások, változások arányokkal, százalékokkal fejezhető ki. A zsúfoltság (16) felbontását logaritmusos skáláról az eredeti abundanciaskálára visszatranszformálva szorzatmodellhez jutunk. A baloldal első tagjából a saját fajhoz tartozó paraziták számának mértani átlaga, a másodikból a saját gazdához tartozó Shannon effektív fajszámok mértani átlaga, míg a jobboldalról a saját gazdán élő paraziták számának mértani átlaga lesz. A szorzatmodell ezek alapján:

$$\begin{array}{ccc} \text{Saját fajhoz tartozó, saját} & & \text{Saját gazdán élő paraziták} \\ \text{gazdán élő paraziták} & \times & \text{Shannon effektív} \\ \text{számának mértani átlaga} & & \text{fajszámainak mértani átlaga} \\ & & = \\ & & \text{Saját gazdán élő} \\ & & \text{összes parazita} \\ & & \text{számának mértani} \\ & & \text{átlaga} \end{array}$$

A második tényezőt nevezzük el szubjektív fajszámnak, ez a paraziták környezetében, lokálisan, saját gazdájukon lévő parazitafajok effektív fajszámainak átlaga. Ez a tényező kapcsolja össze multiplikatív modellben az egy fajhoz tartozó átlagos zsúfoltságot a fajok együttesének zsúfoltságával.

A  $q = 0$  esetben a hiperbolikus zsúfoltság egyszerű függvénye a paraziták átlagos abundanciájának és a paraziták prevalenciájának. A paraziták gazdák közötti eloszlására kevésbé érzékeny, csak a paraziták jelenlétét vagy hiányát veszi számításba. A zsúfoltság (5) felbontása most a

$$\sum_{j=1}^K p_j \left(1 - \frac{\hat{p}_j}{\bar{X}_j} N\right) + \sum_{i=1}^N p_i \frac{S_i - 1}{\bar{X}_i} = 1 - \frac{\hat{P}}{\bar{X}} \quad (20)$$

formulát eredményezi, ahol  $\hat{p}_j$  a  $j$ -ik parazitafaj becsült prevalenciája,  $\hat{P}$  a parazitáltság prevalenciájának becslése, fajra való tekintet nélkül,  $S_i$  az  $i$ -ik gazdán élő parazitafajok száma,  $\bar{X} = \frac{X}{N}$  az egy gazdára jutó paraziták együttes száma. A képletek kiértékelésekor a  $0 \cdot \frac{1}{0} = 0$  ideiglenes szabályt kell alkalmazni. A baloldal első tagja csak a parazitafajok arányától, gazdákra átlagolt egyedszámától és prevalenciájától függ. A második tagban a szubjektív diverzitás denzitásfüggetlen része éppen az  $S_i - 1$  fajgazdagság. A jobboldal csak

az  $\bar{X}$  egy gazdára jutó paraziták számától és a nem fajspecifikus parazitáltság prevalenciájától függ. A felbontás populációs verziója az

$$\sum_{j=1}^K \frac{EX_{1j}}{EX_1} \left(1 - \frac{P_j}{EX_{1j}}\right) + \frac{ES_1 - P}{EX_1} = 1 - \frac{P}{EX_1} \quad (21)$$

egyenlőség. A felbontásban  $P_j$  a  $j$ -ik parazitafaj populációs szintű prevalenciája,  $P$  a gazdák parazitáltságának populációs szintű prevalenciája, parazitafajra tekintet nélkül.

A  $q = 0$  esetben a hiperbolikus zsúfoltság (6) felbontása

$$\sum_{j=1}^K p_j \left(1 - \frac{\hat{p}_j}{x_j} N\right) + \left(1 - \frac{\hat{p}}{\bar{X}}\right) \cdot \frac{\sum_{i=1}^N \max(X_i - 1, 0) \cdot \frac{S_i - 1}{X_i - 1}}{\sum_{i=1}^N \max(X_i - 1, 0)} = 1 - \frac{\hat{p}}{\bar{X}} \quad (22)$$

alakúra hozható, a  $0 \cdot \frac{0}{0} = 0$  ideiglenes szabály mellett. Ennek populációs változata

$$\sum_{j=1}^K \frac{EX_{1j}}{EX_1} \left(1 - \frac{P_j}{EX_{1j}}\right) + \left(1 - \frac{P}{EX_1}\right) \cdot \frac{ES_1 - P}{EX_1 - P} = 1 - \frac{P}{EX_1} \quad (23)$$

Megjegyzendő, hogy 10-nél nagyobb átlagos abundanciánál (20) és (22) baloldalának első tagja közel lesz 1-hez, ezért a második tag szükségképpen kicsi, így a felbontás első ránézésre nehezen értelmezhető. Érdekes ilyenkor a baloldal első tagjának és a jobboldalnak az 1-től való eltéréseit összehasonlítani a baloldal második tagjával, vagyis a szubjektív diverzitással.

A  $q = 0$ -hoz tartozó  $C_0$  hiperbolikus zsúfoltság felbontása a következőképpen értelmezhető:

<i>Saját fajhoz tartozó, saját gazdán élő parazitátársak számarányának átlaga</i>	+	<i>Saját gazdán élő parazitáknak az átlagos abundancia reciprokával súlyozott fajgazdagságainak átlaga</i>	=	<i>Összes, saját gazdán élő parazitátárs számarányának átlaga</i>
---	---	--	---	---

(Az átlag itt is a parazitákra, mint szubjektumokra vett átlagot jelenti.)

A  $q$  kitevőnek az (5) felbontás összetevőire gyakorolt hatásának értékeléséhez előbb adjunk hozzá (5) mindkét oldalához  $\frac{1}{q-1}$ -et, majd osszuk el a kapott egyenlőséget  $X^{q-1}$ -gyel. Így a denzitásfüggetlen

$$\sum_{j=1}^K p_j^q \cdot \frac{\sum_{i=1}^N p_{ij}^q}{q-1} + \sum_{i=1}^N p_i^q \cdot \frac{1 - \sum_{j=1}^K p_{ji}^q}{q-1} = \frac{\sum_{i=1}^N p_i^q}{q-1} \quad (24)$$

összefüggéshez jutunk. Konstans hozzáadása és konstanssal való szorzás nem változtat az összetevők részarányukhoz képesti hatásán, ezért (5) értékelését a (24) egyenlőség alapján végezzük el.

Ha  $q$  értéke 1-nél nagyobb, akkor az (5) és (24) felbontás első, a parazitafajok átlagos zsúfoltságát mérő tagja a nagyobb létszámú fajok zsúfoltságát részarányuknál nagyobb mértékben veszi figyelembe. Egy-egy faj zsúfoltságát ugyanakkor az adott fajjal parazitáltabb

gazdák a parazita-részarányuknál nagyobb mértékben növelik. A felbontás második, a parazitafajok szubjektív diverzitását kifejező tagjára a parazitáltabb gazdák parazita-részarányuknál erősebben hatnak. Emellett egy-egy gazdán a fajdiverzitásra a lokálisan domináns fajok részarányuknál nagyobb mértékben hatnak.

Ha  $q$  értéke 1-nél kisebb, akkor a felbontás fajonkénti átlagos zsúfoltságot mérő tagját a kisebb létszámú fajok zsúfoltsága részarányuknál nagyobb mértékben befolyásolja. Egy-egy faj zsúfoltságát ugyanakkor az adott fajjal kevésbé parazitált gazdák a parazita-részarányukhoz képest nagyobb arányban változtatják. A szubjektív diverzitásra a kisebb arányban parazitált gazdák parazita-részarányuknál erősebben hatnak. Ezzel párhuzamosan egy-egy gazda fajdiverzitásában az ott élő ritkább fajok szerepe részarányukhoz képest nagyobb.

A logaritmikus zsúfoltság (16), (17), (18) és (19) felbontása a  $q = 1$  esetnek felel meg. Az  $f_q$  függvénycsalád és az összes eddig bemutatott, az  $f_q$  skálafüggvényre támaszkodó zsúfoltság- és diverzitáskomponens  $q = 1$ -ben folytonos, ezért megállapítható, hogy a logaritmikus zsúfoltság felbontásának minden komponensében a paraziták szerepe fajonkénti és gazdánkénti részarányuknak éppen megfelelő mértékű.

A Lloyd-féle és a vele ekvivalens lineáris zsúfoltság felbontásán kívül fentiek miatt javasolt a logaritmikus zsúfoltság felbontását is kiszámítani. Ezzel kiértékelhetővé válik, hogy a fajonkénti Lloyd-féle és a lineáris zsúfoltságra, valamint a szubjektív Simpson-diverzitásra a domináns gazdák és parazitafajok milyen, parazitáltságukat ill. számarányukat meghaladó mértékű hatást gyakorolnak.

A kis prevalenciájú parazitafajok lehető legnagyobb mértékű reprezentálását teszi lehetővé a  $q = 0$ -hoz tartozó  $C_0$  hiperbolikus zsúfoltság. Ennek (20) és (22) felbontását, mint extrém határesetet érdemes kiszámolni és a  $q > 0$ -hoz tartozó felbontásokkal összehasonlítani.

### **4.3 A valódi diverzitás és az aggregátsági index multiplikatív felbontása**

Közösségek diverzitásait az  $\alpha$  átlagos diverzitással és az egyesített közösség  $\gamma$  diverzitásával mérjük és hasonlítjuk össze. A közösségek diverzitása változatosságának jellemzésére Whittaker (1960, 1972) bevezette a  $\beta = \frac{\gamma}{\alpha}$  diverzitást. Az előző fejezetben tanulmányozott gazda–parazitafaj–parazitaegyed, vagy általánosabban csoport–egyed típus–egyed hierarchia az egyes csoportok egyed típusaiból álló közösségek együtteseként fogható fel. A csoportonkénti egyed típusok diverzitásait átlagolva az  $\alpha$  diverzitást, az egyesített csoportok közösségén az egyed típusok diverzitása a  $\gamma$  diverzitást eredményezi. A diverzitás



csoportok közötti változatosságát a  $\beta = \frac{\gamma}{\alpha}$  diverzitás jellemzi. Az előző fejezethez hasonlóan most is a gazda–parazitafaj–parazitaegyed terminológiára szorítkozunk, de minden megállapítás érvényes lesz az általános csoport–egyedtípus–egyed hierarchiára is.

Miként a 4.2. fejezetben, a gazda–parazita végtelen populációról most is feltételezzük, hogy benne véges számú, legfeljebb  $K$  parazitafaj fordulhat elő. Válasszunk ki véletlenszerűen és egymástól függetlenül  $N$  gazdát. A gazdák  $N$  elemű mintáján a parazita-abundanciák és relatív gyakoriságok elrendezését és jelölésrendszerét a 4.2. fejezet 1. táblázata mutatja. Ezekon kívül átvesszük a 4.2. fejezetben alkalmazott összes többi jelölést és lényegében változatlanul a gazdák parazitafaj-abundanciáit tartalmazó vektorokra tett valószínűségeloszlási feltevéseket is. Az egyetlen különbség, hogy a mostani fejezetben feltételezzük, hogy egy véletlenszerűen választott gazdán minden parazitafaj abundanciájának, mint valószínűségi változónak minden momentuma véges.

Legyen  $q > 0, q \neq 1$ . A gazdákra összesített  $p_{.j}, j = 1, \dots, K$  fajonkénti relatív gyakoriságokból a  $q$ -hoz tartozó effektív fajszámot, más néven valódi diverzitást a

$$\gamma = (p_{.1}^q + p_{.2}^q + \dots + p_{.K}^q)^{\frac{1}{1-q}}$$

képlettel számítjuk ki (Hill 1973, Jost 2006, 2007, Tuomisto 2010a). A gazdákon való összesített egyedszámokból kapott diverzitás éppen a parazitafajok  $\gamma$  diverzitása (Jost 2006, 2007, Tuomisto 2010a). Az  $i$ -ik gazdán a parazitafajok  $q$ -hoz tartozó valódi diverzitását a

$$\gamma_i = (p_{1|i}^q + p_{2|i}^q + \dots + p_{K|i}^q)^{\frac{1}{1-q}}$$

formula adja meg, ahol  $p_{j|i} = \frac{p_{ij}}{p_i} = \frac{x_{ij}}{x_i}, i = 1, \dots, N$  a  $j$ -ik faj relatív abundanciáját jelenti az  $i$ -edik gazdán.

A gazdánkénti valódi diverzitások súlyozott átlagát véve kapjuk az  $\alpha$ -diverzitást. A súlyok a paraziták gazdánkénti  $p_i$  relatív gyakoriságai. Az átlagolás  $1 - q$  kitevős hatványozással történik. Formulával kifejezve

$$\alpha = (p_1 \cdot \gamma_1^{1-q} + p_2 \cdot \gamma_2^{1-q} + \dots + p_N \cdot \gamma_N^{1-q})^{\frac{1}{1-q}}$$

(Tuomisto 2010a). Az átlagolás miatt  $\alpha$  mindig a  $\gamma_i$  gazdánkénti diverzitások minimuma és maximuma közé esik. Speciálisan, ha a  $\gamma_i$  diverzitások egymással egyenlők, akkor  $\alpha$  megegyezik ezzel a közös értékkel. Egyszerű számítás mutatja, hogy  $\alpha$  felírható

$$\alpha = \left( \sum_{i=1}^N \sum_{j=1}^K p_{ij} \cdot p_{j|i}^{q-1} \right)^{\frac{1}{1-q}}$$

és

$$\alpha = \left( \sum_{i=1}^N p_i \cdot \sum_{j=1}^K p_{j|i}^q \right)^{\frac{1}{1-q}}$$

alakban is, vagyis  $\alpha$  megegyezik a  $p_{j|i}$  feltételes relatív gyakoriságok  $q - 1$  kitevős súlyozott átlagának reciprokával (Tuomisto 2010a).

Megmutatható, hogy a most definiált  $\alpha$  diverzitás mindig kisebb, vagy egyenlő, mint a  $\gamma$  diverzitás. Egyenlőség pontosan akkor áll fenn, ha minden faj esetében a faj gazdánkénti relatív abundanciája minden gazdára ugyanaz az érték. Formálisan, az egyenlőség feltétele  $p_{j|i} = p_j, i = 1, \dots, N, j = 1, \dots, K$ .

Az  $\alpha \leq \gamma$  tulajdonságot Tuomisto (2010a) bizonyítás nélkül említi meg, ezért a bizonyítást itt röviden vázolólok. Legyen  $q > 1$ . A  $p_j = \sum_{i=1}^N p_i \cdot p_{j|i}$  egyenlőségből, az  $f(u) = u^q$  függvény konvexitásából és a Jensen-egyenlőtlenségből következik, hogy  $p_j^q \leq \sum_{i=1}^N p_i \cdot p_{j|i}^q$ . Ezt az egyenlőtlenséget  $j$ -re összegezve adódik az állítás. A  $q < 1$  esetre a bizonyítás hasonlóan történik.

Az  $\alpha \leq \gamma$  egyenlőtlenség következtében a  $\beta = \frac{\gamma}{\alpha}$  diverzitás értéke mindig legalább 1.

Az ismertetett valódi diverzitási formulák az

$$E_q(p_1, \dots, p_k) = \begin{cases} \frac{\ln \sum_{i=1}^k p_i^q}{1-q}, & q \neq 1 \\ -\sum_{i=1}^k p_i \cdot \ln p_i, & q = 1 \end{cases}$$

Rényi-féle általánosított entrópiákból exponenciális transzformációval származtathatók (Rényi 1961, Patil és Taillie 1982). Ezek  $q = 1$ -ben folytonosak,  $q \rightarrow 1$  esetén a határérték nem más, mint a Shannon diverzitáshoz tartozó effektív fajszám vagy más néven Shannon valódi diverzitás. Ezeket a diverzitásokat rendeljük a  $q = 1$  kitevőhöz. A  $\gamma$ -diverzitás  $q = 1$  mellett

$$\gamma = \exp(-p_1 \ln p_1 - p_2 \ln p_2 - \dots - p_K \ln p_K),$$

a valódi diverzitás az  $i$ -ik gazdán

$$\gamma_i = \exp(-p_{1|i} \ln p_{1|i} - p_{2|i} \ln p_{2|i} - \dots - p_{K|i} \ln p_{K|i}),$$

az  $\alpha$ -diverzitás pedig

$$\alpha = \exp\left(-\sum_{i=1}^N p_i \cdot \sum_{j=1}^K p_{j|i} \cdot \ln p_{j|i}\right),$$

vagy másképpen, a gazdánkénti diverzitások

$$\alpha = \exp\left(\sum_{i=1}^N p_i \cdot \ln \gamma_i\right)$$

súlyozott mértani átlagaként fejezhető ki.

A valódi diverzitás formuláinak  $q \downarrow 0$  esetén is van jobboldali határértéke. Legyenek ezek a határértékek definíció szerint a  $q = 0$ -hoz tartozó megfelelő diverzitások. A  $\gamma$ -diverzitás  $q = 0$  mellett a gazdákon előforduló parazitafajok száma,  $\gamma_i$  az  $i$ -ik gazdán élő parazitafajok száma, míg  $\alpha = \sum_{i=1}^N p_i \cdot \gamma_i$  ezek súlyozott aritmetikai átlaga. Az  $\alpha \leq \gamma$  összefüggés természetesen  $q = 1$  és  $q = 0$  esetén is igaz marad. A  $q = 0$ -hoz tartozó  $\alpha$ ,  $\gamma$  és  $\beta = \frac{\gamma}{\alpha}$  diverzitásokat egyszerűségük miatt gyakran alkalmazzák (Lande 1996, Jost 2006, 2007). Hátrányuk azonban, hogy a fajok egyedeinek gazdák közötti megoszlását csupán a fajok jelenléte vagy hiánya szempontjából veszik figyelembe.

A most bemutatott valódi diverzitási formulák a gazdák  $N$  elemű mintáira vonatkoznak. A parazitaabundanciák eloszlására tett feltevések miatt  $N \rightarrow \infty$  esetén az  $\alpha$  és  $\gamma$  diverzitások konvergálnak a megfelelő populációs diverzitásokhoz. Részletesebben, a nagy számok erős törvénye miatt,  $N \rightarrow \infty$  esetén a fajok gazdákra összesített relatív gyakorisága 1 valószínűséggel konvergens és  $\lim_{n \rightarrow \infty} p_j = \pi_j = \frac{EX_{1j}}{EX_1}$ ,  $j = 1, \dots, K$ . Ebből következik, hogy a  $\gamma$  diverzitásnak is van véges határértéke:

$$\lim_{N \rightarrow \infty} \gamma = (\pi_1^q + \pi_2^q + \dots + \pi_K^q)^{\frac{1}{1-q}} = \gamma_{\text{pop}}$$

A „pop” index arra utal, hogy a határérték a gazdák teljes populációjára vonatkozó  $\gamma$  diverzitást eredményezi.

A mintából számolt  $\alpha$  diverzitás

$$\alpha = \left\{ \frac{1}{\frac{X}{N}} \cdot \frac{1}{N} \sum_{i=1}^N X_i \cdot \sum_{j=1}^K \left( \frac{X_{ij}}{X_i} \right)^q \right\}^{\frac{1}{1-q}}$$

A nagy számok erős törvénye miatt ennek  $N \rightarrow \infty$  esetén 1 valószínűséggel létezik véges határértéke, mégpedig

$$\lim_{N \rightarrow \infty} \alpha = \left\{ \frac{1}{EX_1} \cdot E \left( X_1 \cdot \sum_{j=1}^K \left( \frac{X_{1j}}{X_1} \right)^q \right) \right\}^{\frac{1}{1-q}} = \alpha_{\text{pop}}$$

A gazdákra vett mintából becsült  $\alpha$  és  $\gamma$  valódi diverzitások véges határértékhez tartanak, ezért hányadosuk, a  $\beta$  diverzitás is véges határértékhez konvergál:

$$\lim_{N \rightarrow \infty} \frac{\gamma}{\alpha} = \frac{\gamma_{\text{pop}}}{\alpha_{\text{pop}}} = \beta_{\text{pop}}.$$

Az  $\alpha \leq \gamma$  egyenlőtlenségből következik, hogy  $\alpha_{\text{pop}} \leq \gamma_{\text{pop}}$  és  $\beta_{\text{pop}} \geq 1$ .

A  $q = 1$  esetben a  $\gamma$  diverzitás populációs értéke

$$\gamma_{\text{pop}} = \exp(-\pi_{.1} \ln \pi_{.1} - \pi_{.2} \ln \pi_{.2} - \dots - \pi_{.K} \ln \pi_{.K}),$$

az  $\alpha$  diverzitás populációs értéke

$$\alpha_{\text{pop}} = \exp\left(-\frac{1}{EX_{1.}} \cdot E\left(-X_{1.} \ln X_{1.} + \sum_{j=1}^K X_{1j} \ln X_{1j}\right)\right).$$

A  $q = 0$  esetben a  $\gamma$  diverzitás populációs értéke a teljes gazdapopuláción ténylegesen előforduló parazitafajok száma. Az  $\alpha$  diverzitás populációs értéke

$$\alpha_{\text{pop}} = \frac{1}{EX_{1.}} \cdot E\left(X_{1.} \sum_{j=1}^K I(X_{1j} > 0)\right),$$

ahol az események  $I$  indikátorfüggvényét az

$$I(A) = \begin{cases} 1, & \text{ha } A \text{ igaz} \\ 0, & \text{ha } A \text{ hamis} \end{cases}$$

képlet definiálja.

Ha  $\beta$  egész értékű, akkor jelentését szemléltethetjük  $\beta$  számú olyan gazdacsoporttal, melyeken ugyanannyi parazita található, minden gazdacsoportban az összesített diverzitás  $\alpha$  és a gazdacsoportok között nincs átfedés a parazitafajok tekintetében. Ennek az elképzelt közösségnek az  $\alpha, \beta$  és  $\gamma$  diverzitása megegyezik az eredetileg vizsgált közösség megfelelő diverzitásaival (Tuomisto 2010a). A 4.2. fejezetbeli 1. táblázat sémájának megfelelő abundanciastruktúrát  $\beta = 2$  mellett a 2. táblázat tartalmazza.

## 2. táblázat

Parazitafajok megoszlása gazdacsoportonként egyenlő abundancia és fajdiverzitás mellett

		Fajok				Gzadacsoportok súlya
		$S_1 \dots$	$S_t$	$S_{t+1} \dots$	$S_K$	
Gazda- csoportok	$HG_1$	$Z_{1,1}$	$Z_{1,t}$	$0$	$0$	$\frac{Z_{1.}}{Z_{..}} = \frac{1}{2}$
	$HG_2$	$0$	$0$	$Z_{2,t+1}$	$Z_{2,K}$	$\frac{Z_{2.}}{Z_{..}} = \frac{1}{2}$
Fajok súlya		$p_{.1} = \frac{Z_{1,1}}{Z_{..}}$	...	$p_{.t+1} = \frac{Z_{2,t+1}}{Z_{..}}$	...	

A 2. táblázatban  $Z_{i,j}$  az  $i$ -ik gazdacsoport  $j$ -ik parazitafajának összesített abundanciája,  $Z_{..}$  az abundanciák összege. A diverzitás mindkét gazdacsoportban  $\alpha$ , ezért ennek a rendszernek az  $\alpha$  diverzitása is  $\alpha$ . A rendszer  $\gamma$  diverzitására

$$\gamma^{1-q} = p_{.1}^q + \dots + p_{.t}^q + p_{.t+1}^q + \dots + p_{.K}^q = 2 \cdot 2^{-q} \cdot \alpha^{1-q} = 2^{1-q} \cdot \alpha^{1-q},$$

vagyis valóban  $\gamma = 2 \cdot \alpha = \beta \cdot \alpha$ .

A  $\beta$  diverzitás jelentésének előbbi érzékeltetése a populációk szintjén nincs ellentmondásban azzal, hogy az  $X_j = (X_{j1}, X_{j2}, \dots, X_{jK})^T$  parazitafaj-abundancia vektorok egymástól független azonos eloszlású valószínűségi vektorváltozók. Képzeljünk el  $u_i$   $\beta$  számú, közös fajjal páronként nem rendelkező olyan parazitaközösséget, melyek mindegyikének  $\alpha$  a diverzitása. Tegyük fel, hogy ezek valamely előre rögzített egyedszám-kombinációban a gazdákon  $\frac{1}{\beta}$  valószínűséggel telepednek meg oly módon, hogy egy-egy gazdát csupán az egyik parazitaközösség fertőzheti meg, vagyis a közösségek nem tűrik el egymást. Független, véletlen mintavételezéssel  $N$  gazdát kiválasztva, a gazdák parazitáinak fajabundancia-vektorai egymástól független, ugyanolyan eloszlású valószínűségi vektorváltozók lesznek. A mintába került gazdákból a fertőző fajközösség szerint  $\beta$  számú gazdacsoport alakul ki. Ha az  $N$  gazdaszám végtelenbe tart, akkor mindegyik gazdacsoportban ugyanahhoz az  $\alpha_{\text{pop}}$  értékhez tart a valódi diverzitás, az egyesített minta gazdaközösségének valódi diverzitása pedig a  $\gamma_{\text{pop}} = \beta \cdot \alpha_{\text{pop}}$  értékhez konvergál.

A fenti szemléltetési módok nem fejezik ki teljes egészében a  $\beta$  diverzitás jelentését, hiszen számtalan különböző abundancia-szerkezetű gazda–parazita közösség van, amelyekhez az általunk vizsgált közösséggel megegyező  $\alpha$ ,  $\beta$  és  $\gamma$  diverzitások tartoznak.

A paraziták zsúfoltságával, általában csoportok méretével kapcsolatos indexeknek a valódi diverzitáshoz hasonló multiplikatív felbontása érdekében cseréljük fel a gazdák és a parazitafajok szerepét. A zsúfoltság skálafüggvényének a 4.1. fejezetben bevezetett  $f_q$  függvénycsaládot választjuk. A gazdák „valódi diverzitása” az effektív gazdaszám, ami – mint azt a 4.1. fejezetben láttuk – az ugyanolyan mértékben parazitált gazdák számát jelenti egy olyan elképzelt gazda–parazita közösségben, ahol a zsúfoltság megegyezik az eredetileg vizsgált adatokból számolt zsúfoltsággal. Az effektív gazdaszám a gazdák  $N$  számának növekedésekor  $N$ -nel arányosan nő, ezért az effektív gazdaszám helyett az aggregáltsági indexekre adunk meg a valódi diverzitás multiplikatív felbontásával analóg felbontást.

A számításokhoz nem kell mást tennünk, mint a 4.2. fejezet 1. táblázatában felcserélni a sorok és az oszlopok szerepét. A 4.1. fejezet effektív gazdaszám formuláját az  $f_q$  függvényre alkalmazva, a gazdákon élő parazita-kolóniák fajonként összesített abundanciáiból nyerjük a  $\gamma$ -effektív gazdaszám becslését:

$$\gamma^* = (p_{1.}^q + p_{2.}^q + \dots + p_{N.}^q)^{\frac{1}{1-q}}.$$

Az effektív gazdaszámot a  $j$ -ik parazitafaj vonatkozásában a

$$\gamma_j^* = (p_{1|j}^q + p_{2|j}^q + \dots + p_{N|j}^q)^{\frac{1}{1-q}}$$

képlet adja,  $j = 1, \dots, K$ .

Vegyük észre a valódi diverzitás és az effektív gazdaszám formulái között meglévő formai hasonlóságot. Egyik a másikba átírható az  $i$  és  $j$  indexek, valamint  $N$  és  $K$  felcserélésével. Erre építve alakítjuk ki a még szükséges fogalmak definícióit.

A fajonkénti effektív gazdaszám  $1 - q$  kitevős súlyozott átlagával értelmezzük az  $\alpha$ -effektív gazdaszámot:

$$\alpha^* = (p_{.1} \cdot \gamma_1^{*(1-q)} + p_{.2} \cdot \gamma_2^{*(1-q)} + \dots + p_{.K} \cdot \gamma_K^{*(1-q)})^{\frac{1}{1-q}}$$

Ebből egyszerű algebrai átalakítással kapjuk, hogy

$$\alpha^* = \left( \sum_{j=1}^K \sum_{i=1}^N p_{ij} \cdot p_{i|j}^{q-1} \right)^{\frac{1}{1-q}}$$

és

$$\alpha^* = \left( \sum_{j=1}^K p_{.j} \sum_{i=1}^N p_{i|j}^q \right)^{\frac{1}{1-q}}.$$

A  $\beta$  diverzitásnak megfelelő index a  $\gamma$ -effektív gazdaszám és az  $\alpha$ -effektív gazdaszám hányadosa:  $\beta^* = \frac{\gamma^*}{\alpha^*}$ . Tuomisto (2010a) alapján  $\alpha^* \leq \gamma^*$ , így  $\beta^* \geq 1$ .

A 4.1. fejezetben láttuk, hogy az effektív gazdaszámok az aggregáltsági indexekkel hozhatók kapcsolatba. Értelmezzük a  $\gamma$ -aggregáltságot az  $A_\gamma = \frac{N}{\gamma^*}$ , az  $\alpha$ -aggregáltságot az  $A_\alpha = \frac{N}{\alpha^*}$  formulákkal. Azt kapjuk, hogy

$$A_\gamma = \frac{A_\alpha}{\beta^*},$$

azaz a  $\gamma$ -aggregáltság kifejezhető, mint az  $\alpha$ -aggregáltság és a  $\beta^*$  index hányadosa. A továbbiakban a  $\beta^*$  indexet  $\beta$ -aggregáltságnak nevezzük és  $A_\beta$ -val jelöljük. Az  $\alpha^* \leq \gamma^*$  egyenlőtlenségből  $A_\gamma \leq A_\alpha$  következik, ami matematikailag azt a tapasztalattal összhangban álló jelenséget fejezi ki, hogy több parazitafaj együttesen nem lehet aggregáltabb, mint fajonként külön-külön.

Helyettesítsük be az abundanciákat a kapott összefüggésekbe. A  $\gamma$ -effektív gazdaszámra

$$\gamma^* = N \cdot \left\{ \left( \frac{X_{\cdot}}{N} \right)^{-q} \cdot \frac{1}{N} \sum_{i=1}^N X_i^q \right\}^{\frac{1}{1-q}}.$$

Ha mintában a gazdák  $N$  száma tart a végtelenhez, akkor a  $\gamma$ -effektív gazdaszám is végtelenhez tart, azonban a  $\gamma$ -aggregáltság véges határértékhez konvergál, mely a populáció szintű  $\gamma$ -aggregáltság:

$$A_{\gamma, \text{pop}} = \lim_{N \rightarrow \infty} \frac{N}{\gamma^*} = \left\{ \frac{(EX_1)^q}{E(X_1^q)} \right\}^{\frac{1}{1-q}}.$$

Ha  $N \rightarrow \infty$ , akkor az  $\alpha$ -effektív gazdaszám

$$\alpha^* = N \cdot \left\{ \sum_{j=1}^K \frac{X_j}{X_{\cdot}} \cdot \left( \frac{X_j}{N} \right)^{-q} \cdot \frac{1}{N} \sum_{i=1}^N X_{ij}^q \right\}^{\frac{1}{1-q}}$$

is tart a végtelenbe, de az  $\alpha$ -aggregáltság véges értékhez, a populáció szintű

$$A_{\alpha, \text{pop}} = \lim_{N \rightarrow \infty} \frac{N}{\alpha^*} = \left\{ \sum_{j=1}^K \frac{EX_{1j}}{EX_1} \cdot \frac{E(X_{1j}^q)}{(EX_{1j})^q} \right\}^{\frac{1}{1-q}}$$

$\alpha$ -aggregáltsághoz konvergál. Az előbbieket alapján, ha  $N \rightarrow \infty$ , akkor a  $\beta$ -aggregáltság véges határértékhez tart, melynek értéke

$$A_{\beta, \text{pop}} = \frac{A_{\alpha, \text{pop}}}{A_{\gamma, \text{pop}}}.$$

A  $\gamma$ -effektív gazdaszám a  $q = 1$  melletti logaritmikus esetben

$$\gamma^* = \exp(-p_1 \ln p_1 - p_2 \ln p_2 - \dots - p_N \ln p_N),$$

az effektív gazdaszám a  $j$ -ik parazitafaj tekintetében

$$\gamma_j^* = \exp(-p_{1|j} \ln p_{1|j} - p_{2|j} \ln p_{2|j} - \dots - p_{N|j} \ln p_{N|j}),$$

az  $\alpha$ -effektív gazdaszám pedig

$$\alpha^* = \exp\left(-\sum_{j=1}^K p_j \sum_{i=1}^N p_{i|j} \cdot \ln p_{i|j}\right),$$

vagy másképpen, a fajonkénti effektív gazdaszámok

$$\alpha^* = \exp\left(\sum_{j=1}^K p_j \cdot \ln \gamma_j^*\right)$$

súlyozott mértani átlagaként fejezhető ki. A logaritmikus  $\beta$ -aggregáltság most is  $A_\beta = \frac{\gamma^*}{\alpha^*}$ . A formulák a már tárgyalt  $q \neq 1$  feltétel melletti képletekből  $q \rightarrow 1$  határátmenettel kaphatók, ennek következtében itt is  $\alpha^* \leq \gamma^*$  és  $A_\beta \geq 1$ .

Értelmezzük ebben az esetben is a logaritmikus  $\gamma$ -aggregáltságot az  $A_\gamma = \frac{N}{\gamma^*}$ , az  $\alpha$ -aggregáltságot az  $A_\alpha = \frac{N}{\alpha^*}$  formulákkal. Egyszerű számítás mutatja, hogy  $N \rightarrow \infty$  esetén a logaritmikus  $\gamma$ -aggregáltság az

$$A_{\gamma, \text{pop}} = \frac{\exp\left(\frac{E(X_1 \cdot \ln X_1)}{EX_1}\right)}{EX_1}$$

véges populációs  $\gamma$ -aggregáltsághoz, a logaritmikus  $\alpha$ -aggregáltság az

$$A_{\alpha, \text{pop}} = \exp\left(\frac{1}{EX_1} \cdot \sum_{j=1}^K \{E(X_{1j} \cdot \ln X_{1j}) - E(X_{1j}) \cdot \ln E(X_{1j})\}\right)$$

véges populációs  $\alpha$ -aggregáltsághoz konvergál. A kettő hányadosa a populációs  $\beta$ -aggregáltság,  $A_{\beta, \text{pop}} = \frac{A_{\alpha, \text{pop}}}{A_{\gamma, \text{pop}}}$ , ami szintén véges.

A  $\gamma$ -effektív gazdaszám a  $q = 0$  melletti hiperbolikus esetben  $\gamma^* = N^+$ , a mintában lévő parazitált gazdák száma. Az effektív gazdaszám a  $j$ -ik parazitafaj tekintetében  $\gamma_j^* = N^{+j}$ , a mintában lévő,  $j$ -ik fajjal parazitált gazdák száma. A hiperbolikus  $\alpha$ -effektív gazdaszám a fajonkénti effektív gazdaszámok  $\alpha^* = \sum_{j=1}^K p_j \cdot \gamma_j^*$  súlyozott átlagaként fejezhető ki. A hiperbolikus  $\beta$ -aggregáltság  $A_\beta = \frac{\gamma^*}{\alpha^*} = \frac{N^+}{\sum_{j=1}^K p_j \cdot N^{+j}}$ .

A formulák a már tárgyalt képletekből  $q \downarrow 0$  határátmenettel kaphatók, ennek következtében itt is  $\alpha^* \leq \gamma^*$  és  $A_\beta \geq 1$ . A  $\gamma$ -aggregáltság  $A_\gamma = \frac{N}{N^+} = \frac{1}{\hat{P}}$ , ahol  $\hat{P}$  a nem fajspecifikus parazitáltság becsült prevalenciája. Az  $\alpha$ -aggregáltság  $A_\alpha = \frac{N}{\alpha^*} = \frac{N}{\sum_{j=1}^K p_j \cdot N^{+j}}$ . A  $\beta$ -aggregáltságot az  $\alpha$ - és  $\gamma$ -aggregáltsággal kifejezve, az

$$A_\beta = \frac{A_\alpha}{A_\gamma} = \frac{\hat{P}}{\sum_{j=1}^K p_j \cdot \hat{P}_j}$$



képletet nyerjük, ahol  $\hat{P}_j$  a gazdáknak a  $j$ -ik parazitafajjal való fertőzöttségre becsült prevalenciája. Ez alapján megállapítható, hogy a hiperbolikus  $\beta$ -aggregáltság a fajspecifikus és nem fajspecifikus parazitáltság prevalenciájától, valamint a parazitafajok összes gazdára számolt részarányától függ.

A nagy számok erős törvénye miatt  $N \rightarrow \infty$  esetén a hiperbolikus  $\gamma$ -aggregáltság  $A_{\gamma, \text{pop}} = \frac{1}{P}$ , az  $\alpha$ -aggregáltság  $A_{\alpha, \text{pop}} = \frac{1}{\sum_{j=1}^K \pi_j \cdot P_j}$  véges populációs értékekhez tart, ahol  $P_j$  a gazdák  $j$ -ik parazitafajjal való fertőzöttségének populációs prevalenciája,  $P$  pedig a nem fajspecifikus parazitáltság populációs prevalenciája. A hiperbolikus populációs  $\beta$ -aggregáltság

$$A_{\beta, \text{pop}} = \frac{A_{\alpha, \text{pop}}}{A_{\gamma, \text{pop}}} = \frac{P}{\sum_{j=1}^K \pi_j \cdot P_j}$$

A  $\beta$ -aggregáltság  $A_{\beta}$  értékének jelentése érzékeltethető olyan, egymást közös gazdán nem elviselő parazita fajcsoportok számával, melyeknek ugyanannyi egyedük van, mindegyik fajcsoportnak  $A_{\alpha}$  az  $\alpha$ -aggregáltsága, az együttes, nem fajspecifikus aggregáltságuk pedig  $A_{\gamma}$ . Illusztrációként az  $A_{\beta} = 2$  esetre a 3. táblázat két egymást kizáró, ugyanannyi egyedet számláló, ugyanakkora  $A_{\alpha}$   $\alpha$ -aggregáltságú fajcsoport gazdánkénti abundanciáit tartalmazza.

### 3. táblázat

*Parazitafajok megoszlása fajcsoportonként egyenlő abundancia és  $\alpha$ -aggregáltság mellett*

	Fajcsoportok		Gazdák súlya	
	$SG_1$	$SG_2$		
Gazdák	$H_1$	$Y_{1,1}$	$0$	$p_{1.} = \frac{Y_{1,1}}{Y_{..}}$
	...			...
	$H_n$	$Y_{n,1}$	$0$	$p_{n.} = \frac{Y_{n,1}}{Y_{..}}$
	$H_{n+1}$	$0$	$Y_{n+1,2}$	$p_{n+1.} = \frac{Y_{n+1,2}}{Y_{..}}$
	...			...
	$H_N$	$0$	$Y_{N,2}$	$p_{N.} = \frac{Y_{N,2}}{Y_{..}}$
Fajcsoportok súlya	$p_{.1} = \frac{1}{2}$	$p_{.2} = \frac{1}{2}$		

A 3. táblázatban  $Y_{i,j}$  az  $i$ -ik gazda  $j$ -ik parazitafaj-csoportjának összesített abundanciája,  $Y_{i\cdot}$  az abundanciák összege. Az  $\alpha$ -aggregáltság mindkét fajcsoportban  $A_\alpha$ , ezért a rendszer  $\alpha$ -aggregáltsága is  $A_\alpha$ . A rendszer  $\gamma$ -effektív gazdaszámára

$$\gamma^{*1-q} = p_1^q + \dots + p_n^q + p_{n+1}^q + \dots + p_N^q = 2^{1-q} \cdot \alpha^{*1-q},$$

amiből  $\gamma^* = 2 \cdot \alpha^* = A_\beta \cdot \alpha^*$  és  $A_\beta = \frac{A_\alpha}{A_\gamma}$ .

A  $\beta$ -aggregáltság előbbi szemléltetése nem mond ellent a gazdánkenti fajabundancia-vektorok függetlenségének és azonos eloszlásának. Képzeljünk el  $u_i$   $A_\beta$  számú, páronként diszjunkt, közös fajt nem tartalmazó parazita fajcsoportot, melyek mindegyikének ugyanaz az  $A_\alpha$  érték az  $\alpha$ -aggregáltsága. Tegyük fel, hogy ezek a fajcsoportok a gazdákon  $\frac{1}{A_\beta}$  valószínűséggel telepednek meg, de úgy, hogy egymás jelenlétét nem viselik el és megtelepedésük esetén egyedszámaik adott eloszlásúak. Ebben a modellben a gazdák parazitáinak abundanciavektorai egymástól független, azonos eloszlású valószínűségi vektorváltozók, a gazdák  $N$  számának növelésével pedig határértékben, populáció szinten  $A_{\beta,\text{pop}} = \frac{A_{\alpha,\text{pop}}}{A_{\gamma,\text{pop}}}$ .

A diverzitási és aggregáltsági indexek a gazda–parazita közösségeket eltérő szemlélettel mutatják be, amelyek azonban egymást kiegészíthetik. Mindkét felbontást érdemes felhasználni a közösség minél árnyaltabb bemutatása érdekében.

Az  $\alpha$  és  $\gamma$  diverzitások és aggregáltságok általában különböznek és nem állnak szoros kapcsolatban. Mit mondhatunk a  $\beta$  diverzitás és a  $\beta$ -aggregáltság kapcsolatáról? Ha  $q = 1$ , tehát a Shannon diverzitásról, illetve a logaritmusos zsúfoltságról van szó, akkor egyszerű számolással igazolható, hogy  $\beta = A_\beta$ . Tegyük fel, hogy  $q \neq 1$  és vizsgáljunk meg egy olyan egyszerű felépítésű gazda–parazita közösséget, amelyben két különböző egyedszámú parazitafaj fordul elő. Tegyük fel, hogy a parazitafajok nem túrik el egymást ugyanazon a gazdán. Az abundanciák szerkezetét a 4. táblázat adja meg.

#### 4. táblázat

Két eltérő egyedszámú parazitafaj abundancia-megoszlása

	Fajok		Gazdák súlya	
	$S_1$	$S_2$		
Gazdák	$H_1$	$U_{1,1}$	$0$	$p_{1.} = \frac{U_{1,1}}{U_{..}}$
	...			...
	$H_r$	$U_{r,1}$	$0$	$p_{r.} = \frac{U_{r,1}}{U_{..}}$
	$H_{r+1}$	$0$	$U_{r+1,2}$	$p_{r+1.} = \frac{U_{r+1,2}}{U_{..}}$
	...			...
	$H_N$	$0$	$U_{N,2}$	$p_{N.} = \frac{U_{N,2}}{U_{..}}$
Fajok súlya	$p_{.1} = w$	$p_{.2}$ $= 1 - w$		

A 4. táblázatban  $U_{i,j}$  az  $i$ -ik gazda  $j$ -ik parazitafajának abundanciája,  $U_{..}$  az abundanciák összege.

A fajokhoz tartozó effektív gazdaszámok

$$\gamma_1^* = \left\{ \frac{1}{w^q} (p_{1.}^q + \dots + p_{r.}^q) \right\}^{\frac{1}{1-q}}$$

és

$$\gamma_2^* = \left\{ \frac{1}{(1-w)^q} (p_{r+1.}^q + \dots + p_{N.}^q) \right\}^{\frac{1}{1-q}}.$$

Az  $\alpha$ -effektív gazdaszám

$$\alpha^* = (w \cdot \gamma_1^{*1-q} + (1-w) \cdot \gamma_2^{*1-q})^{\frac{1}{1-q}},$$

a  $\gamma$ -effektív gazdaszám

$$\gamma^* = (p_{1.}^q + \dots + p_{r.}^q + p_{r+1.}^q + \dots + p_{N.}^q)^{\frac{1}{1-q}} = (w^q \cdot \gamma_1^{*1-q} + (1-w)^q \cdot \gamma_2^{*1-q})^{\frac{1}{1-q}}.$$

Ebből a  $\beta$ -aggregáltság

$$A_\beta = \left\{ \frac{w^q \cdot \gamma_1^{*1-q} + (1-w)^q \cdot \gamma_2^{*1-q}}{w \cdot \gamma_1^{*1-q} + (1-w) \cdot \gamma_2^{*1-q}} \right\}^{\frac{1}{1-q}}.$$

A fajok száma minden gazdán 1, ezért az  $\alpha$ -diverzitás 1. A  $\gamma$ -diverzitás

$$\gamma = (w^q + (1 - w)^q)^{\frac{1}{1-q}},$$

ugyanennyi a  $\beta$ -diverzitás is,

$$\beta = \frac{\gamma}{\alpha} = (w^q + (1 - w)^q)^{\frac{1}{1-q}}.$$

Ha  $\gamma_1^* = \gamma_2^*$  vagy  $w = \frac{1}{2}$ , akkor  $A_\beta = \beta$ . Egyéb esetekben  $A_\beta$  nem feltétlenül egyezik meg  $\beta$ -val, ugyanis  $\beta$  nem függ a  $\gamma_1^*$  és  $\gamma_2^*$  fajonkénti effektív gazdaszámoktól. Általában  $A_\beta$  az  $\frac{1}{w}$  és  $\frac{1}{1-w}$  között bármilyen értéket felvehet, a  $\frac{\gamma_1^*}{\gamma_2^*}$  aránytól függően.

Az illusztratív példák mutatják, hogy az  $\alpha$ ,  $\beta$ ,  $\gamma$  diverzitások és az  $A_\alpha$ ,  $A_\beta$ ,  $A_\gamma$  aggregáltságok értelmezése egyszerűbbé válik, ha a gazda–parazita párok abundanciaszerkezete blokkokba rendeződik, azaz az abundanciák táblázatában, mátrixában az abundanciák nagy része csak néhány, a fődiagonális mentén lévő blokkban található. Vegyük észre, hogy sem az  $\alpha$ ,  $\beta$ ,  $\gamma$  diverzitások, sem az  $A_\alpha$ ,  $A_\beta$ ,  $A_\gamma$  aggregáltságok nem változnak, ha az abundanciák táblázatában két sort vagy két oszlopot felcserélünk. A sorok és az oszlopok többszöri permutálásával az abundanciák táblázatát sokszor sikerül közelítőleg blokkos szerkezetűvé alakítani. Erre szolgálnak az ún. blokk-klaszterező eljárások (block clustering methods, Govaert és Nadif 2010). A blokk-klaszterezés az R 3.0.2 statisztikai szoftver *blockcluster* csomagjával valósítható meg (R Core Team 2014, Bhatia et al. 2014).

Az abundanciablokkok megtalálását, a diverzitások és aggregáltságok értelmezését segítheti, ha azonosítjuk azokat a gazda–parazita párokat, melyek együttes előfordulási gyakoriságai pozitív asszociációt fejeznek ki, azaz nagyobbak annál, mintha a társulás teljesen véletlenszerű lenne. Korrespondencia analízissel a gazdákat és a parazitákat a síkon vagy a térben olyan közös koordinátarendszerben ábrázolhatjuk, amelyben a pozitív kapcsolatban álló gazdák és paraziták pontjai egymáshoz közel kerülnek (Greenacre 2007, Bolla 1987, Reiczigel 1988). Az R 3.0.2 szoftver *anacor* és *ca* csomagjával végezhetünk korrespondencia analízist (R Core Team 2014, Leeuw és Mair 2009, Nenadic és Greenacre 2007).

A  $\beta$ -diverzitás és a  $\beta$ -aggregáltság alkalmazásához fontos tudni, hogy a  $q$  kitevő hogyan befolyásolja értéküket. A  $\beta$ -diverzitás felírható

$$\beta = \frac{\gamma}{\alpha} = \frac{\left(\sum_{i=1}^N \sum_{j=1}^K p_{ij} \cdot p_{ji}^{q-1}\right)^{\frac{1}{q-1}}}{\left(\sum_{j=1}^K p_{.j} \cdot p_j^{q-1}\right)^{\frac{1}{q-1}}}$$

alakban (Tuomisto 2010a). A nevező a parazitafajok részarányainak  $q - 1$  kitevős súlyozott átlaga, a számláló a parazitafajok gazdánkénti részarányainak  $q - 1$  kitevős súlyozott átlaga. Ha  $q > 1$ , akkor a  $\beta$ -diverzitás értékének alakulásában a nagyobb részarányú parazitafajok és azok a parazitafajok, melyek részaránya valamelyik gazdán a többi fajhoz képest

nagyobb, abundancia-arányukhoz képest nagyobb súllyal szerepelnek. Ha  $q < 1$ , akkor az alacsonyabb részarányú parazitafajok és azok a parazitafajok, melyek részaránya valamelyik gazdán a többi fajhoz képest kisebb, a  $\beta$ -diverzitás értékét abundancia-arányukhoz képest relatíve jobban befolyásolják.

A  $\beta$ -aggregáltság az

$$A_{\beta} = \frac{A_{\alpha}}{A_{\gamma}} = \frac{\left(\sum_{i=1}^N \sum_{j=1}^K p_{ij} \cdot p_{ij}^{q-1}\right)^{\frac{1}{q-1}}}{\left(\sum_{i=1}^N p_i \cdot p_i^{q-1}\right)^{\frac{1}{q-1}}}$$

formulával fejezhető ki. A nevező a paraziták gazdák közötti részarányainak  $q - 1$  kitevős súlyozott átlaga, a számláló egy-egy parazitafaj parazitái gazdák közötti részarányainak  $q - 1$  kitevős súlyozott átlaga. Ha  $q > 1$ , akkor a  $\beta$ -aggregáltság értékére parazitáltságukhoz képest nagyobb hatással vannak azok a gazdák, melyek parazitáltsága nagy, vagy valamelyik parazitafaj a többi gazdához képest nagyobb arányban fordul elő rajtuk. Ha  $q < 1$ , akkor azok a gazdák, melyek fajspecifikus vagy nem fajspecifikus parazitáltsága a többi gazdához képest kisebb, parazitaabundancia-arányukhoz viszonyítva relatíve nagyobb szerepet kapnak a  $\beta$ -aggregáltság értékének alakulásában.

A  $q = 1$  esetben láttuk, hogy a Shannon  $\beta$ -diverzitás és a logaritmikus  $\beta$ -aggregáltság értéke megegyezik. Ezt a közös értéket a parazitafajok fajspecifikus és nem fajspecifikus abundanciaarányai mind a gazdák, mind a paraziták körében a részarányokkal éppen megegyező mértékben befolyásolják. A parazita-részarányok hatása tekintetében tehát a  $\beta$ -diverzitás és a vele egyező logaritmikus  $\beta$ -aggregáltság a legkiegyensúlyozottabb, kiszámítása, alkalmazása mindenképpen javasolt. A  $q = 0$  esethez tartozó  $\beta$ -diverzitást csak a fajok jelenléte vagy hiánya és a paraziták gazdák közötti részaránya befolyásolja. A hiperbolikus  $\beta$ -aggregáltság esetében pedig csak a paraziták fajonkénti részaránya és a parazitáknak a gazdákon való jelenléte vagy hiánya számít, a gazdánkénti részarány mértéke nem játszik szerepet. A  $q = 2$  esetben a nagyobb parazita-részarányok szerepe nagyobb. A parazita-részarányok hatásának vizsgálatához érdemes mindhárom kitevővel kiszámolni a  $\beta$ -diverzitást és a  $\beta$ -aggregáltságot.

## 5 Egyedekből és ezek diszjunkt csoportjaiból álló populációk valószínűségelméleti modellezése

Csoportokba rendeződő egyedek biológiai közösségének vizsgálatakor felmerül a kérdés, vajon a csoportok vagy az egyedek legyenek a megfigyelési egységek. A csoport szintű megközelítés elsősorban a csoportok jellemzőit hangsúlyozza, az egyedek tulajdonságait

azok csoportonkénti összesítésével veszi figyelembe, az egyedeket a csoportot alkotó komponenseknek tekintve. Ha a megfigyelés egysége az egyed, akkor ezek populációs jellemzőin, átlagain, szórásain, eloszlásain kívül főként annak a csoportnak a sajátosságai lehetnek fontosak, amelyikhez az egyed tartozik. (Ezért az egyedszempon-tú skála megfigyelési egysége valójában az egyed és az egyedet tartalmazó csoport által alkotott páros.) A kétféle skála eltérő szemléletnek felel meg, olyannyira, hogy ugyanahhoz az átlaghoz, arányhoz, valószínűséghez általában numerikusan eltérő érték tartozik. Például a csoportok átlagos egyedszáma különbözik a csoportok egyedenként átlagolt egyedszámától, ami az egyedek lineáris zsúfoltsága, átlagos csoportmérete. A zsúfoltság általában nagyobb, mint az átlagos egyedszám (Reiczigel et al. 2005, 2008).

Tételezzük fel az egyszerűség kedvéért, hogy a közösség véges  $N$  számú egyedből áll, melyek összesen  $K$  csoportba tartoznak. A véges egyedszám és csoportszám feltételezése csak a tárgyalásmódot egyszerűsíti, valójában nem szükségszerű, mert az alább megfogalmazásra kerülő állításokban a populációs átlagok, várható értékek és valószínűségek nem függenek  $N$  és  $K$  értékétől, így határértékben érvényesek lesznek végtelen sok egyedből és csoportból álló elméleti közösségekre is. Feltesszük, hogy minden egyed pontosan egy csoporthoz, a gazdacsoportjához tartozik. Ez azt jelenti, hogy a csoportok az egyedek populációjának egymásba át nem nyúló, a teljes populációt lefedő partícióját alkotják. Megengedjük ugyanakkor, hogy egyes csoportokhoz ne tartozzon egyed. Gazda-parazita közösségekben ugyanis a csoport a gazdaállatot jelentheti, ha a gazdának megfeleltetjük a rajta élő paraziták összességét. Ilyen közösségekben előfordulhat, hogy a gazdák egy részén nincs parazita, a nekik megfelelő csoportokban nincs egyed.

Válasszunk ki az  $N$  egyed és a  $K$  csoport közül egyet-egyet egymástól függetlenül, egyenletes eloszlás szerint. A kiválasztott egyed és csoport legyen  $\omega$  ill.  $G$ . Formálisan a kiválasztás valószínűsége  $P(\omega, G) = \frac{1}{N \cdot K}$ . Legyen  $X = |G|$  a kiválasztott csoport egyedszáma. A közösség egyedszempon-tú eseményei és valószínűségei legyenek *definíció szerint* az  $\omega \in G$  eseményre vett feltételes események, valószínűségek. Ez a feltétel azt jelenti, hogy az egyeddel és csoportkörnyezetével kapcsolatos, szubjektív eseményeket vizsgálunk. A közösség általános, feltétel nélküli eseményei nem kapcsolják össze az egyedeket a csoportjukkal. Az egyedeket kiintegrálva csoportszintű, a csoportokat kiintegrálva egyedszintű marginális eloszlásokhoz jutunk.

Az  $\omega \in G$  feltételnek megfelelő valószínűségi mezőben az elemi események olyan  $(u, g)$  elem-csoport párok, ahol  $u \in g$ . A feltételes valószínűségi mérték egyenletes, azaz minden  $(u, g)$  elemi eseménynek ugyanakkora a valószínűsége, ui. a feltétel nélküli valószínűségi mezőben is egyenletes a valószínűségi mérték. Minden  $u$  egyedhez pontosan egy olyan  $g$

csoport létezik, amelyre  $u \in g$ , így a feltételes egyenletes valószínűség éppen az egyedek összességére vett, feltétel nélküli egyenletes valószínűségi mértékkel egyezik meg. Ez a tulajdonság fejezi ki azt, hogy az események, valószínűségek egyedszemponitúak. Legyen  $A$  tetszőleges esemény a feltételes valószínűségi mezőben, azaz álljon olyan  $(u, g)$  elemi eseményekből, ahol  $u \in g$ . A  $P(A|\omega \in G)$  feltételes valószínűséget becsülhetjük az egyedek közül visszatevéssel, egymástól függetlenül, ugyanakkora valószínűséggel kiválasztott  $u_1, u_2, \dots, u_n$  mintából az  $A$  esemény relatív gyakoriságával. A nagy számok Bernoulli-féle törvénye szerint a relatív gyakoriság az elméleti egyedszemponitú  $P(A|\omega \in G)$  valószínűséghez tart, ha  $n \rightarrow \infty$ .

Legyen  $Y$  csoportszintű véletlen vektormező, azaz  $Y(g)$  minden  $g$  csoport esetében valószínűségi vektorváltozó. Az  $Y(G)$  valószínűségi vektorváltozó eloszlását fogjuk tanulmányozni az  $\omega \in G$  eseményre feltételesen és e feltétel nélkül. A továbbiak során feltesszük, hogy  $\omega$ ,  $G$  és  $Y$  egymástól teljesen függetlenek. Az  $Y(g)$  lehet például a  $g$  csoportból visszatevés nélkül kisorsolt  $n$  egyed körében a hímek száma, ha  $|g| \geq n$ , különben 0.  $Y(g)$  a csoportszintű adatok mellett kiegészíthető a  $g$  csoporthoz tartozó egyedek jellemzőit magába foglaló vektorral is, ilyenkor  $Y(g)$  dimenziója függhet a  $g$  csoporttól. Ebben az esetben  $Y(g)$  csoportszinten összegyűjti, információvesztés nélkül tárolja a  $g$  csoport egyedeihez tartozó adatokat is.

A példák érzékeltetik, hogy míg  $Y$  az egyed–csoport közösséget leíró adatok, tényezők, magyarázó változók vektora, az  $\omega$  megfigyelő egyedet és  $G$  tesztcsoportot kizárólag a matematikai modellezés céljából, a közösség jellemzőitől függetlenül értelmezzük. Ezért tesszük fel, hogy  $\omega$ ,  $G$  és  $Y$  egymástól teljesen függetlenek.

Elsőként az  $X = |G|$  változó kétféle szemlélet melletti eloszlását hasonlítjuk össze. A fejezet során végig feltételezzük, hogy  $P(X > 0) \neq 0$ , vagyis a közösségben pozitív azoknak a csoportoknak az aránya, amelyekben van egyed.

**1. Állítás:**  $P(X = x|\omega \in G) = P(X = x) \cdot \frac{x}{EX}$ . (1)

**Bizonyítás:** Az  $X = x$  és  $\omega \in G$  események együttes valószínűsége

$$P(\omega \in G, |G| = x) = \sum_{k=1}^K \frac{1}{K} \cdot P(\omega \in g_k, |g_k| = x | G = g_k).$$

Az  $\omega$  és  $G$  függetlensége miatt a jobboldalon a feltételek elhagyhatók. Az összegben csak azok a tagok lehetnek pozitívak, melyekre  $|g_k| = x$  teljesül. Így

$$P(\omega \in G, |G| = x) = \sum_{k: |g_k|=x} \frac{1}{K} \cdot \frac{x}{N}.$$

Másrészt

$$P(|G| = x) = \sum_{k=1}^K \frac{1}{K} \cdot P(|g_k| = x | G = g_k) = \sum_{k: |g_k|=x} \frac{1}{K},$$

hiszen a  $|g_k| = x$  esemény vagy a biztos, vagy a lehetetlen esemény, így független a  $G = g_k$  eseménytől. Az eddigiek alapján megállapíthatjuk, hogy

$$P(\omega \in G | X = x) = \frac{x}{N}.$$

Ebből

$$P(\omega \in G) = \sum_{u=0}^N P(\omega \in G | X = u) \cdot P(X = u) = \sum_{u=0}^N \frac{u}{N} \cdot P(X = u) = \frac{EX}{N}.$$

Bayes tétele szerint

$$P(X = x | \omega \in G) = \frac{P(\omega \in G | X=x) \cdot P(X=x)}{P(\omega \in G)} = P(X = x) \cdot \frac{x}{EX}.$$

A  $P(X > 0) \neq 0$  feltétel miatt  $EX > 0$ , az (1) formula minden  $x$  mellett kiértékelhető.

Az 1. állítás szerint az egyedek saját csoportja egyedszámának eloszlása paradox módon különbözik a csoportok egyedszámának eloszlásától. A paradoxon magyarázata az, hogy az egyedek nagyobb eséllyel tartoznak nagyobb csoportokhoz, mint kisebbekhez, így az egyedek saját, szubjektív csoportjának nagysága átlagosan nagyobb, mint az átlagos csoportnagyság. A szubjektív csoportnagyság eloszlásában (1) szerint a nagyobb értékek valószínűsége nő a csoportszintű eloszláshoz képest. A csoportok egyedszáma sok alkalmazásban aggregált, jobbra ferde. Ilyenkor az egyedek saját csoportja egyedszámának eloszlása egyenletesebb lesz (kevésbé lesz jobbra ferde).

A kétféle szempont szerinti átlagos csoportnagyságot pontosan ki is tudjuk számítani.

**1. Következmény:** A egyed saját csoportja egyedszámának várható értéke

$$E(X | \omega \in G) = \frac{EX^2}{EX}, \quad (2)$$

emellett tetszőleges  $f$  valós függvényre

$$E(f(X) | \omega \in G) = \frac{EXf(X)}{EX}. \quad (3)$$

Az első formula a lineáris, a második az általános  $f$  skálafüggvényű zsúfoltság populációs értéke, ha  $f$  monoton növekvő (Reiczigel et al. 2005, 2008).

**Bizonyítás:** Elég a második egyenlőséget bizonyítani, az első ennek speciális esete. A feltételes várható érték definíciója és az 1. állítás szerint

$$E(f(X) | \omega \in G) = \sum_{x=0}^N f(x) \cdot P(X = x | \omega \in G) = \sum_{x=0}^N f(x) \frac{x}{EX} \cdot P(X = x) = \frac{EXf(X)}{EX}.$$

Az előző állítást könnyen általánosíthatjuk  $Y(G)$  és  $X$  egyedszemponitű és csoportszintű együttes eloszlásának összehasonlítására.

**2. Állítás:**  $P(Y(G) = y, X = x | \omega \in G) = P(Y(G) = y, X = x) \cdot \frac{x}{EX}.$  (4)

**Bizonyítás:** Definíció szerint

$$P(Y(G) = y, X = x | \omega \in G) = \frac{P(Y(G)=y, X=x, \omega \in G)}{P(\omega \in G)}.$$



A számlálót átalakítva,

$$\begin{aligned} P(Y(G) = y, |G| = x, \omega \in G) &= \sum_{k=1}^K \frac{1}{K} P(Y(g_k) = y, |g_k| = x, \omega \in g_k | G = g_k) = \\ &= \sum_{k=1}^K \frac{1}{K} P(Y(g_k) = y, |g_k| = x, \omega \in g_k) = \\ &= \sum_{k=1}^K \frac{1}{K} P(Y(g_k) = y, |g_k| = x) \cdot P(|g_k| = x, \omega \in g_k) = P(Y(G) = y, X = x) \cdot \frac{x}{N}. \end{aligned}$$

Az átalakítások során felhasználtuk, hogy  $G$  független  $Y$ -tól és  $\omega$ -tól, továbbá  $|g_k| = x$  vagy a biztos, vagy a lehetetlen esemény.

A nevezőre az 1. állítás bizonyítása során nyert

$$P(\omega \in G) = \frac{EX}{N}$$

összefüggés teljesül. Az állításban szereplő formulát kapjuk, ha a számlálóba és a nevezőbe visszahelyettesítjük a kapott kifejezéseket.

Ha  $x$ -et (4)-ből kiintegráljuk, akkor  $Y(G)$  eloszlását nyerjük.

$$\mathbf{3. \text{ Állítás: }} P(Y(G) = y | \omega \in G) = P(Y(G) = y) \cdot \frac{E(X|Y(G)=y)}{EX}. \quad (5)$$

Ennek az összefüggésnek az a jelentősége, hogy ha van egy  $P(Y(G) = y)$  eloszlásmodellünk és az  $X$  csoportméretre egy  $E(X|Y(G) = y)$  regressziós modellünk, akkor ezekből  $Y(G)$  egyedszemponitú eloszlásmodelljét ki tudjuk számítani. Ha például a csoportméret regressziós modellje a

$$g(E(X|Y(G) = y)) = \alpha + \beta^T \cdot y$$

általánosított lineáris modell, akkor  $Y(G)$  egyedszemponitú eloszlásmodellje

$$P(Y(G) = y | \omega \in G) = \frac{g^{-1}(\alpha + \beta^T y)}{E g^{-1}(\alpha + \beta^T Y(G))} \cdot P(Y(G) = y). \quad (6)$$

**2. Következmény:** Legyen  $Y = (U, V)$ . Ha  $P(X > 0 | V(G) = v) \neq 0$ , akkor

$$P(U(G) = u | V(G) = v, \omega \in G) = P(U(G) = u | V(G) = v) \cdot \frac{E(X|U(G)=u, V(G)=v)}{E(X|V(G)=v)}. \quad (7)$$

**Bizonyítás:** Alkalmazzuk (5)-öt az  $(U, V)$  és a  $V$  változókra. A kapott egyenleteket elosztva egymással, megkapjuk (7)-t.

Érdeemes az (5) eloszlás várható értékét kiszámolni.

**4. Állítás:** Tegyük fel, hogy  $Y(g)$  dimenziója nem függ a  $g$  csoporttól. Ekkor

$$E(Y(G) | \omega \in G) = \frac{E(X \cdot Y(G))}{EX}. \quad (8)$$

**Bizonyítás:** A várható érték (5) alapján

$$E(Y(G) | \omega \in G) = E\left(\frac{Y(G) \cdot E(X|Y(G))}{EX}\right) = \frac{E(E(X \cdot Y(G) | Y(G)))}{EX} = \frac{E(X \cdot Y(G))}{EX}.$$

A (8) összefüggés nem más, mint a populációs szintű Lloyd-féle keresztzsúfoltság (Lloyd 1967).

A 4. állítás az egyedszemponitú és a csoportszintű várható érték kapcsolatát fejezi ki. A nagy számok erős törvénye szerint, ha az  $u_1, u_2, \dots, u_n$  mintát az egyedek közül visszatevéssel, egymástól függetlenül, ugyanakkora valószínűséggel választjuk ki, továbbá  $u_i \in g_i, i = 1, \dots, n$ , akkor  $\frac{1}{n} \sum_{i=1}^n Y(g_i) \rightarrow E(Y(G)|\omega \in G)$  majdnem mindig (azaz egy valószínűséggel), ha  $n \rightarrow \infty$ . Válasszunk ki most egy  $g_1, g_2, \dots, g_m$  mintát a csoportok közül visszatevéssel, egymástól függetlenül, ugyanakkora valószínűséggel. Legyen  $x_i = |g_i|, i = 1, \dots, m$ . A nagy számok erős törvénye miatt  $\frac{\sum_{i=1}^m x_i \cdot Y(g_i)}{\sum_{i=1}^m x_i} = \frac{\frac{1}{m} \sum_{i=1}^m x_i \cdot Y(g_i)}{\frac{1}{m} \sum_{i=1}^m x_i} \rightarrow \frac{E(X \cdot Y(G))}{EX} = E(Y(G)|\omega \in G)$  majdnem mindig, ha  $m \rightarrow \infty$ . Az egyedszemponitú és csoportszintű minták és a hozzájuk tartozó adatok összetétele más, a mintaátlagok azonban határértékben megegyeznek.

Ha  $Y(G) = (U(G), V(G))$ , akkor az  $U(G)$  és  $V(G)$  változók közötti kapcsolat, pl. korreláció vagy regresszió más lehet az egyedek szemponitjából és a csoportok szintjén. Előfordulhat, hogy csoportszinten korrelálatlan a két változó, az egyedek szemponitjából pedig nem nulla a korreláció, vagy fordítva.

**5. Állítás:** Legyen  $Y(G) = (U(G), V(G))$  kétdimenziós vektorváltozó. Az  $U(G)$  és  $V(G)$  változók egyedszemponitú kovarianciája

$$Cov(U(G), V(G)|\omega \in G) = \frac{EU(G)V(G)X}{EX} - \frac{EU(G)X}{EX} \cdot \frac{EV(G)X}{EX}. \quad (9)$$

Az 5. állítás egyszerű következménye a 4. állításnak. Ha  $U(G) = V(G)$ , akkor (9) speciális esete az egyedszemponitú variancia

$$Var(U(G)|\omega \in G) = \frac{EU(G)^2X}{EX} - \left(\frac{EU(G)X}{EX}\right)^2. \quad (10)$$

Az egyedszemponitú korreláció

$$Corr(U(G), V(G)|\omega \in G) = \frac{\frac{EU(G)V(G)X}{EX} - \frac{EU(G)X}{EX} \cdot \frac{EV(G)X}{EX}}{\left(\frac{EU(G)^2X}{EX} - \left(\frac{EU(G)X}{EX}\right)^2\right)^{\frac{1}{2}} \cdot \left(\frac{EV(G)^2X}{EX} - \left(\frac{EV(G)X}{EX}\right)^2\right)^{\frac{1}{2}}}. \quad (11)$$

Most példát adunk olyan csoportszinten korrelálatlan változókra, melyek az egyedek szemponitjából korreláltak. Legyen  $U(G)$  független  $X = |G|$ -től és  $V(G) = U(G) \cdot (X - EX)$ . Egyszerű számolással ellenőrizhető, hogy a csoportszintű kovariancia

$$Cov(U(G), V(G)) = EU(G)V(G) - EU(G) \cdot EV(G) = 0,$$

tehát  $U(G)$  és  $V(G)$  csoportszinten korrelálatlan. Az egyedszemponitú kovariancia (9) alapján, egyszerű átalakítással

$$\text{Cov}(U(G), V(G)|\omega \in G) = \text{Var}(U(G)) \cdot \frac{\text{Var}(X)}{EX}$$

pozitív, ha  $U(G)$  és  $X$  szórása pozitív. Ilyenkor tehát  $U(G)$  és  $V(G)$  között az egyedek szempontjából pozitív a korreláció. A konstrukció mintájára könnyedén adható példa az egyedek szempontjából korrelálatlan és csoportszinten korrelált, illetve egyedszempontból és csoportszinten ellentétes előjellel korrelált változókra. Ha például  $W(G) = U(G) \cdot \left(1 - \frac{EX}{X}\right)$ , akkor  $U(G)$  és  $W(G)$  az egyedek szempontjából korrelálatlan, csoportszinten pedig negatív a korreláció közöttük.

Fentiekkel összevetve, némileg meglepő a következő.

**6. Állítás:** Ha  $Y(G)$  és  $X$  független csoportszinten, akkor független az egyedek szempontjából is.

**Bizonyítás:** Tegyük fel a csoportszintű függetlenséget. Az egyedszempontú függetlenség a

$$P(Y(G) = y, X = x|\omega \in G) = P(Y(G) = y|\omega \in G) \cdot P(X = x|\omega \in G) \quad (12)$$

egyenlőséget jelenti. A 2. állítás és  $Y(G)$ ,  $X$  csoportszintű függetlensége miatt

$$\begin{aligned} P(Y(G) = y, X = x|\omega \in G) &= P(Y(G) = y, X = x) \cdot \frac{x}{EX} = \\ P(Y(G) = y) \cdot P(X = x) \cdot \frac{x}{EX} &= P(Y(G) = y) \cdot P(X = x) \cdot \frac{x}{EX}, \end{aligned}$$

továbbá a 3. állítás szerint

$$\begin{aligned} P(Y(G) = y|\omega \in G) &= P(Y(G) = y) \cdot \frac{E(X|Y(G)=y)}{EX} = P(Y(G) = y) \cdot \frac{EX}{EX} = \\ P(Y(G) = y), \end{aligned}$$

ezen kívül az 1. állítás alapján

$$P(X = x|\omega \in G) = P(X = x) \cdot \frac{x}{EX}.$$

A 6. állítás megfordítása csak további feltétel mellett igaz.

**7. Állítás:** Ha  $Y(G)$  és  $X$  függetlenek az egyedek szempontjából és  $P(X > 0) = 1$ , akkor függetlenek csoportszinten is.

**Bizonyítás:** Tegyük fel, hogy (12) igaz. Az 1., 2. és 3. állítás miatt (12)-ből a

$$P(Y(G) = y, X = x) \cdot \frac{x}{EX} = P(Y(G) = y) \cdot \frac{E(X|Y(G)=y)}{EX} \cdot P(X = x) \cdot \frac{x}{EX}$$

egyenlőséget kapjuk. Ezt  $x > 0$  miatt  $\frac{x}{EX}$ -szel egyszerűsíthetjük, az

$$P(Y(G) = y, X = x) = P(Y(G) = y) \cdot \frac{E(X|Y(G)=y)}{EX} \cdot P(X = x) \quad (13)$$

összefüggés adódik. Kiintegráljuk  $x$ -et, az eredményül kapott

$$P(Y(G) = y) = P(Y(G) = y) \cdot \frac{E(X|Y(G)=y)}{EX}$$

egyenlőséget pedig visszahelyettesítjük (13)-ba. Az így nyert

$$P(Y(G) = y, X = x) = P(Y(G) = y) \cdot P(X = x)$$

egyenlőség bizonyítja a csoportszintű függetlenséget.

Vizsgáljuk meg általánosan, milyen feltételek teljesülése esetén egyeznek meg a csoportszintű és egyedszemponitú eloszlásmodellek.

**8. Állítás:** Legyen  $Y = (U, V)$ . Ha  $E(X|U(G) = u, V(G) = v) = E(X|V(G) = v)$  és  $P(X > 0|V(G) = v) \neq 0$ , akkor

$$P(U(G) = u|V(G) = v, \omega \in G) = P(U(G) = u|V(G) = v).$$

**Bizonyítás:** A 3. állítás alapján

$$P(U(G) = u, V(G) = v|\omega \in G) = \frac{E(X|U(G)=u, V(G)=v)}{EX} \cdot P(U(G) = u, V(G) = v).$$

Másrészt

$$P(V(G) = v|\omega \in G) = \frac{E(X|V(G)=v)}{EX} \cdot P(V(G) = v).$$

Az  $E(X|U(G) = u, V(G) = v) = E(X|V(G) = v)$  feltételt az első egyenlőségbe behelyettesítve, majd az első két egyenlőséget egymással elosztva kapjuk a bizonyítandó összefüggést.

Megjegyzés: Ha  $U(G) = u$  és  $X$  függetlenek a  $V(G) = v$  feltételre vonatkozóan, akkor a 8. állításban szereplő  $E(X|U(G) = u, V(G) = v) = E(X|V(G) = v)$  összefüggés teljesül.

**3. Következmény:** Ha  $x > 0$ , akkor

$$P(U(G) = u|V(G) = v, X = x, \omega \in G) = P(U(G) = u|V(G) = v, X = x).$$

**Bizonyítás:**  $U(G) = u$  nyilvánvalóan független  $X$ -től a  $V(G) = v, X = x$  feltétel mellett, így alkalmazható a 8. állítás.

A 3. következmény azt fogalmazza meg, hogy ha a feltételes eloszlás feltételei között szerepel az  $X = |G|$  csoportméret, akkor a csoportszintű és az egyedek szempontjából vett eloszlások megegyeznek. Lényegében arról van szó, hogy az  $X = x$  feltételnek eleget tevő csoportok rétegében a csoportok egyenlő egyedszáma miatt mindegy, hogy csoportszintű vagy egyedszemponitú valószínűségeket számolunk.

A továbbiak során feltesszük, hogy  $U(g)$  dimenziója nem függ  $g$ -től,  $V(g)$  dimenziója viszont változhat a  $g$  csoport függvényében. Tekintsük a  $\varphi(V(G)) = E(U(G)|V(G))$  nemlineáris regressziós modellt. Itt a  $V(G)$  vektor csoportszintű fix és random hatásokat, faktorokat és kovariánsokat tartalmazhat. A modellcsalád a

$$g\left(E(U(G)|V(G))\right) = \alpha + \beta^T V(G)$$

általánosított lineáris kevert modelleket is magában foglalja. A következő két állítás elégséges feltételt ad arra, hogy a regressziós modell ugyanaz legyen csoportszinten és az egyedek szempontjából. Mindkét állítás egyszerűen levezethető a 8. állításból.

**9. Állítás:** Ha  $U(G)$  és  $X$  feltételesen függetlenek a  $V(G) = v$  eseményre vonatkozóan és  $P(X > 0|V(G) = v) \neq 0$ , akkor

$$E(U(G)|V(G) = v, \omega \in G) = E(U(G)|V(G) = v).$$

A 9. állítás az egyedszemponitú és csoportszintú regressziós modellek egybeesésére ad elégséges feltételt. Egyedszemponitú regressziós modellt úgy illesztünk, hogy visszatevéssel, egymástól függetlenül, ugyanakkora valószínűséggel egyedekből álló mintát választunk ki, majd az egyedekhez tartozó csoportok adatait modellezzük. Csoportszintú regressziós modellhez a csoportokból veszünk ki néhányat visszatevéssel, egymástól függetlenül, ugyanakkora valószínűséggel. Az így kapott csoportok adataira illesztjük a regressziót. A két megközelítés más mintavételi eljárást jelent, eltérő adatszerkezetet eredményez, az eredmények értelmezése is eltérő, még akkor is, ha az eredményül kapott modellek egybeesnek.

Az alkalmazások során előfordul, hogy csak egyedszemponitú vagy csak csoportszintú mintavételezésre van lehetőség. Ilyen esetekben a kísérlet, az adatgyűjtés tervezésekor nagyon fontos annak biztosítása, hogy a regressziós modell lehetőleg mindkét szemlélet mellett ugyanaz legyen. Most erre fogalmazunk meg egyszerűen teljesíthető feltételt.

**10. Állítás:** Ha  $x > 0$ , akkor

$$E(U(G)|V(G) = v, X = x, \omega \in G) = E(U(G)|V(G) = v, X = x).$$

A 10. állítást a 3. következményből nyerjük.

A 10. állítás szerint minden, a csoport–egyed közösségre vonatkozó regressziós modellbe érdemes az  $X = |G|$  csoportméretet magyarázó változóként bevenni (szükség szerint a többi fix és random hatásos magyarázó változóval interakcióban), mert ez garantálja, hogy az egyedek szempontjából ugyanazt a modellt kapjuk, mint a csoportok szintjén. Az  $X$  csoportméret akkor hagyható el a regressziós modellből, ha elhagyva a modell változása nem releváns. Ha a modell lényegesen megváltozik, akkor mind a csoportok, mind az egyedek szempontjából  $X$  hiányzó kovariáns, zavaró változó (confounder variable) lehet, ami a modellek torzítását és téves következtetések levonását eredményezheti.

**11. Állítás:** Ha  $P(X > 0|V(G) = v) \neq 0$ , akkor a csoportszintű és egyedszempon-tú regressziós modellek általános kapcsolatát az

$$E(U(G)|V(G) = v, \omega \in G) = \frac{E(X \cdot U(G)|V(G)=v)}{E(X|V(G)=v)} \quad (14)$$

egyenlőség fejezi ki.

**Bizonyítás:** A 2. következmény (7) eloszlásformulájával kiszámítjuk (14) baloldalát:

$$E(U(G)|V(G) = v, \omega \in G) = E \left\{ U(G) \cdot \frac{E(X|U(G),V(G)=v)}{E(X|V(G)=v)} \middle| V(G) = v \right\} =$$

$$E \left\{ \frac{E(X \cdot U(G)|U(G),V(G)=v)}{E(X|V(G)=v)} \middle| V(G) = v \right\} = \frac{E(X \cdot U(G)|V(G)=v)}{E(X|V(G)=v)}.$$

A (14) formula arra is alkalmas, hogy egy regressziós modellben bizonyos kovariánsok melletti marginális eloszlások várható értékét kiszámíthassuk. Ha a kiválasztott kovariánsok között nem szerepel minden modellbeli  $X$ -től függő,  $X$ -szel interakcióban lévő magyarázó változó, akkor a marginális várható értékek különbözhetnek csoportszinten és az egyedek szempontjából akkor is, ha a modell maga megegyezik a két szinten.

A 11. állítás alapján szükséges és elégséges feltételt adhatunk arra, hogy ugyanazt a regressziós modellt kapjuk csoportszinten és az egyedek szempontjából.

**12. Állítás:** Tegyük fel, hogy  $P(X > 0|V(G) = v) \neq 0$ . Az egyedszempon-tú és a csoportszintű regressziós modellek pontosan akkor esnek egybe, azaz

$$E(U(G)|V(G) = v, \omega \in G) = E(U(G)|V(G) = v), \quad (15)$$

ha az  $U(G)$  vektor minden koordinátája és  $X$  feltételesen korrelálatlan a  $V(G) = v$  eseményre vonatkozóan.

**Bizonyítás:**

A 11. állítás szerint (15) pontosan akkor teljesül, ha

$$E(X \cdot U(G)|V(G) = v) = E(X|V(G) = v) \cdot E(U(G)|V(G) = v),$$

ami bizonyítja az állítást.

A 12. állítás a gyakorlati alkalmazások szempontjából azért lehet jelentős, mert a korrelálatlanság statisztikai vizsgálata sokszor jóval egyszerűbb, mint a függetlenség ellenőrzése. A következő állítások a gyakorlatban rutinszerűen alkalmazható elégséges feltételeket adnak a 12. állításban szereplő korrelálatlanságra.

**13. Állítás:** Tegyük fel, hogy  $E(U(G)|X, V(G) = v) = E(U(G)|V(G) = v)$ . Ekkor az  $U(G)$  vektor minden koordinátája és  $X$  feltételesen korrelálatlan a  $V(G) = v$  eseményre vonatkozóan.

**Bizonyítás:**

$$\begin{aligned}
E(X \cdot U(G)|V(G) = v) &= E\{E(X \cdot U(G)|X, V(G) = v)|V(G) = v\} = \\
E\{X \cdot E(U(G)|X, V(G) = v)|V(G) = v\} &= E\{X \cdot E(U(G)|V(G) = v)|V(G) = v\} = \\
E\{X|V(G) = v\} \cdot E(U(G)|V(G) = v). &
\end{aligned}$$

**14. Állítás:** Tegyük fel, hogy  $E(X|U(G), V(G) = v) = E(X|V(G) = v)$ . Ekkor az  $U(G)$  vektor minden koordinátája és  $X$  feltételesen korrelálatlan a  $V(G) = v$  eseményre vonatkozóan.

**Bizonyítás:**

$$\begin{aligned}
E(X \cdot U(G)|V(G) = v) &= E\{E(X \cdot U(G)|U(G), V(G) = v)|V(G) = v\} = \\
E\{U(G)E(X|U(G), V(G) = v)|V(G) = v\} &= E\{U(G)E(X|V(G) = v)|V(G) = v\} = \\
E\{X|V(G) = v\} \cdot E(U(G)|V(G) = v). &
\end{aligned}$$

A 13. állítás feltétele azt jelenti, hogy  $U(G)$ -nek  $X$ -re és  $V(G)$ -re vett regressziós modelljéből  $X$  elhagyható. A 14. állítás feltétele pedig akkor teljesül, ha  $X$ -nek  $U(G)$ -re és  $V(G)$ -re vett regressziós modelljéből  $U(G)$  elhagyható. Magyarázó változók regressziós modellből való elhagyására vagy modellben tartására közismert statisztikai módszerek szolgálnak, ezekkel az említett feltételek ellenőrizhetők.

Összefoglalva, az egyedszemponitú és a csoportszintű regressziós és eloszlásmodellek egybeesését úgy tudjuk a legegyszerűbben és legtermészetesebb módon garantálni, ha az adatok gyűjtése során minden mintába választott csoportról feljegyezzük a bennük lévő összes egyed számát és ezt beépítjük a modellekbe. Ha az adatgyűjtés ilyen módon nehézségekbe ütközik, akkor érdemes lehet a csoport fogalmát úgy módosítani, hogy az egyedszámot meg tudjuk állapítani. Gazdaállatok egész testfelületének mintázása helyett szorítkozhatunk egy meghatározott testtájékra, területek teljes körű vizsgálata helyett egy-egy kvadrátba, kihelyezett csapdába kerülő egyedeket tekinthetünk csoportnak. Ilyenkor természetesen a vizsgált közösség fogalma és az eredmények értelmezése is módosul.

A gyakorlati alkalmazások egy részében az  $X$  csoportméret nem ismert, mert a csoportok mintavételezése csak a csoport egy részére, pl. néhány területre, kvadrátra terjedt ki. Ha a csoportok egyedeinek csak egy részét mintavételezzük és a vizsgált célváltozó értékét csak ebből a részmintából és a csoport általános jellemzőiből határozzuk meg, akkor sokszor feltehető, biológiai szempontból alátámasztható, hogy a célváltozó független a csoport nem mintavételezett egyedeitől. A 12. állítás módosításával szükséges és elégséges feltételt kapunk olyan csoportszintű és egyedszemponitú modellek egybeesésére, amelyekben a csoportok teljes egyedszáma nem ismert.

**15. Állítás:** Tegyük fel, hogy  $S \subset G, |S| = T, |G| = X$  és  $P(X > 0 | T = t, V(S, G) = v) \neq 0$ . Az  $U(S, G)$  célváltozó egyedszemponitú és csoportszintű regressziós modellje a  $T = t$  és  $V(S, G) = v$  feltétel mellett pontosan akkor esik egybe, azaz

$$E(U(S, G) | T = t, V(S, G) = v, \omega \in G) = E(U(S, G) | T = t, V(S, G) = v), \quad (16)$$

ha az  $U(S, G)$  célváltozó korrelálatlan a nem mintavételezett egyedek  $X - T$  számával a  $T = t, V(S, G) = v$  feltétel mellett.

**Bizonyítás:** A  $T = t$  feltétel miatt  $X - T = X - t$ , ezért  $U(S, G)$  és  $X - T$  korrelálatlansága  $U(S, G)$  és  $X$  korrelálatlanságával ekvivalens. Az állítás ezért következik a 12. állításból.

Megjegyzések:

1. Ha  $U(S, G)$  csak az  $S$  mintától és a  $G$  csoport egészének jellemzőitől függ, ezen kívül a magyarázó változók  $V(S, G)$  vektora „elég gazdag”, elegendő információt tartalmaz, akkor az esetek nagy részében elérhető, hogy a 15. állításban szereplő korrelálatlanság feltétele teljesüljön.
2. A 13. és 14. állítások alapján a 15. állításban szereplő korrelálatlanság teljesül, ha az  $E(U(S, G) | X, T = t, V(G) = v) = E(U(S, G) | T = t, V(G) = v)$  összefüggés vagy pedig az  $E(X | U(S, G), T = t, V(G) = v) = E(X | T = t, V(G) = v)$  összefüggés fennáll. Ha pl.  $T$  a  $G$  véletlenszerűen kiválasztott egységnyi területű kvadrátján lévő egyedek száma és a  $G$  csoport  $A(G) = a$  összterülete szerepel a  $V(G) = v$  feltételek között, akkor sok esetben  $E(X | U(S, G), T = t, V(G) = v) = E(X | T = t, V(G) = v) = t \cdot a$ , tehát a 15. állításban a korrelálatlanság feltétele teljesül.

**16. Állítás:** Tegyük fel, hogy  $U(G)$  és  $X$  feltételesen korrelálatlanok  $V(G)$ -re nézve. Legyen  $V^* = \psi(V)$  a  $V$  determinisztikus függvénye. Ha  $P(X > 0 | V^*(G) = v^*) \neq 0$ , akkor

$$E(U(G) | V^*(G) = v^*, \omega \in G) = E(E(U(G) | V(G)) | V^*(G) = v^*, \omega \in G) = \frac{E(X \cdot E(U(G) | V(G)) | V^*(G) = v^*)}{E(X | V^*(G) = v^*)} \quad (17)$$

és

$$E(U(G) | V^*(G) = v^*) = E(E(U(G) | V(G)) | V^*(G) = v^*). \quad (18)$$

**Bizonyítás:** A 11. állítás szerint

$$E(U(G) | V^*(G) = v^*, \omega \in G) = \frac{E(X \cdot U(G) | V^*(G) = v^*)}{E(X | V^*(G) = v^*)}.$$



A kapott tört számlálója

$$\begin{aligned} E(X \cdot U(G) | V^*(G) = v^*) &= E(E(X \cdot U(G) | V(G)) | V^*(G) = v^*) = \\ E(E(X | V(G)) \cdot E(U(G) | V(G)) | V^*(G) = v^*) &= \\ E\left(E\left(X \cdot E(U(G) | V(G)) | V(G)\right) | V^*(G) = v^*\right) &= E(X \cdot E(U(G) | V(G)) | V^*(G) = v^*), \end{aligned}$$

kihasználva  $U(G)$  és  $X$  feltételes korrelálatlanságát  $V(G)$ -re nézve. A (18) összefüggés egyszerű következménye a feltételes várható érték tulajdonságainak.

A 16. állítás jelentése az, hogy  $U(G)$  egyedszemponitú és csoportszintű átlaga a  $V^*(G) = v^*$  részpopuláción egybeesik  $E(U(G) | V(G))$  egyedek szemponitjából, ill. csoportszinten vett átlagával ezen a részpopuláción. Az állítás jelentőségét az adja, hogy az  $E(U(G) | V(G))$  feltételes várható érték varianciája kisebb vagy egyenlő, mint  $U(G)$  varianciája, ezért remélhető, hogy az egyedek szemponitjából vett átlag statisztikai becslésének kisebb lesz a standard hibája, ha  $U(G)$  helyett az  $E(U(G) | V(G))$  feltételes várható érték regressziós modelljéből nyert predikciót átlagoljuk.

## 6 A vetési varjún (*Corvus frugilegus*) és a dolmányos varjún (*Corvus cornix*) élő tetűparaziták abundanciája, zsúfoltsága, diverzitása és aggregáltsága

A két varjófaj és számos további madárfaj tetűparazitáltságára vonatkozó adatokat dr. Kiss János Botond és dr. Rékási József gyűjtötték Romániában és Magyarországon az 1970-es évektől az 1990-es évekig (Rékási 1973, 1978, 1984, 1986, 1991, 1993, Rékási és Kiss 1977, 1980, 1984, további publikálatlan adatok). Több olyan elemző, összehasonlító tanulmány is készült, amelyek részben vagy teljesen ezekre az adatokra támaszkodnak (Rózsa et al. 1996, Rékási et al. 1997, Rózsa 1997). Mindkét varjófajon 5–5 tetűfaj él, ezek génuszai egymással megegyező párokat alkotnak. A génuszokra építve a paraziták abundanciái biológiai és statisztikai értelemben is összehasonlíthatók (Rózsa et al. 1996). A továbbiak során a két varjófaj parazitáltságát először külön, majd egymással összehasonlítva vizsgáljuk. A tetvek elnevezésében Price et al. (2003) összefoglalóját követjük.

### 6.1 Vetési varjú

#### 6.1.1 Adatok, leíró statisztikák

Vetési varjúból 37 példány került a mintába. Ezeken a gazdaállatokon összesen 1973 tetűegyedet találtak, melyek 5 fajhoz tartoznak. Az abundanciák összegének, átlagának, szórásának, variációs együtthatójának és a lineáris zsúfoltságnak (4.1. fejezet) fajok közötti megoszlását mutatja be az 1. táblázat.

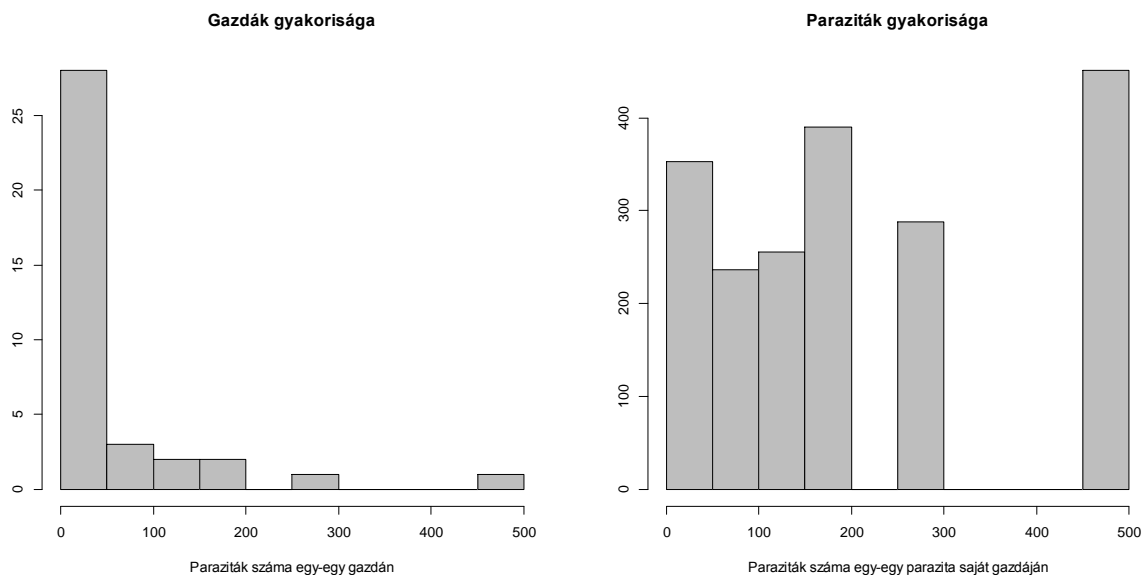
#### 1. táblázat

*Paraziták fajonkénti és együttes prevalenciája, intenzitása, abundanciájának összege, átlaga, szórása, variációs együtthatója, és a fajonkénti lineáris zsúfoltság*

Tetűfaj	Prevalencia (%)	Intenzitás átlaga	Abundancia				Lineáris zsúfoltság
			összege	átlaga	szórása	variációs együtthatója (%)	
<i>Brueelia tasniamae</i>	48.65	43.72	787	21.27	43.78	205.81	108.93
<i>Myrsidea isostoma</i>	70.27	22.00	572	15.46	33.93	219.46	87.91
<i>Menacanthus gonophaeus</i>	64.86	23.29	559	15.11	32.80	217.07	84.37
<i>Philopterus atratus</i>	24.32	4.22	38	1.03	2.66	259.13	7.74
<i>Colpocephalum fregili</i>	8.11	5.67	17	0.46	1.77	385.93	7.12
<i>Együttes</i>	<i>91.89</i>	<i>58.03</i>	<i>1973</i>	<i>53.32</i>	<i>93.98</i>	<i>176.24</i>	<i>214.48</i>

Az 5 tetűfaj közül 3 domináns, ezek a *Brueelia tasniamae*, *Myrsidea isostoma* és *Menacanthus gonophaeus* fajok. Közülük a *Brueelia tasniamae* átlagos abundanciája a legnagyobb, de a másik kettőé is hasonló nagyságrendű. A 3 domináns tetűfaj adja az együttes abundancia 97%-át. Az abundanciák szórása nagy, minden fajra és együtt is jóval meghaladja az átlagos abundanciát. A variációs együttható 200-300%-os. Ez arra utal, hogy a fajonkénti és együttes parazitáltság erősen aggregált a vetési varjakon. A lineáris zsúfoltság minden tetűfajra és együttesen is jóval nagyobb, mint az átlagos abundancia.

A tetűparaziták számának eloszlását vizsgálhatjuk gazdaállatok között és a paraziták szempontjából (5. fejezet, 1. Állítás). Az első szemléletnek megfelelően varjúegyenedenként megállapítjuk a tetvek számát és ezek hisztogramját készítjük el. A második megközelítésben tetűegyenedenként a gazdán lévő összes tetű számából állítunk össze hisztogramot. A két hisztogram az 1. ábrán látható.



**1. ábra**

*Tetűparaziták számának hisztogramja a gazdák és a paraziták szemszögéből*

A baloldali hisztogram a különbözőképpen parazitált gazdák gyakoriságát ábrázolja. A hisztogram jobbra ferde, mutatva, hogy a gazdákon a tetvek eloszlása aggregált, azaz a tetvek jelentős része csupán néhány varjún él. Az összesen 1973 tetűből 938 egyed (47%) mindössze 3 varjún (8%) található meg. A jobboldali hisztogram a különböző mértékben parazitált gazdájú tetvek gyakoriságát mutatja be. Az eloszlás többé-kevésbé egyenletes, ami azt jelenti, hogy egy véletlenszerűen kiválasztott tetű kb. ugyanakkora eséllyel él kevésbé és nagyon fertőzött gazdán.

## 6.1.2 Zsúfoltságok additív felbontása

A tetűparaziták fajonkénti átlagos zsúfoltságát, átlagos gazdánkénti szubjektív diverzitását és ezek összegét, az együttes zsúfoltságot tartalmazza a 2. táblázat a Lloyd, lineáris, logaritmusos és hiperbolikus zsúfoltságokra (4.1. és 4.2. fejezet).

### 2. táblázat

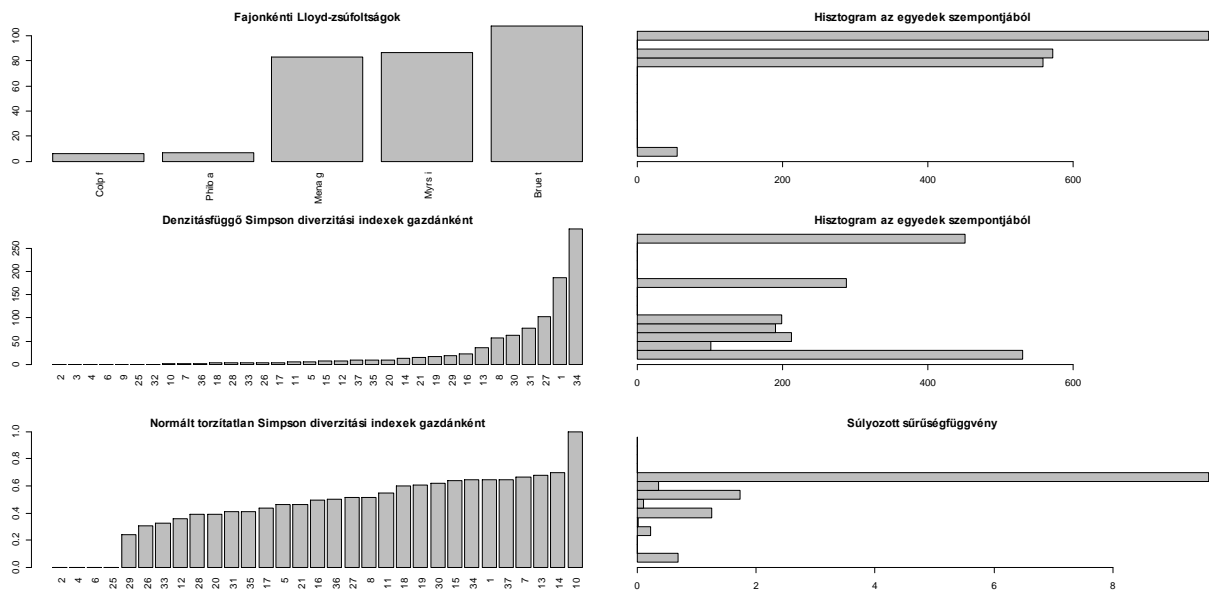
*Zsúfoltságok additív összetevői*

	Fajonkénti átlagos (eredeti skálán)	Szubjektív diverzitás		Együttes (eredeti skálán)
		Denzitásfüggő	Normált (%)	
Lloyd	92.05 (93.05)	121.43	56.88	213.48 (214.48)
Lineáris	93.05 (93.05)	121.43	56.62	214.48 (214.48)
Logaritmusos	4.11 (61.11)	0.84 (2.31*)	16.90	4.95 (141.01)
Hiperbolikus	0.96 (24.66)	0.02	2.37	0.98 (58.03)

*\*Eredeti exponenciális skálán, szorzatmodellben.*

A Lloyd és lineáris zsúfoltságok értéke – mint mindig – csak egy egységnyivel tér el egymástól. A domináns tetűfajok és az erősebben parazitált varjak részarányukhoz képest nagyobb mértékben befolyásolják ezeket a zsúfoltságokat, ennek megfelelően értékük a többihez képest nagyobb. Ez a megállapítás igaz a zsúfoltságok eredeti, parazitaszámra vonatkozó skáláján is. Az együttes, nem fajspecifikus zsúfoltságban a szubjektív diverzitás részaránya nagy, 50%-ot meghaladó mértékű. A lineáris zsúfoltság felbontása szerint egy tetű gazdáján átlagosan 93.05 saját fajhoz tartozó és 121.43 más fajhoz tartozó tetűegyed van. A logaritmusos zsúfoltságban a tetűfajok éppen részarányuknak megfelelő, a varjak pedig parazitaszámuknak megfelelő arányban súlyozottak, ezért a kisebb fajok és a kevésbé parazitált gazdák szerepe kifejezettebb a Lloyd és a lineáris zsúfoltsághoz képest. A logaritmusos zsúfoltság esetében a szubjektív diverzitás részaránya 16.90%, így a fajonkénti zsúfoltság súlya nagyobb az additív felbontásban, mint a Lloyd vagy a lineáris zsúfoltság esetében. Az értelmezés az eredeti skálán egyszerűbb, itt szorzatmodellt kapunk. Eredeti skálán a fajonkénti zsúfoltság 61.11, ezt szorozzuk az átlagos Shannon effektív fajszámmal, 2.31-gyel. Az eredmény 141.01, ami eredeti skálán az együttes zsúfoltság. A hiperbolikus zsúfoltság a ritka fajokat és a kevésbé parazitált gazdákat arányukhoz képest felülreprezentálja. A zsúfoltságok – mint általában – csak kevéssel maradnak 1 alatt, ezért 1-től való eltérésüket érdemes vizsgálni (4.2. fejezet, (22)). A fajonkénti zsúfoltsághoz képest az együttes zsúfoltság 1-től való eltérése felére csökkent, ezért az átlagos szubjektív diverzitás ebben a megközelítésben is jelentősen befolyásolja az együttes zsúfoltságot. Ezt a megállapítást az eredeti skálás fajonkénti és együttes zsúfoltságok nagy eltérése is alátámasztja.

A zsúfoltság additív felbontásában szereplő tagok átlagok, érdemes ezért az összetevőik eloszlását, homogenitását vagy heterogenitását grafikusán megvizsgálni. A Lloyd zsúfoltság összetevőinek eloszlásait a 2. ábra alapján elemezzük.



**2. ábra**

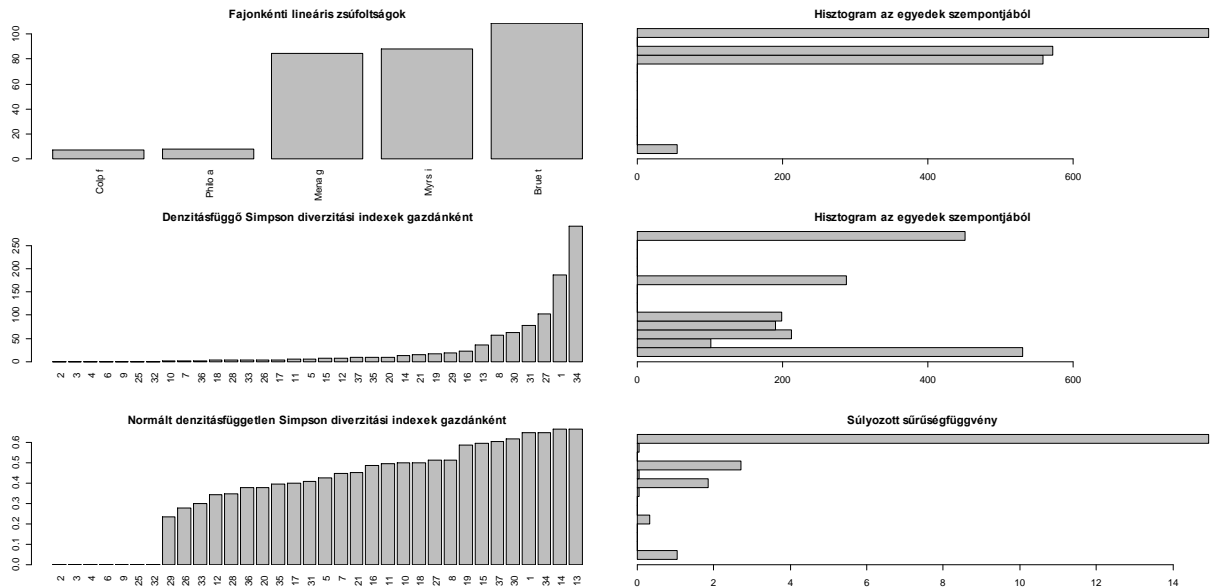
*Lloyd zsúfoltság összetevőinek eloszlása*

Az első sor két ábrája a fajonkénti zsúfoltságokat és ezek egyedszemponjú gyakoriságait jeleníti meg (4.2. fejezet, (9)). Éppen annak a 3 fajnak a zsúfoltsága és egyedszemponjú gyakorisága domináns, amelyeknek az átlagos abundanciája is a legnagyobb (*Brueelia tasniamae*, *Myrsidea isostoma* és *Menacanthus gonophaeus*).

A második sor ábrái a gazdánkénti denzitásfüggő Simpson diverzitási indexek nagyság szerint növekvő sorrendben és ezek egyedszemponjú gyakorisági hisztogramja (4.2. fejezet, (9)). Jelentős a kis szubjektív diverzitású gazdák szerepe. Ezeken kevés a parazita vagy kicsi a lokális Simpson-diverzitás (a következő bekezdésben látni fogjuk, hogy körükben inkább a kis parazitaszám a jellemző). A 34-es és az 1-es sorszámú gazdák kiugróan nagy szubjektív diverzitást mutatnak. A gazdák harmadik, az előző kettővel kb. azonos súlyú, elkülöníthető csoportja a középmezőnybe tartozik.

A harmadik sor első ábrája az egységre normált gazdánkénti Simpson diverzitási indexeket ábrázolja növekvő sorrendben. A mellette lévő ábra a diverzitási indexek súlyozott sűrűségfüggvénye (4.2 fejezet, (11)). A kis diverzitású gazdák súlya kisebb, mint a denzitásfüggő diverzitás esetében. Ez azt jelzi, hogy a kevés tetűfajjal fertőzött varjak súlya csekély. Meghatározó a szerepe a 0.6-0.8 diverzitású varjaknak. Mások mellett ide tartozik az előző bekezdésben kiemelt 34-es és 1-es sorszámú gazda.

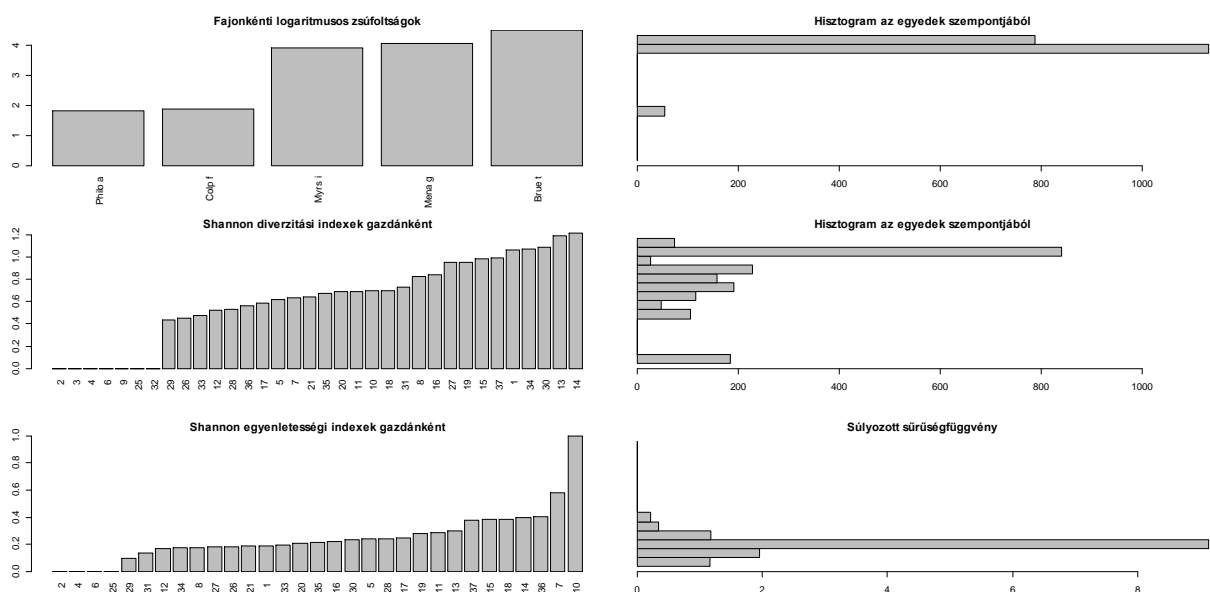
A lineáris zsúfoltság összetevőit a 3. ábrán tanulmányozhatjuk. A kép szinte semmiben se különbözik a Lloyd zsúfoltság szerkezetétől. Talán az egységre normált Simpson diverzitások képe stabilabb, strukturáltabb (3. sor 1. grafikon). A kis és nagy diverzitású gazdák mellett itt is elkülöníthető egy köztes csoport.



3. ábra

*Lineáris zsúfoltság összetevőinek eloszlása*

A logaritmusos zsúfoltságok felbontását és eloszlás szerkezetét a 4. ábrán vizsgáljuk meg.

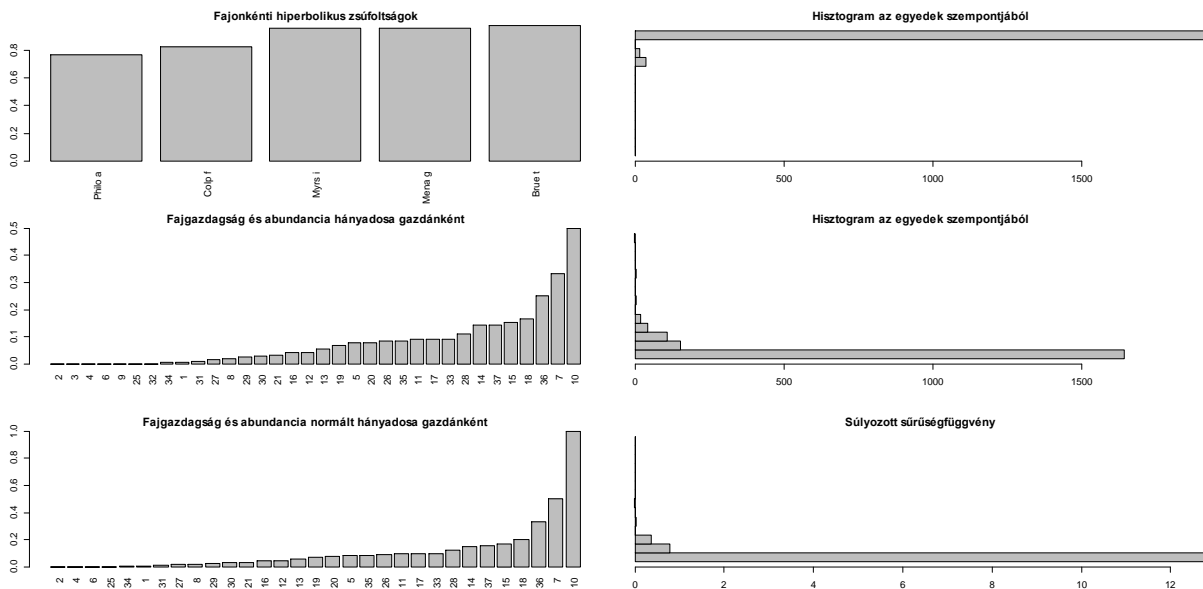


4. ábra

*Logaritmusos zsúfoltság összetevőinek eloszlása*

A fajonkénti zsúfoltságokban a két alacsonyabb részarányú parazita faj relatíve nagyobb értékkel jelenik meg, mint a Lloyd és a lineáris zsúfoltság esetében. A második sorban lévő szubjektív Shannon diverzitás denzitásfüggetlen (4.2. fejezet, (16)). A gazdák két jól elkülönülő csoportba oszthatók, a 0 diverzitásúak (azaz 1 parazitafajjal fertőzöttek) és a 0.5-1.2 diverzitásúak. Ez utóbbiban nagy súlyt képviselnek az 1.0-1.2 diverzitású gazdák. A harmadik ábrában az együttes zsúfoltság szubjektív diverzitáshoz kapcsolható százalékarányát a gazdánkénti Shannon egyenletességi indexek megoszlásával szemléltetjük (4.2. fejezet, (18)). Az egyenletességi indexek eloszlása homogén, egycsúcsú, a legnagyobb súlyú értéktartomány a 0.2-0.3. Az egyenletességi index a Shannon diverzitásból a parazitaszám logaritmusával való normálással számítható ki, ez magyarázhatja, hogy a diverzitáshoz képest az eloszlása koncentráltabb.

A hiperbolikus zsúfoltság összetevőinek szerkezetét az 5. ábra mutatja.



5. ábra

*Hiperbolikus zsúfoltság összetevőinek eloszlása*

A fajonkénti zsúfoltságok értékei között csekély a különbség annak megfelelően, hogy gazdáról gazdára itt csak a tetűfaj jelenléte vagy hiánya számít. A fajonkénti zsúfoltságok egyszempontú gyakorisága a 3 domináns fajra együtt kiemelkedően nagy. A szubjektív diverzitás összetevői most gazdánként a fajgazdagság és az abundancia hányadosai (4.2. fejezet, (20)). Ezek közül legnagyobb súlya a kicsi, 0.0-0.05 tartományba eső értékeknek van. A harmadik ábrában bemutatott normált szubjektív diverzitásoknál (4.2. fejezet, (22)) elsősorban a kicsi, 0.0-0.2 közötti értékek határozzák meg az együttes zsúfoltságban a szubjektív diverzitásra jutó százalékarányt.

### 6.1.3 A valódi diverzitás és az aggregáltság multiplikatív felbontása

A tetűfajok varjankénti átlagos alfa valódi diverzitását, a teljes mintából számolt gamma valódi diverzitását és e kettő hányadosát, a  $\beta = \text{gamma}/\text{alfa}$  diverzitást vizsgáljuk. A tényleges gazdaszám és az effektív gazdaszám hányadosaként értelmezett aggregáltság tetűfajonkénti és fajokra együtt vett átlaga az alfa és gamma aggregáltság. Az alfa/gamma hányados a béta aggregáltság. A diverzitások és aggregáltságok három fő típusát, a lineáris, logaritmusos és hiperbolikus változatot alkalmazzuk (4.3. fejezet). E mennyiségek mintából becsült értékeit a 3. táblázat tartalmazza.

#### 3. táblázat

*A valódi diverzitás és az aggregáltság tényezői*

	Valódi diverzitás			Aggregáltság		
	Alfa	Béta	Gamma	Alfa	Béta	Gamma
Lineáris	1.97	1.56	3.09	5.55	1.38	4.02
Logaritmusos	2.31	1.44	3.32	3.80	1.44	2.64
Hiperbolikus	2.93	1.70	5.00	1.70	1.57	1.09

Ha a diverzitás ritkasági függvénye lineáris, akkor a valódi Simpson diverzitást kapjuk eredményül. Az ennek megfelelő lineáris zsúfoltsági függvény alkalmazása a lineáris zsúfoltságra, aggregáltságra vezet. A Simpson diverzitásban számarányukhoz képest felülreprezentáltak a domináns fajok, a lineáris aggregáltságban pedig a domináns, erősen parazitált gazdák (4.3. fejezet). Az alfa diverzitás 1.97, eszerint a gazdánkénti átlagos effektív tetűfaj-szám kb. 2. A gamma diverzitás 3.09, ami azt jelenti, hogy a gazdák együttesén, a teljes tetűkolóniában az effektív tetűfaj-szám kb. 3. A béta diverzitás 1.56, ez azt a lehetőséget sugallja, hogy a tetűkolónia fajai két olyan csoportra oszthatók, amelyek egyedei közös gazdán viszonylag ritkán fordulnak elő. Az alfa aggregáltság 5.55, ez elég nagy érték, jelzi, hogy a tetűfajok külön-külön a domináns gazdák között aggregáltan oszlanak el. A gamma aggregáltság 4.02, ez is nagy érték, mutatja, hogy a tetűket faj szerint el nem különítve a gazdák közötti aggregáltság továbbra is nagy, nem csökken lényegesen. A béta aggregáltság 1.38, ez utalhat a domináns gazdák két olyan csoportjára, amiken a tetűfajok részarányai eltérőek.

Ha logaritmus a ritkasági függvény, akkor hozzá a Shannon valódi diverzitás és a logaritmusos aggregáltság kapcsolódik. Az alfa és gamma diverzitások a Simpson diverzitásokhoz képest valamivel nagyobbak, ez megfelel annak, hogy most a domináns tetűfajok nem nyomják el a többi fajt, a Shannon valódi diverzitásban mindegyik faj a részarányának megfelelő mértékben reprezentált. A béta diverzitás 1.44, amiből a Simpson



béta diverzitáshoz hasonló következtetésre juthatunk, vagyis elképzelhető, hogy a tetűkolónia fajai két, gazdánként kevéssé átfedő csoportra oszthatók. A logaritmikus alfa és gamma aggregáltság kisebb, mint a lineáris, eszerint a gazdákat parazitáltságuknak megfelelő súllyal figyelembe véve, a tetűfajok külön-külön és együtt is kevésbé aggregáltan oszlanak meg a gazdák között. A logaritmikus béta aggregáltság értéke – mint minden esetben – pontosan megegyezik a béta diverzitás értékével (4.3. fejezet). A béta aggregáltság most 1.44, nem különbözik jelentősen a lineáris béta aggregáltságtól. Szintén a gazdák két olyan csoportjának létezését jelzi, melyeken a tetvek fajonkénti részaránya eltérő.

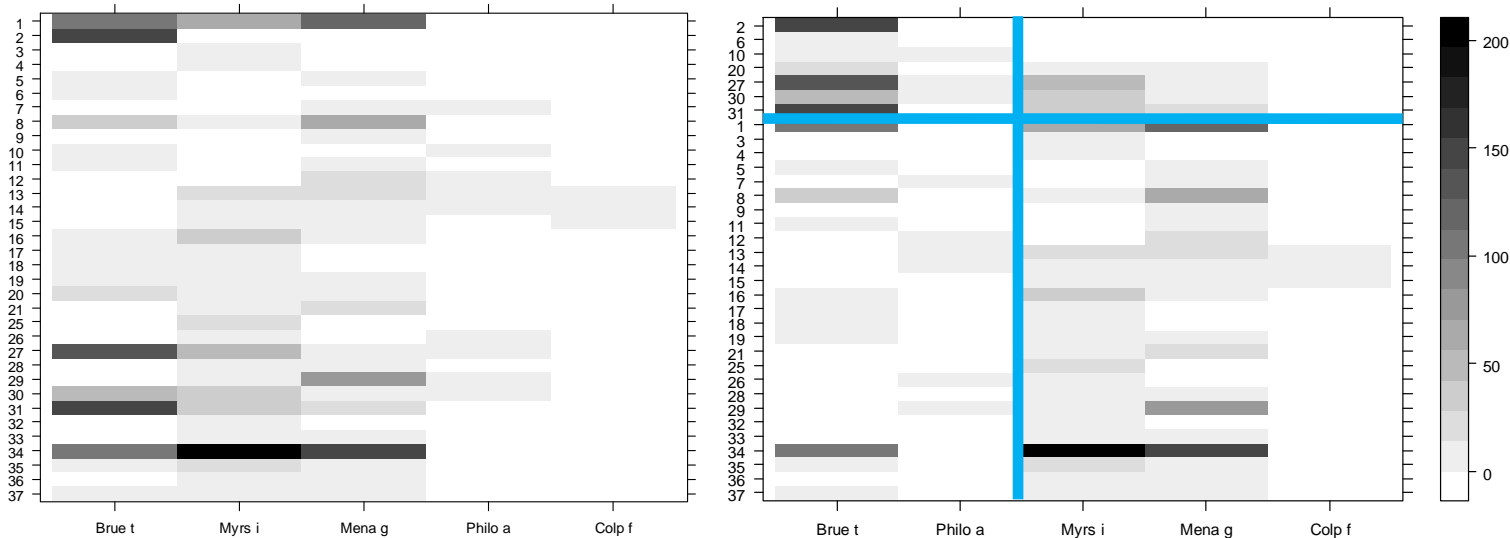
Hiperbolikus ritkasági függvény mellett a valódi diverzitás éppen a tetűfajok száma. Az alfa és a gamma diverzitások itt a legnagyobbak, ami érthető, hiszen a ritkán előforduló fajok ugyanúgy szerepelnek a fajszámban, mint a többi. Ez magyarázza azt is, hogy a béta diverzitás ebben az esetben a legnagyobb. Az alfa diverzitás 2.93, a gamma diverzitás 5, ezért gazdánként átlagosan kb. 3, a teljes kolónián 5 tetűfaj fordul elő. A béta diverzitás 1.7, ami most is a tetűfajok két, gazdákon kevéssé átfedő, hipotetikus csoportjára utal. A hiperbolikus alfa és gamma aggregáltságok kisebbek, mint a megfelelő lineáris és logaritmikus aggregáltságok. Ez megfelel annak, hogy a kevéssé fertőzött gazdákat parazitáltságuknál nagyobb mértékben reprezentálja a hiperbolikus aggregáltság. A béta aggregáltság 1.57, ami az előző két béta aggregáltsággal szinkronban, a gazdák két, tetűfajonként eltérő arányban parazitált csoportjának létezését prognosztizálja.

#### **6.1.4 Gazdák és paraziták csoportosítása**

A valódi béta diverzitások és béta aggregáltságok értékei alapján azt a hipotézist fogalmazhatjuk meg, hogy a tetűfajok és a varjak két-két csoportra oszthatók úgy, hogy az első tetűfaj-csoport elsősorban az első varjúcsoporton, a második tetűfaj-csoport inkább a második varjúcsoporton él. Blokk klaszterező eljárással mód van ennek a feltételezésnek az ellenőrzésére. Az eljárás a kérdéses tetűfaj-csoportokat és varjúcsoportokat a gazda–parazita abundanciátáblázat sorainak és oszlopainak egymás utáni cserélgetésével keresi meg oly módon, hogy a nagy abundanciaértékeket egymás közelébe mozgatja (4.3. fejezet). Az áttekinthetőség érdekében a nem fertőzött gazdák sorai kikerültek a táblázatból, összesen 34 gazda adatai maradtak meg. A tetűmentes gazdák elhagyása nem változtatja meg a valódi diverzitások és az aggregáltságok értékét. A csoportok keresését az így redukált abundanciák táblázatán elvégezve, az eredményt a 6. ábra mutatja.

A bal oldali kép az eredeti elrendezésben, a jobb oldali a blokkos klaszterek szerinti varjú- és tetűfaj-sorrendben ábrázolja az abundanciákat. Az abundanciák értékeit az ábra jobb szélén lévő színskála jelzi. Az eredmény alátámasztja a hipotézist, valóban elkülöníthető két tetűfaj-

csoport és nekik megfelelően két varjúcsoport. Az első tetűfaj-csoport domináns faja a *Brueelia tasniamae*, a csoport másik, jóval ritkább faja a *Philoaterus atratus*. A második tetűfaj-csoport domináns fajai a *Myrsidea isostoma* és a *Menacanthus gonophaeus*. A csoport harmadik faja az igen ritka *Colpocephalum fregili*. A tetűfaj-csoportok a varjaknak csak viszonylag kis hányadán, 6 egyeden (18%) mutatnak jelentősebb átfedést. Kisebb mértékű, másodlagos szegregáció megfigyelhető a második tetűfaj-csoporton belül a *Myrsidea isostoma* és a *Menacanthus gonophaeus* fajok között is.



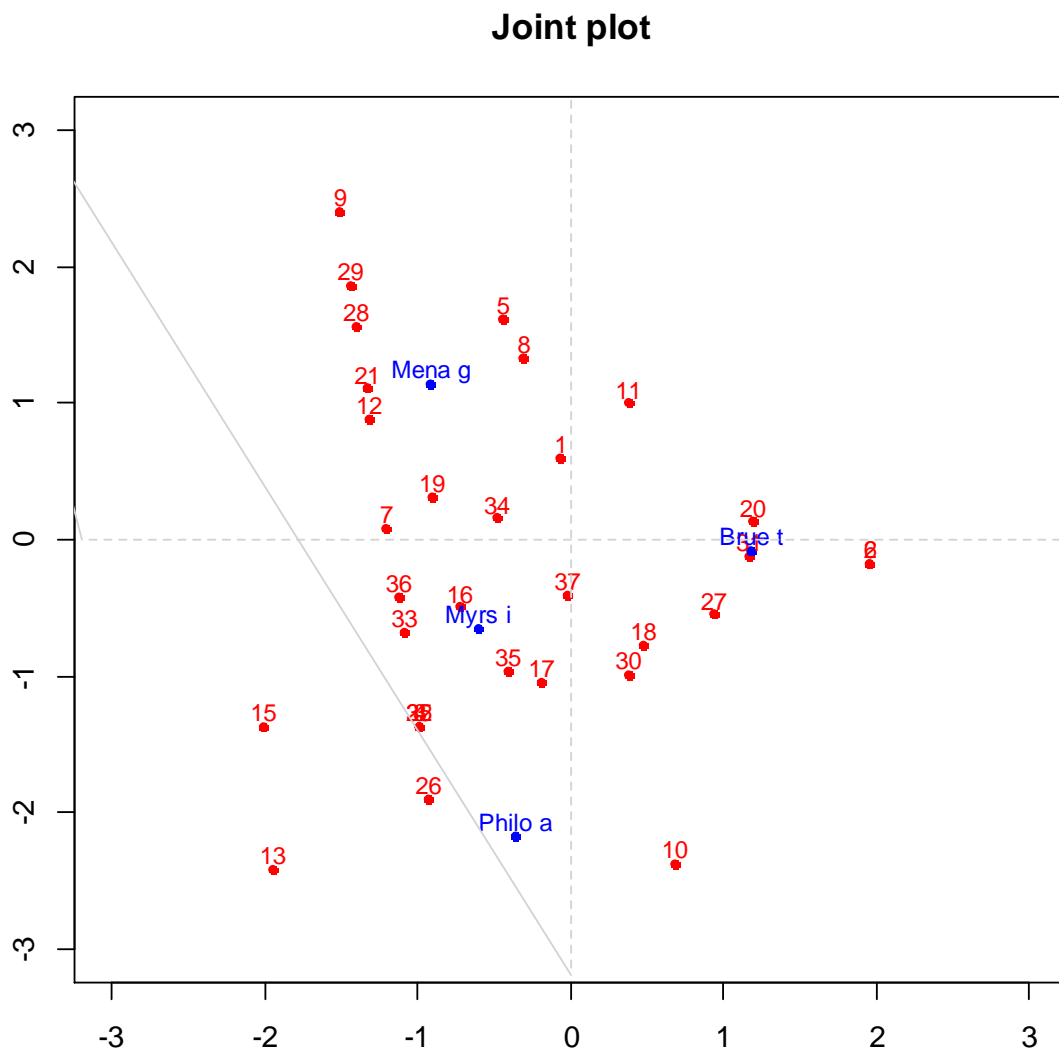
6. ábra

*A varjak tetűparazita abundanciáinak táblázata blokkos csoportosítás előtt és utána*

### 6.1.5 Korrespondenciaanalízis

A béta diverzitások és béta aggregáltságok értékei alapján sejthető tetűfaj- és gazdacsoportokat korrespondenciaanalízis segítségével is megkereshetjük. Ezzel a statisztikai eljárással közös síkbeli (esetleg térbeli) koordinárendszerben pontokkal ábrázoljuk a gazdákat és a parazitafajokat. Azok a gazdák és parazitafajok kerülnek egymáshoz közel, ahol a gazdán az adott faj parazitái a véletlen, esetleges előfordulási arányhoz képest nagyobb arányban fordulnak elő. Az ábrán a gazdák egymástól való távolsága a rajtuk lévő parazitafajok eloszlásának hasonlóságát, a parazitafajok távolsága pedig a gazdák közötti eloszlásuk hasonlóságát mutatja (4.3. fejezet). A varjú–tetűfaj abundanciátáblázatra alkalmazott, az R 3.0.2 szoftver *anacor* eljáráscsomagjával (4.3. fejezet) készített korrespondenciaanalízis grafikus eredményét, a korrespondencia térképet a 7. ábra mutatja. A térképen látható, hogy a gazdák nagy része vagy a *Myrsidea isostoma* és a *Menacanthus gonophaeus* tetűfajok közelében, vagy a *Brueelia tasniamae* tetűfaj környezetében található. A gazdák és a tetűfajok egymástól való távolsága, ill. közelsége nagyjából megfelel a blokk klaszterezés szerinti csoportosításnak. Ez újabb bizonyítékát adja

a béta diverzitások és béta aggregáltságok alapján sejtett gazda- és tetűfaj-csoportok létezésének.



**7. ábra**

*A varjak tetűparazita abundanciáinak korrespondancia térképe*

## 6.2 Dolmányos varjú

### 6.2.1 Adatok, leíró statisztikák

Dolmányos varjúból 89 példány került a mintába. Ezekon a gazdaállatokon összesen 833 tetűegységet találtak, melyek 5 fajhoz tartoznak. Az abundanciák összegének, átlagának, szórásának, variációs együtthatójának és a lineáris zsúfoltságnak (4.1. fejezet) fajok közötti megoszlását mutatja be az 4. táblázat.

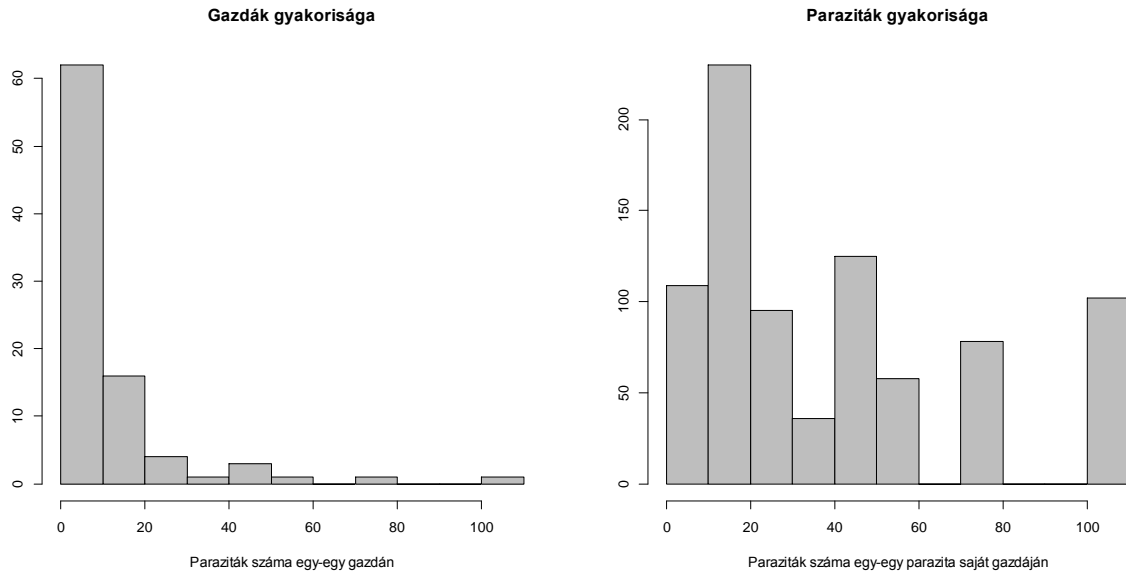
#### 4. táblázat

*Paraziták fajonkénti és együttes prevalenciája, intenzitása, abundanciájának összege, átlaga, szórása, variációs együtthatója, és a fajonkénti lineáris zsúfoltság*

Tetűfaj	Prevalencia (%)	Intenzitás átlaga	Abundancia				Lineáris zsúfoltság
			összege	átlaga	szórása	variációs együtthatója (%)	
<i>Brueelia uncinosa</i>	7.87	2.00	14	0.16	0.60	382.14	2.43
<i>Myrsidea cornicis</i>	32.58	11.83	343	3.85	12.16	315.63	41.82
<i>Menacanthus eurysternus</i>	43.82	11.18	436	4.90	9.73	198.58	24.00
<i>Philopterus ocellatus</i>	7.87	5.57	39	0.44	1.66	378.37	6.64
<i>Colpocephalum fregili</i>	1.12	1.00	1	0.01	0.11	943.40	1.00
<i>Együttes</i>	56.18	16.66	833	9.36	16.84	179.91	39.31

Az 5 tetűfaj közül 2 domináns, ezek a *Myrsidea cornicis* és a *Menacanthus eurysternus* fajok. A 2 domináns tetűfaj adja az együttes abundancia 94%-át. Az abundanciák szórása nagy, minden fajra és együtt is jóval meghaladja az átlagos abundanciát. A variációs együttható 200-400%-os. (A különösen ritka *Colpocephalum fregili* fajt nem számítva.) Ez arra utal, hogy a fajonkénti és együttes parazitáltság erősen aggregált a dolmányos varjakon. A lineáris zsúfoltság minden tetűfajra és együttesen is jóval nagyobb, mint az átlagos abundancia, emellett a fajonkénti zsúfoltságok nagyságrendi sorrendje nem egyezik az átlagos abundanciákéval. Legnagyobb átlagos abundanciája a *Menacanthus eurysternus* fajnak van, míg a legnagyobb zsúfoltság a *Myrsidea cornicis* fajt jellemzi.

A tetűparaziták számának eloszlását vizsgálhatjuk gazdaállatok között és a paraziták szempontjából (5. fejezet, 1. Állítás). Az első szemléletnek megfelelően varjúegyedenként megállapítjuk a tetvek számát és ezek hisztogramját készítjük el. A második megközelítésben tetűegyedenként az egyed gazdáján lévő összes tetű számából állítunk össze hisztogramot. A két hisztogram a 8. ábrán látható.



**8. ábra**

*Tetűparaziták számának hisztogramja a gazdák és a paraziták szemszögéből*

A baloldali hisztogram a különbözőképpen parazitált gazdák gyakoriságát ábrázolja. A hisztogram jobbra ferde, mutatva, hogy a gazdákon a tetvek eloszlása aggregált, azaz a tetvek jelentős része viszonylag kevés varjún él. A 833 egyedből álló tetűkolóniából 724 egyed (87%) összesen 27 varjúról (30%) gyűjtöttek be. A jobboldali hisztogram a különböző mértékben parazitált gazdájú tetvek gyakoriságát mutatja be. Az eloszlás sokkal kevésbé jobbra ferde, csaknem egyenes, ami azt jelenti, hogy egy véletlenszerűen kiválasztott tetű durva közelítésben ugyanolyan eséllyel él kevésbé és nagyon fertőzött gazdán.

**6.2.2 Zsúfoltságok additív felbontása**

A tetűparaziták fajonkénti átlagos zsúfoltságát, átlagos gazdánkénti szubjektív diverzitását és ezek összegét, az együttes zsúfoltságot tartalmazza az 5. táblázat a Lloyd, lineáris, logaritmikus és hiperbolikus zsúfoltságokra (4.1. és 4.2. fejezet).

## 5. táblázat

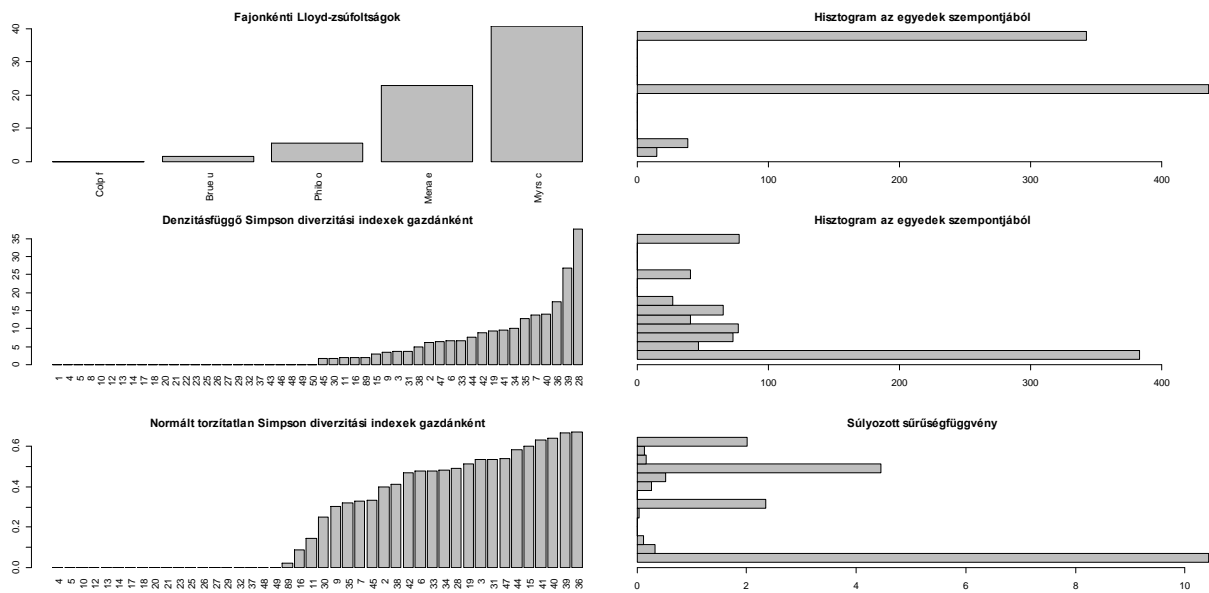
### Zsúfoltságok additív összetevői

	Fajonkénti átlagos (eredeti skálán)	Szubjektív diverzitás		Együttes (eredeti skálán)
		Denzitásfüggő	Normált (%)	
Lloyd	29.13 (30.13)	9.18	23.96	38.31 (39.31)
Lineáris	30.13 (30.13)	9.18	23.35	39.31 (39.31)
Logaritmusos	2.88 (17.89)	0.42 (1.52*)	12.74	3.31 (27.26)
Hiperbolikus	0.90 (10.04)	0.04	4.21	0.94 (16.66)

\*Eredeti exponenciális skálán, szorzatmodellben.

A Lloyd és lineáris zsúfoltságok értéke – mint mindig – csak egy egységnyivel tér el egymástól. A domináns tetűfajok és az erősebben parazitált varjak részarányukhoz képest nagyobb mértékben befolyásolják ezeket a zsúfoltságokat, ennek megfelelően értékük a többihez képest nagyobb. Ez a megállapítás igaz a zsúfoltságok eredeti, parazitaszámra vonatkozó skáláján is. Az együttes, nem fajspecifikus zsúfoltságban a szubjektív diverzitás részaránya nem túl magas, mintegy 23-24%-os. A lineáris zsúfoltság felbontása szerint egy tetű gazdján átlagosan 30.13 saját fajhoz tartozó és 9.18 más fajhoz tartozó tetűegyed van. A logaritmusos zsúfoltságban a tetűfajok éppen részarányuknak megfelelő, a varjak pedig parazitaszámuknak megfelelő arányban súlyozottak, ezért a kisebb fajok és a kevésbé parazitált gazdák szerepe kifejezettebb a Lloyd és a lineáris zsúfoltsághoz képest. A logaritmusos zsúfoltság esetében a szubjektív diverzitás részaránya 12.74%, így az átlagos fajonkénti zsúfoltság súlya nagyobb az additív felbontásban, mint a Lloyd vagy a lineáris zsúfoltság esetében. Az értelmezés az eredeti skálán egyszerűbb, itt szorzatmodellt kapunk. Eredeti skálán a fajonkénti zsúfoltság 17.89, ezt szorozzuk az átlagos Shannon effektív fajszámmal, 1.52-vel. Az eredmény 27.26, ami eredeti skálán az együttes zsúfoltság. A hiperbolikus zsúfoltság a ritka fajokat és a kevésbé parazitált gazdákat arányukhoz képest felülreprezentálja. A zsúfoltságok – mint általában – csak kevéssel maradnak 1 alatt, ezért 1-től való eltérésüket érdemes vizsgálni (4.2. fejezet, (22)). A fajonkénti zsúfoltsághoz képest az együttes zsúfoltság 1-től való eltérése csaknem felére csökkent, ezért az átlagos szubjektív diverzitás ebben a megközelítésben is jelentősen befolyásolja az együttes zsúfoltságot. Ezt a megállapítást az eredeti skálás fajonkénti és együttes zsúfoltságok nagy eltérése is alátámasztja.

A zsúfoltság additív felbontásában szereplő tagok átlagok, érdemes ezért az összetevőik eloszlását, homogenitását vagy heterogenitását grafikusán megvizsgálni. A Lloyd zsúfoltság összetevőinek eloszlásait a 9. ábrán elemezzük.



9. ábra

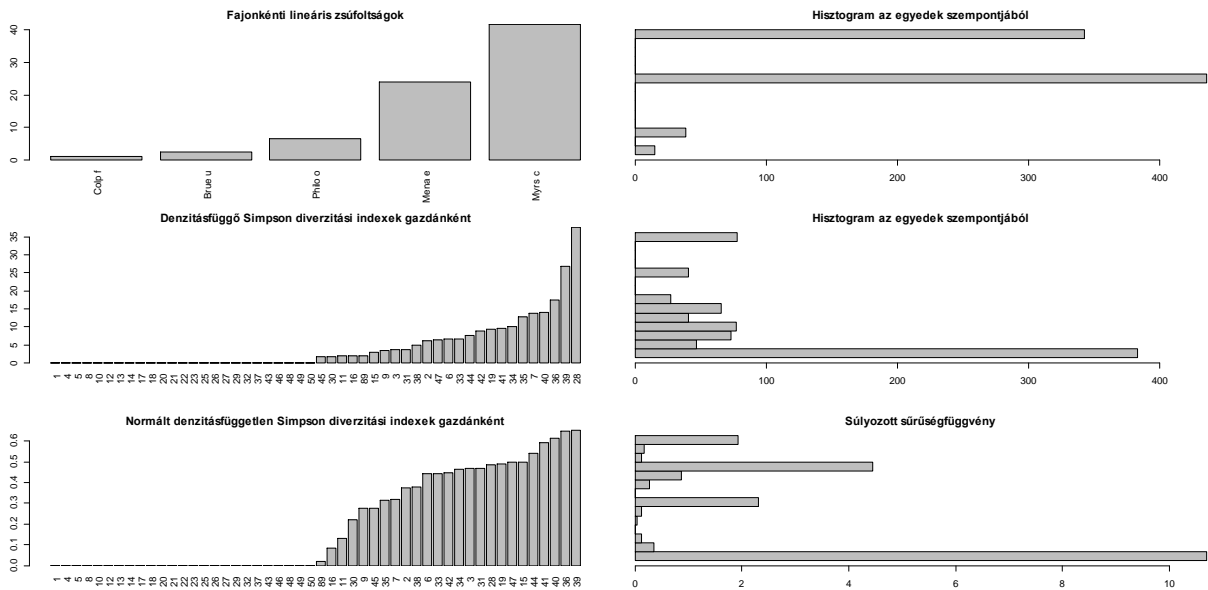
*Lloyd zsúfoltság összetevőinek eloszlása*

Az első sor két ábrája a fajonkénti zsúfoltságokat és ezek egyedszempon-tú gyakoriságait jeleníti meg (4.2. fejezet, (9)). Annak a 2 fajnak a zsúfoltsága és egyedszempon-tú gyakorisága domináns, amelyeknek az átlagos abundanciája is a legnagyobb (*Myrsidea cornicis* és *Menacanthus eurysternus*). Az átlagos fajonkénti zsúfoltság leginkább meghatározó, legnagyobb zsúfoltságú összetevője a *Myrsidea cornicis* tetűfaj, annak ellenére, hogy átlagos abundanciájánál a *Menacanthus eurysternus* átlagos abundanciája nagyobb.

A második sor ábrái a gazdánkénti denzitásfüggő Simpson diverzitási indexek nagyság szerint növekvő sorrendben és ezek egyedszempon-tú gyakorisági hisztogramja (4.2. fejezet, (9)). Jelentős a kis szubjektív diverzitású gazdák szerepe. Ezekon kevés a parazita és/vagy a kis fajszám miatt kicsi a lokális Simpson-diverzitás (a következő bekezdésben látni fogjuk, hogy körükben jellemző a kis parazita fajszám). A 28-as és a 39-es sorszámú gazdák kiugróan nagy szubjektív diverzitást mutatnak. A gazdák harmadik, jelentős súlyú, elkülöníthető csoportja a középmezőnybe tartozik.

A harmadik sor első ábrája az egységre normált gazdánkénti Simpson diverzitási indexeket ábrázolja növekvő sorrendben. A mellette lévő ábra a diverzitási indexek súlyozott sűrűségfüggvénye (4.2. fejezet, (11)). A nulla vagy kis diverzitású gazdák súlya jelentős, valamivel nagyobb, mint a denzitásfüggő diverzitás esetében. Ez azt jelzi, hogy számottevő a kevés tetűfajjal fertőzött varjak súlya. Meghatározó a szerepe a 0.3 körüli és a 0.5-0.6 diverzitású varjaknak. Mások mellett ez utóbbihoz tartozik az előző bekezdésben kiemelt 28-as és 39-es sorszámú gazda.

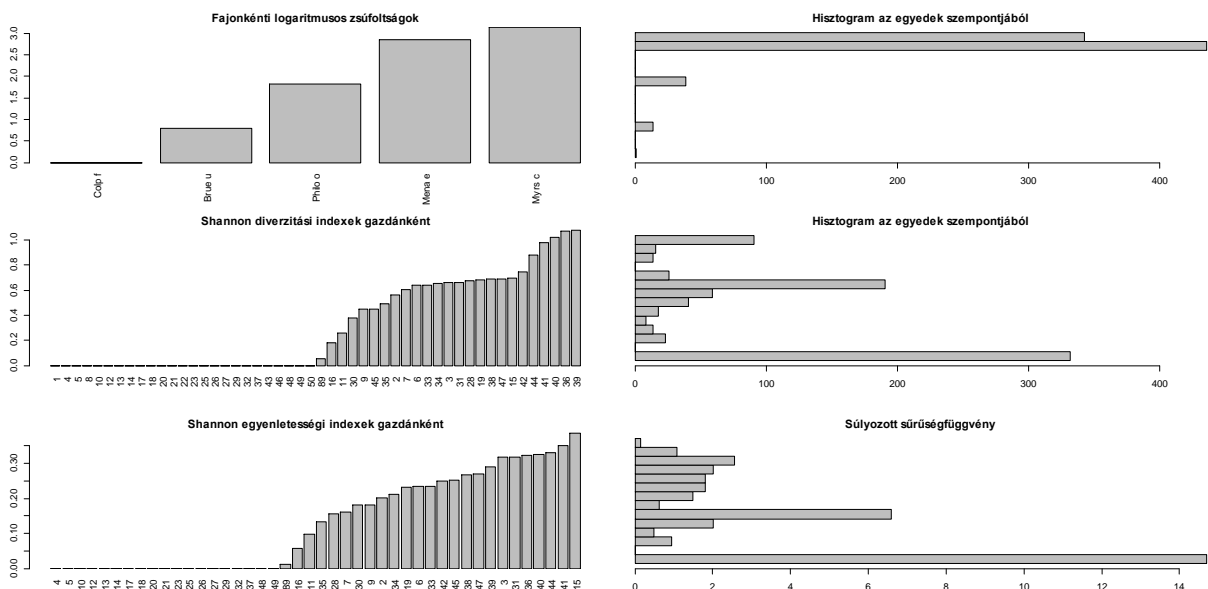
A lineáris zsúfoltság összetevőit a 10. ábrán tanulmányozhatjuk. A kép szinte semmiben se különbözik a Lloyd zsúfoltság szerkezetétől.



10. ábra

*Lineáris zsúfoltság összetevőinek eloszlása*

A logaritmusos zsúfoltságok felbontását és eloszlás szerkezetét a 11. ábrán vizsgáljuk meg.



11. ábra

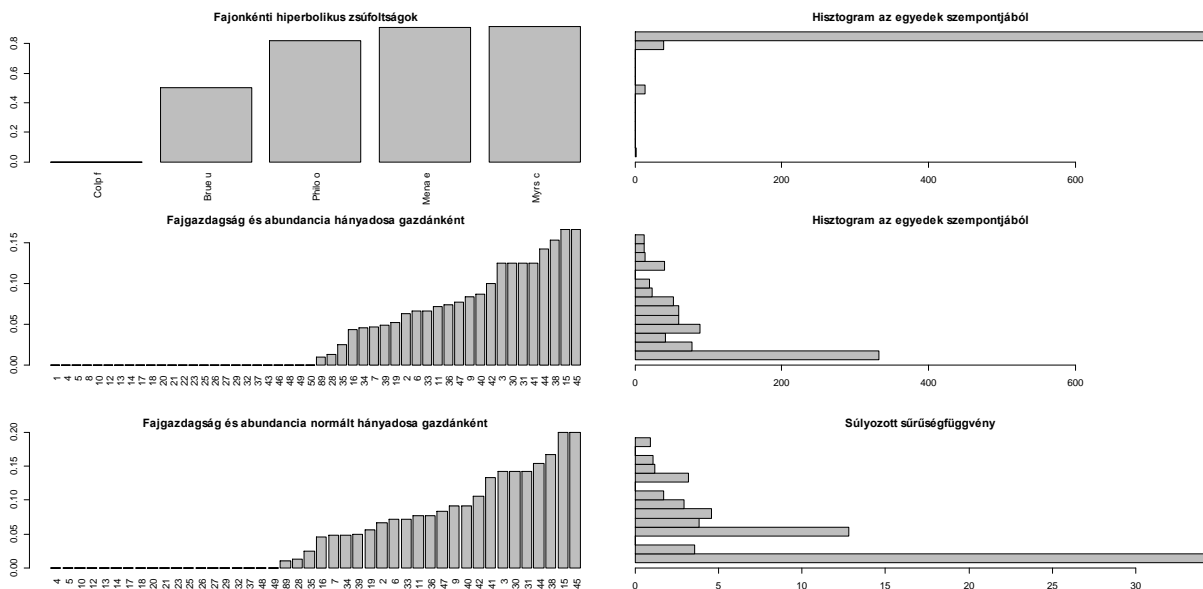
*Logaritmusos zsúfoltság összetevőinek eloszlása*

A fajonkénti zsúfoltságokban a két alacsonyabb átlagos abundanciájú parazitafaj, a *Brueelia uncinosa* és a *Phlopterus ocellatus* relatíve nagyobb értékkel jelenik meg, mint a Lloyd és a lineáris zsúfoltság esetében. (A *Colpocephalum fregili* arányaiban továbbra is jelentéktelen,



ui. összesen 1 egyedét gyűjtötték belőle.) A második sorban a szubjektív Shannon diverzitások vannak, melyek denzitásfüggetlenek (4.2. fejezet, (16)). A gazdák három jól elkülönülő csoportba oszthatók, a 0 diverzitásúak (azaz 1 tetűfajjal fertőzöttek), a 0.4-0.6 diverzitásúak és a 0.8-1.0 diverzitásúak. Az ábra harmadik sorában az együttes zsúfoltság szubjektív diverzitáshoz kapcsolható százalékarányát a gazdánkénti Shannon egyenletességi indexek megoszlásával szemléltetjük (4.2. fejezet, (18)). Az egyenletességi indexek eloszlása is három homogén csoport keveréke, ezek a 0 diverzitású (1 tetűfajjal fertőzött), a 0.05-0.15 egyenletességi indexű és a 0.15-0.40 egyenletességi indexű gazdák. Az egyenletességi index a Shannon diverzitásból a parazitaszám logaritmusával való normálással számítható ki, ez magyarázza, hogy a diverzitáshoz képest az eloszlása némileg koncentráltabb.

A hiperbolikus zsúfoltság összetevőinek szerkezetét a 12. ábra mutatja.



12. ábra

*Hiperbolikus zsúfoltság összetevőinek eloszlása*

A fajonkénti zsúfoltságok értékei között csekély a különbség annak megfelelően, hogy gazdáról gazdára itt csak a tetűfaj jelenléte vagy hiánya számít. A fajonkénti zsúfoltságok egyedszemponjú gyakorisága a 2 domináns fajra és a *Phlopterus ocellatus* fajra együtt kiemelkedően nagy. A szubjektív diverzitás összetevői most gazdánként a fajgazdagság és az abundancia hányadosai (4.2. fejezet, (20)). Ezek közül legnagyobb súlya a 0.0 körüli tartományba eső értékeknek van, ezen kívül a 0.05-0.10 és 0.15-0.20 közötti diverzitás-összetevők is jelentősek. A harmadik ábrsorban bemutatott normált szubjektív diverzitások (4.2. fejezet, (22)) alapján a 0.0-0.2, 0.05-0.10 és a 0.12-0.20 közötti lokális diverzitású

gazdacsoportok határozzák meg az együttes zsúfoltságban a szubjektív diverzitásra jutó százalékarányt.

### 6.2.3 A valódi diverzitás és az aggregáltság multiplikatív felbontása

A tetűfajok varjankénti átlagos alfa valódi diverzitását, a teljes mintából számolt gamma valódi diverzitását és e kettő hányadosát, a  $\beta = \text{gamma}/\text{alfa}$  diverzitást vizsgáljuk. A tényleges gazdaszám és az effektív gazdaszám hányadosaként értelmezett aggregáltság tetűfajonkénti és fajokra együtt vett átlaga az alfa és gamma aggregáltság. Az alfa/gamma hányados a béta aggregáltság. A diverzitások és aggregáltságok három fő típusát, a lineáris, logaritmusos és hiperbolikus változatot alkalmazzuk (4.3. fejezet). E mennyiségek mintából becsült értékeit a 6. táblázat tartalmazza.

#### 6. táblázat

*A valódi diverzitás és az aggregáltság tényezői*

	Valódi diverzitás			Aggregáltság		
	Alfa	Béta	Gamma	Alfa	Béta	Gamma
Lineáris	1.36	1.65	2.24	8.11	1.93	4.20
Logaritmusos	1.52	1.65	2.52	4.82	1.65	2.91
Hiperbolikus	1.96	2.55	5.00	2.71	1.52	1.78

Ha a diverzitás ritkasági függvénye lineáris, akkor a valódi Simpson diverzitást kapjuk eredményül. Az ennek megfelelő lineáris zsúfoltsági függvény alkalmazása a lineáris zsúfoltságra, aggregáltságra vezet. A Simpson diverzitásban számarányukhoz képest felülreprezentáltak a domináns fajok, a lineáris aggregáltságban pedig a domináns, erősen parazitált gazdák (4.3. fejezet). Az alfa diverzitás 1.36, eszerint a gazdánkénti átlagos effektív tetűfaj-szám kb. 1-2. A gamma diverzitás 2.24, ennyi a gazdák együttesén, a teljes tetűkolóniában az effektív tetűfaj-szám. A béta diverzitás 1.65, ez azt a lehetőséget sugallja, hogy a tetűkolónia fajai két olyan csoportra oszthatók, amelyek egyedei közös gazdán viszonylag ritkán fordulnak elő. Az alfa aggregáltság 8.11, ez nagy érték, jelzi, hogy a tetűfajok külön-külön a domináns gazdák között erősen aggregáltan oszlanak el. A gamma aggregáltság 4.20, ez is nagy érték, mutatja, hogy a tetveket faj szerint el nem különítve a gazdák közötti aggregáltság továbbra is nagy. A béta aggregáltság 1.93, ez utalhat a domináns gazdák két olyan csoportjára, amiken a tetűfajok részarányai markánsan eltérőek.

Ha logaritmus a ritkasági függvény, akkor hozzá a Shannon valódi diverzitás és a logaritmusos aggregáltság kapcsolódik. Az alfa és gamma diverzitások a Simpson diverzitásokhoz képest valamivel nagyobbak, ez megfelel annak, hogy most a domináns

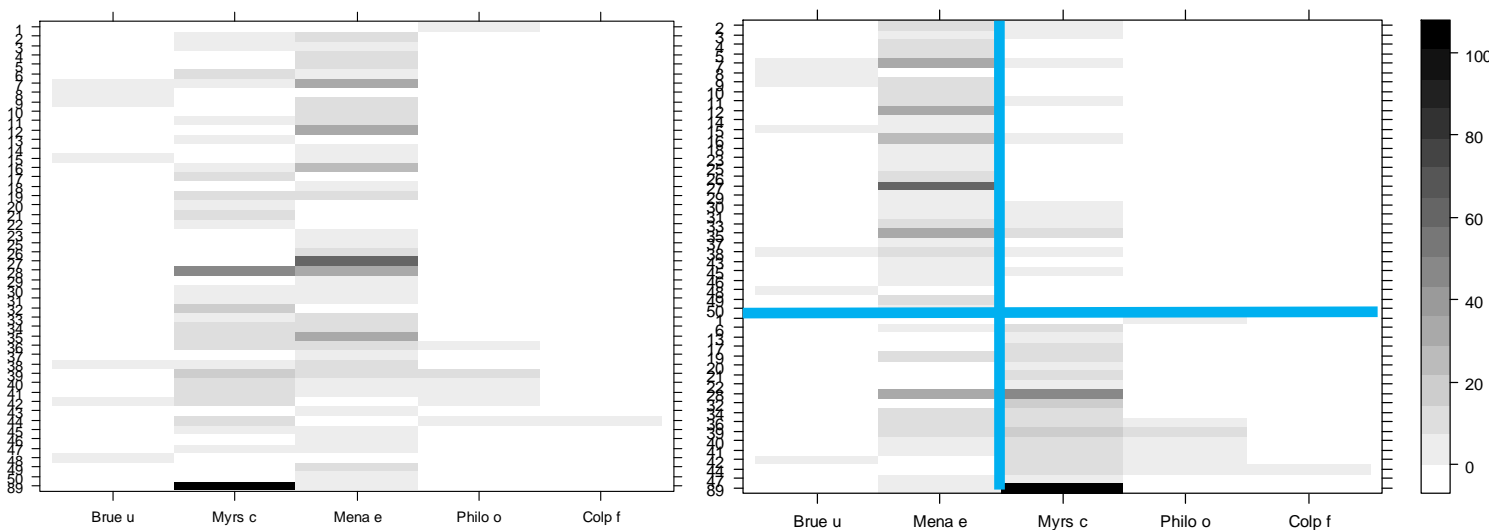
tetűfajok nem nyomják el a többi fajt, a Shannon valódi diverzitásban mindegyik faj a részarányának megfelelő mértékben reprezentált. A béta diverzitás 1.65, amiből a Simpson béta diverzitáshoz hasonló következtetésre juthatunk, vagyis elképzelhető, hogy a tetűkolónia fajai két, gazdánként kevéssé átfedő csoportra oszthatók. A logaritmikus alfa és gamma aggregáltság kisebb, mint a lineáris, eszerint a gazdákat parazitáltságuknak megfelelő súllyal figyelembe véve, a tetűfajok külön-külön és együtt is kevésbé aggregáltak oszlanak meg a gazdák között. A logaritmikus béta aggregáltság értéke – mint minden esetben – pontosan megegyezik a béta diverzitás értékével (4.3. fejezet). A béta aggregáltság most 1.65, valamivel kisebb, mint a lineáris béta aggregáltság. Szintén a gazdák két olyan csoportjának létezését jelzi, melyeken a tetvek fajonkénti részaránya eltérő.

Hiperbolikus ritkasági függvény mellett a valódi diverzitás éppen a tetűfajok száma. Az alfa és a gamma diverzitások itt a legnagyobbak, ami érthető, hiszen a ritkán előforduló fajok ugyanúgy szerepelnek a fajszámban, mint a többi. Ez magyarázhatja azt is, hogy a béta diverzitás ebben az esetben a legnagyobb. Az alfa diverzitás 1.96, a gamma diverzitás 5, ezért gazdánként átlagosan kb. 2, a teljes kolónián 5 tetűfaj fordul elő. A béta diverzitás 2.55, ami most a tetűfajok két vagy három, gazdákon kevéssé átfedő, hipotetikus csoportjára utal. A hiperbolikus alfa és gamma aggregáltságok kisebbek, mint a megfelelő lineáris és logaritmikus aggregáltságok. Ez megfelel annak, hogy a kevéssé fertőzött gazdákat parazitáltságuknál nagyobb mértékben reprezentálja a hiperbolikus aggregáltság. A béta aggregáltság 1.52, ami az előző két béta aggregáltsággal többé-kevésbé szinkronban, a gazdák két, tetűfajonként eltérő arányban parazitált csoportjának létezését prognosztizálja.

#### **6.2.4 Gazdák és paraziták csoportosítása**

A valódi béta diverzitások és a béta aggregáltságok értékei alapján azt a hipotézist fogalmazhatjuk meg, hogy a tetűfajok és a varjak két-két csoportra oszthatók úgy, hogy az első tetűfaj-csoport elsősorban az első varjúcsoporton, a második tetűfaj-csoport inkább a második varjúcsoporton él. Blokk klaszterező eljárással mód van ennek a feltételezésnek az ellenőrzésére. Az eljárás a kérdéses tetűfaj-csoportokat és varjúcsoportokat a gazda–parazita abundanciátáblázat sorainak és oszlopainak egymás utáni cserélgetésével keresi meg oly módon, hogy a nagy abundanciaértékeket egymás közelébe mozgatja (4.3. fejezet). Az áttekinthetőség érdekében a nem fertőzött gazdák abundanciasorai kikerültek a táblázatból, összesen 50 gazda adatai maradtak meg. A nulla tetűabundanciájú gazdák elhagyása nem változtatja meg a valódi diverzitások és az aggregáltságok értékét. A csoportok keresését az így redukált abundanciátáblázaton elvégezve, az eredményt a 13. ábra mutatja.

A bal oldali kép az eredeti elrendezésben, a jobb oldali a blokkos klaszterek szerinti varjú- és tetűfaj-sorrendben ábrázolja az abundanciákat. Az abundanciák értékeit az ábra jobb szélén lévő színskála jelzi. Az eredmény alátámasztja a hipotézist, valóban elkülöníthető két tetűfaj-csoport és nekik megfelelően két varjúcsoport. Az első tetűfaj-csoport domináns faja a *Menacanthus eurysternus*, a csoport másik, jóval ritkább faja a *Brueelia uncinosa*. A második tetűfaj-csoport domináns faja a *Myrsidea cornicis*. A csoporthoz tartozik még a ritka *Philopterus ocellatus* és az igen ritka *Colpocephalum fregili*. A tetűfaj-csoportok a varjaknak csak kis hányadán, 5 egyeden (10%) mutatnak jelentősebb átfedést.



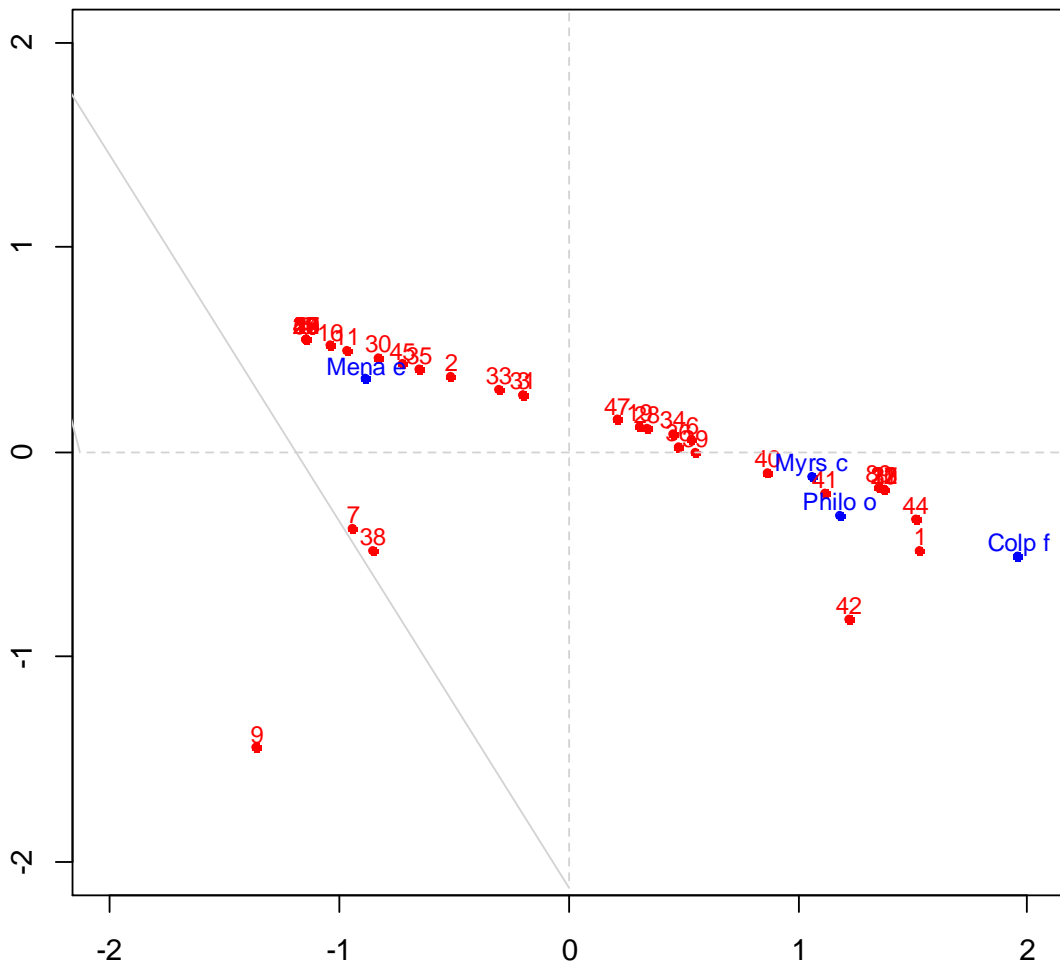
13. ábra

A varjak tetűparazita abundanciáinak táblázata blokkos csoportosítás előtt és utána

## 6.2.5 Korrespondenciaanalízis

A béta diverzitások és béta aggregáltságok értékei alapján sejthető tetűfaj- és gazdacsoportokat korrespondenciaanalízis segítségével is megkeressük. A varjú–tetűfaj abundanciátáblázatra alkalmazott, az R 3.0.2 szoftver *anacor* eljáráscsomagjával (4.3. fejezet) készített korrespondenciaanalízis grafikus eredményét, a korrespondencia térképet a 14. ábra mutatja. A térképen látható, hogy a gazdák nagy része vagy a *Menacanthus eurysternus*, vagy a *Myrsidea cornicis* tetűfajok közelében található. A gazdák és a tetűfajok egymástól való távolsága, ill. közelsége nagyjából megfelel a blokk klaszterezés szerinti csoportosításnak. Ez újabb bizonyítékát adja a béta diverzitások és béta aggregáltságok alapján sejtett gazda- és tetűfaj-csoportok létezésének.

### Joint plot



**14. ábra**

*A varjak tetűparazita abundanciáinak korrespondancia térképe*

### 6.3 A vetési varjú és a dolmányos varjú parazitáltságának összehasonlítása

Az adatmintában mindkét varjófajon ugyanazt az 5 tetűgénuszt (Brueelia, Myrsidea, Menacanthus, Philopterus, Colpocephalum) találták meg. Összességében a vetési varjakon az átlagos tetűabundancia és a zsúfoltság minden génuszra nagyobb, mint a dolmányos varjakon. A génuszok abundanciaeloszlása a két varjófajon annyiban hasonló, hogy a Myrsidea és Menacanthus génuszok mindkét varjófajon dominánsak, átlagos abundanciájuk és zsúfoltságuk nagy, a Philopterus és Colpocephalum génuszok viszont ritkák. A vetési varjakon a Brueelia génusz is domináns, ennek van a mintába került vetési varjakon az 5 génusz közül a legnagyobb az átlagos abundanciája és zsúfoltsága. A dolmányos varjakon ezzel szemben a Brueelia génusz átlagos abundanciája és zsúfoltsága kicsi (1. és 4. táblázat).

A tetvek gazdák közötti eloszlása mindkét varjófaj esetében nagyon aggregált. A tetűkolónia jelentős része a gazdák aránylag kis részére koncentrálódik. A tetűegyedek szempontjából ugyanakkor a saját gazdák parazitászama sokkal kevésbé aggregált, szinte egyenletes eloszlású (1. és 8. ábra, 5. fejezet 1. állítás).

Az együttes zsúfoltság additív felbontásában a szubjektív diverzitás komponense a vetési varjakon nagyobb arányú, mint a dolmányos varjakon. A Lloyd és lineáris zsúfoltság esetében ez az arány vetési varjakra kb. 56%, dolmányos varjakra 23-24%. A logaritmusos zsúfoltságra a megfelelő arányok 17% és 13% (2. és 5. táblázat). A varjófajok közötti eltérést magyarázhatja, hogy a fertőzött dolmányos varjak körében nagyobb arányú az egyetlen tetűfajjal fertőzöttek, így 0 diverzitásúak aránya (50%), mint a fertőzött vetési varjakon (21%).

A szubjektív diverzitás komponensében mindkét varjófaj esetében elkülöníthető a közepesen és a nagyon diverz egyedek gazdacsoportja is. A 0 diverzitású csoporttal együtt ezért mindkét varjófajnál 3 gazdacsoport állapítható meg. A csoportok gyakoriságai és súlyai a dolmányos varjaknál tisztábban elhatárolódnak, kevésbé fedik át egymást, mint a vetési varjaknál (2-5. és 9-12. ábrák).

A tetvek valódi diverzitásai a vetési varjak esetében valamivel nagyobbak, mint a dolmányos varjaknál. A béta diverzitások a két varjófajnál nagyjából megegyeznek, értékük 1.5-2.5 között van. A tetűfajok átlagos, azaz alfa aggregáltsága a dolmányos varjak körében nagyobb, mint a vetési varjaknál. Ez azt jelenti, hogy a dolmányos varjak tetűfajai átlagosan aggregáltabban oszlanak meg a gazdaállatok között, mint a vetési varjak tetűfajai. A béta aggregáltságok a dolmányos varjaknál 2-höz közelebbi értékűek, mint a vetési varjaknál (3.

és 6. táblázat). A gamma aggregáltság is nagyobb a dolmányos varjakon, az eltérés azonban kisebb, mint az alfa aggregáltság esetében.

A béta aggregáltságok értéke mindkét varjófajnál nagyjából 1 és 2 közötti, ez alapján a vetési varjak génuszait két, gazdákon kevéssé átfedő csoportra, a *Brueelia*, illetve a *Myrsidea* és *Menacanthus* együttesére bontottuk blokk klaszterezéssel és korrespondanciaelemzéssel (6-7. és 13-14. ábra). A dolmányos varjakon a két domináns génusz, a *Myrsidea* és a *Menacanthus* alkot a gazdákon csak kismértékű átfedést mutató csoportot (6-7. és 13-14. ábra). Érdekesség, hogy a vetési varjakon a dolmányos varjakkal ellentétben a *Myrsidea* és *Menacanthus* génuszok nem különülnek el, gyakran előfordulnak közös gazdán.

## **7 A szarvaskullancslégy (*Lipoptena cervi*) ivararányának és a gímszarvas gazda (*Cervus elaphus*) parazitáltságának kapcsolata**

Egyszerre egyedszemponútú és csoportszintű regressziós modell alkalmazásával megvizsgáljuk a gímszarvas (*Cervus elaphus*) gazdákat parazitáló szarvaskullancslegyek (*Lipoptena cervi*) ivararánya és a parazitáltság foka közötti kapcsolatot. Szarvaskullancslegyek gímszarvasokon való előfordulásáról Rigó Eszter gyűjtött adatokat szakdolgozatához, melynek konzulense Prof. dr. Egri Borisz volt (Rigó 2011).

### **7.1 Adatok**

A Hanságban 2010 novembere és 2011 novembere között terítékre hozott gímszarvasok nyakának jobboldali tájékaról begyűjtötték az összes ott talált szarvaskullancslegyet. Feljegyezték a gímszarvasok elejtésének helyét és idejét, az állatok ivarát és korát, a rajtuk talált szarvaskullancslegyek számát ivar szerinti bontásban (Rigó 2011).

### **7.2 Statisztikai módszerek**

Random hatásos logisztikus regresszióval vizsgáltuk a hím kullancslegyek gímszarvasonkénti részarányát (Agresti 2002). A célváltozó az, hogy egy gímszarvason véletlenszerűen kiválasztott szarvaskullancslégy hím. Ennek valószínűségét modelleztük. A random szubjektum a gímszarvas, a modell fix kovariánsai a szarvaskullancslegyek száma a gímszarvason és az évszak a gímszarvas elejtésének időpontjában (ősz, tél). A gazdánkénti szarvaskullancslegyek számának magyarázó változóként való szerepeltetésével tártuk fel a szarvaskullancslegyek ivararánya és a gímszarvasok parazitáltsága közötti kapcsolatot. Ez a magyarázó változó egyúttal biztosítja, hogy a regressziós modell ugyanaz a szarvaskullancslegyek és a gímszarvas gazdák szintjén (5. fejezet 10. állítás). A modellezésből kizártuk azokat a gímszarvasokat, amiken nem volt szarvaskullancslégy. Az ivararány és a parazita nemek részaránya ui. rajtuk nem értelmezhető. A számításokat az R 3.0.2 statisztikai szoftver *lme4* programcsomagjával végeztük el (R Core Team 2014, Bates et al. 2013).

A logisztikus regressziós modellből kiszámítottuk a hímek részarányainak gazdánkénti és parazitánkénti átlagát. A modellből minden gímszarvasra kaptunk egy-egy predikciót és ezek átlagát vettük. Az általános számítási képlet és levezetése az 5. fejezet 11. állításában szerepel. A gímszarvasok szintjén és szarvaskullancslegyek szempontjából átlagolt



valószínűségeket összehasonlítottuk egymással. Összevetettük ezen kívül a mintabeli részarányoknak a két szempont szerinti egyszerű átlagait is. Percentilis és BCa bootstrap konfidenciaintervallumokat becsültünk a hímek valószínűségének gazdánkénti és szarvaskullancslegyenkénti populációs átlagára és ezek különbségére (Davison és Hinkley 1997). Összesen 2000 bootstrap mintát generáltunk, mindegyiken újraillesztettük a kevert hatásos logisztikus regressziós modellt, majd kiszámítottuk a modell szerinti valószínűségek és a nyers arányok gazdánkénti és parazitánkénti átlagát, ezek különbségét. Az így kapott 2 x 2000 tripletból jöttek ki a bootstrap konfidenciaintervallumok. A bootstrap becsléseket az R 3.0.2 statisztikai szoftver *boot* programcsomagjával végeztük el (R Core Team 2014, Canty és Ripley 2013).

### 7.3 Leíró statisztikák

Összesen 31 gímszarvason 1994 szarvaskullancslegyet gyűjtöttek. A paraziták abundanciája összegének, átlagának, szórásának, variációs együtthatójának és a lineáris zsúfoltságnak (4.1. fejezet) az ivarok közötti megoszlását mutatja be az 1. táblázat. Részletesebb és teljesebb körű leíró statisztikák találhatóak a (Rigó 2011) közleményben.

#### 1. táblázat

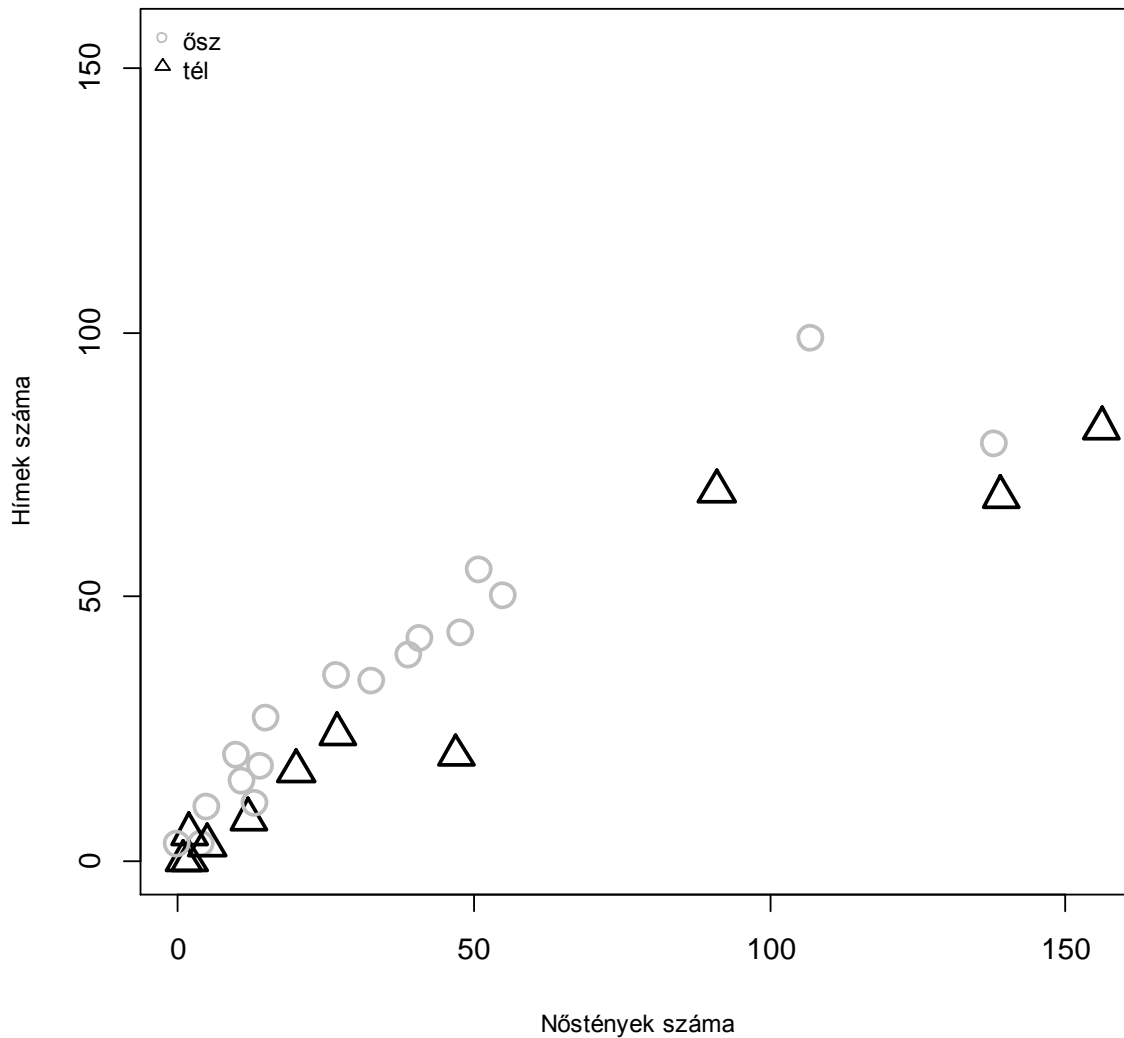
*Szarvaskullancslegyek ivaronkénti és együttes prevalenciája, átlagos intenzitása, abundanciájának összege, átlaga, szórása, variációs együtthatója, és az ivaronkénti lineáris zsúfoltság*

Ivar	Prevalencia (%)	Intenzitás átlaga	Abundancia			Lineáris zsúfoltság	
			összege	átlaga	szórása		
L. cervi ♀	87.10	41.22	1113	35.90	44.64	124.34	89.62
L. cervi ♂	83.87	33.88	881	28.42	28.11	98.92	55.34
<i>Együttes</i>	<i>90.32</i>	<i>71.21</i>	<i>1994</i>	<i>64.32</i>	<i>71.45</i>	<i>111.08</i>	<i>141.12</i>

Összességében megállapítható, hogy a nőstény szarvaskullancslegyek abundanciájának átlaga, szórása, variációs együtthatója és a lineáris zsúfoltság értéke mind nagyobb a hímekénél. A variációs együttható 100% körüli vagy inkább nagyobb, ami mindkét ivar esetében és együtt is a gazdánkénti abundanciák aggregált eloszlását jelzi. Ugyanerre utal az, hogy a zsúfoltságok kb. kétszer nagyobbak az átlagos abundanciánál. (A 4.1. fejezetben láttuk, hogy a lineáris aggregáltság éppen a lineáris zsúfoltság és az átlagos abundancia hányadosa.)

A szarvaskullancslegyek ivararánya és a gímszarvasok parazitáltsága közötti kapcsolat grafikus vizsgálatához szóródási diagramon ábrázoltuk a nő- és hímivarú

szarvaskullancslegyek gímszarvasonkénti abundanciáját, megkülönböztetve a gímszarvas elejtésének évszakát (ősz, tél) is (1. ábra).



**1. ábra**

*Nőtény és hím szarvaskullancslegyek gímszarvasonkénti abundanciája*

Az 1. ábráról megállapítható, hogy a szarvaskullancslegyek abundanciájának növekedésével párhuzamosan az ivararány a nőstények irányába tolódik el. Télen az őszhöz képest az ivararány szintén elmozdul a nőstények felé.

## 7.4 A szarvaskullancslegyek ivararányának és abundanciájának kapcsolata

Az ivararányt a hím szarvaskullancslegyek arányával számszerűsítettük. Az ezt modellező random hatásos logisztikus regresszió eredménye szerint a hím szarvaskullancslegyek aránya negatív kapcsolatban van a szarvaskullancslegyek abundanciájával. Tízfel több parazita esetén annak esélye, oddsza, hogy egy szarvaskullancslégy hím, 0.97-szeresre csökken. A szorzó értéke 1-nél szignifikánsan kisebb ( $p=0.0001$ ), a 95%-os konfidenciaintervallum [0.96, 0.99]. Az évszaknak szintén szignifikáns a kapcsolata az ivararányal, télen annak oddsza, hogy egy véletlenszerűen választott szarvaskullancslégy hím, 0.69-szeres az őszhöz képest. A szorzó értéke 1-nél szignifikánsan kisebb ( $p=0.0004$ ), a 95%-os konfidenciaintervallum [0.57, 0.85]. A modell random hatású tényezői a gazdaállatok. Az oddszok gímszarvasonkénti szorzótényezője (a szarvasnak, mint szubjektumnak a hatása) átlagosan 0.94 és 1.06 között van, ami átlagosan  $\pm 6\%$  ingadozást jelent.

Az illesztett regressziós modellből kiszámítottuk a hímek gazdánkénti részarányának gazdaszintű és parazita szempontú (parazitákra vett) átlagát. A két átlag nem szükségképpen egyenlő annak ellenére, hogy a regressziós modell a gímszarvasok és a szarvaskullancslegyek szintjén egybeesik (5. fejezet 11. állítás). Összevetettük az átlagos modellezett részarányokat és a mintabeli arányok egyszerű gazdaszintű és parazita szempontú átlagait is. Az eredményeket a 2. táblázatban foglaltuk össze.

### 2. táblázat

*A hím szarvaskullancslegyek átlagos részarányainak becslései és 95%-os bootstrap konfidenciaintervallumai.*

	Szempont	Prevalencia	Percentilis CI		BCa CI	
Regressziós modell	Gímszarvas	0.487	0.447	0.529	0.446	0.527
	Szarvaskullancslégy	0.442	0.405	0.490	0.398	0.481
	Különbség	0.046	0.008	0.080	0.012	0.085
Mintabeli részarány	Gímszarvas	0.472	0.396	0.548	0.397	0.544
	Szarvaskullancslégy	0.442	0.405	0.490	0.398	0.481
	Különbség	0.030	-0.045	0.095	-0.039	0.102

A hím szarvaskullancslegyek részarányának gímszarvasokra vett átlagai mindkét módszerrel nagyobbak, mint a parazitákra számolt átlagok. A gímszarvasok szempontjából, vagy

másképp mondva, a gímszarvasok populációján más a szarvaskullancslegyek ivararánya, mint a szarvaskullancslegyek szintjén, populációján. Az előbbin 5%-os szinten az ivararány nem különbözik szignifikánsan az 1:1 aránytól, az utóbbin viszont az ivararány szignifikánsan eltolódik a nőstények irányába, ui. a hímek átlagos részarányának 95%-os bootstrap konfidenciaintervallumai 0.5 alatt maradnak. A tipikus gímszarvason a szarvaskullancslégy-paraziták ivararánya 1:1 körüli, míg a tipikus szarvaskullancslégy gazdáján a hím paraziták részaránya kisebb a nőstényekénél.

A regressziós modellből nyert, gímszarvasokra átlagolt részarány konfidenciaintervallumai rövidebbek, pontosabbak, mint a nyers arányok átlagához tartozó konfidenciaintervallumok. A szarvaskullancslegyekre (parazitákra) átlagolt részarány modellből számított és a mintából egyszerű átlagolással kapott becslései is, konfidenciaintervallumai is szinte teljesen megegyeznek. Összességében a regressziós modellből származó becslések pontossága legalább olyan jó lett, mint a nyers arányok átlagai – ez a várakozásoknak megfelelő (5. fejezet 16. állítás).

## 7.5 Megvitatás

A gímszarvasok és szarvaskullancslégy-parazitáik populációin más az átlagos parazita ivararány. Ha ezt a különbséget nem vesszük tekintetbe és a szokásos eljárással csupán a gazdaállatokra átlagolt értékekre támaszkodunk, akkor a biológiai értelmezés nem lesz megalapozott és téves következtetésekre vezethet.

A kapott regressziós modell ugyanaz, ha a megfigyelés egysége a gímszarvas, és ha a megfigyelés egysége a szarvaskullancslégy. Az első esetben a regressziós összefüggés a gímszarvasokra, mint egymástól különböző entitásokra vonatkozik. Ebben a csoportszintű megközelítésben mondhatjuk, hogy ha egy gímszarvason 10-zel több parazita van, akkor a hím paraziták esélye 0.97-szeresre csökken. A második esetben az individuum a szarvaskullancslégy, a regresszió a kullancslegyek saját gazdáit, csoportkörnyezetét modellezi. Az egyedszempontrú értelmezés szerint, ha egy szarvaskullancslégy saját gazdáján 10-zel több a parazita, akkor a gazdán a hím paraziták oddsza 0.97-szeresre változik.

A szarvaskullancslegyek gímszarvasonkénti számának felvétele a regressziós modell kovariánsai közé önmagában nem feltétlenül biztosítja a gazdaszintű és parazita szempontú modellek egybeesését. Ehhez még szükség lehet a parazitaszám és a modell egyéb magyarázó változói közötti interakciók modellbe ágyazására, beleértve a gímszarvassal, mint random szubjektummal való interakciót is. Az adatmintára illesztés után azonban ezek a kiterjesztett modellek nem különböztek szignifikánsan az interakciókat nem tartalmazó

modelltől. A modellek között az Akaike-féle információs kritérium (Akaike 1974) alapján sem találtunk lényeges különbséget, ezért az interakciós tagokat a modellből eltávolítottuk.

A szarvaskullancslegyeket az elejtett gímszarvasok nyakának jobboldali részéről gyűjtötték (Rigó 2011), ezért valójában ez a terület azonosítható az 1.4 fejezetbeli egyedcsoporttal. A kimutatott eredmények szorosan véve a gímszarvasoknak csak erre a testtájékára érvényesek, nem feltétlenül általánosíthatók az egész testfelületükre, az ott lévő összes szarvaskullancslégyre.

A gímszarvasok jobboldali nyakrészének átlagos felszíne lényegesen különbözött a bikáknál, teheneknél és a borjaknál, ezért egy módosított regressziós modellbe az egységnyi felületre jutó paraziták számát vettük be magyarázó változóként. Ez a modell lényegében ugyanolyan eredményt adott, mint az eredeti, viszont arra a nem ellenőrizhető feltételezésre épült, hogy a parazitaszám szoros kapcsolatban van a nyakrész felszínének nagyságával. Ezért ezt a megközelítést elvetettük.

A szarvaskullancslegyek hímjeinek téli megritkulása ismert jelenség. A kullancslegyek (Hippoboscidae) körében megfigyelték, hogy a mérsékelt és hideg égövben a hímek aránya az év vége felé haladva lecsökken (Marshall 1981). A hímek arányának téli csökkenése és az arány parazitáltságtól való függése között nem mutattunk ki interakciót, az illesztett regressziós modellben az évszak és a parazitaszám hatása egymástól független.

Végezetül megemlíjtük, hogy a gímszarvasok mintába kerülése nem tekinthető teljesen véletlenszerűnek, különösen az évszakok vonatkozásában. A mintába jutott tíz bika mindegyikét ősszel ejtették el (Rigó 2011). A minta szelektált jellege valamelyest torzíthatja az eredményeket.

## 8 Új tudományos eredmények

1. Új módszert ismertettünk a valódi prevalencia közelítő konfidenciaintervallumának megszerkesztésére abban az esetben, amikor a diagnosztikai teszt szenzitivitását és specifikusságát független mintákból becsüljük.
2. Számítógépes szimulációs kísérletekkel igazoltuk, hogy az új konfidenciaintervallum a jelenleg elfogadott Wald-Rogan-Gladen konfidenciaintervallumnál pontosabban tartja meg az előírt megbízhatósági szintet még kicsi, 30 elemű mintákra is; a lefedés legalább 88%, 93% és 98%, ha az előírt szint rendre 90%, 95% és 99%.
3. Elméleti úton bizonyítottuk, hogy a valódi prevalencia Rogan-Gladen pontbecslése az új konfidenciaintervallumba esik, ha a diagnosztikai eljárás becsült szenzitivitása és specifikussága legalább 0.5, és a mintanagyság a szenzitivitásra  $\geq 26$ , a specifikusságra  $\geq 26$ , a célpopuláció esetében  $\geq 16$ .
4. Az új konfidenciaintervallum-szerkesztő eljárás előnyeit valós megfigyelések publikált adatainak újbóli feldolgozásával illusztráltuk.
5. Szimulációs kísérletekkel vizsgáltuk a pontosság romlását mintákból becsült szenzitivitás és specifikusság mellett Blaker és Clopper-Pearson módszerével szerkesztett konfidenciaintervallumoknál. (Ezek az eljárások ismert szenzitivitást és specifikusságot feltételeznek.) Azt tapasztaltuk, hogy a pontosság romlása csak akkor válik elhanyagolhatóvá, ha a szenzitivitás és specifikusság becslésére kiválasztott minták nagysága legalább ötször nagyobb, mint a prevalencia becslésére szolgáló minta. A pontosság romlásának mértéke ezen túlmenően függ a prevalencia mintájának nagyságától és a szenzitivitás és specifikusság valódi értékétől is.
6. Számítógépes szimulációs kísérletekkel igazoltuk, hogy a Wald-Rogan-Gladen konfidenciaintervallum javítását célzó, széles körben alkalmazott és publikált  $\hat{P} \pm 3 \cdot \text{var}(\hat{P})^{\frac{1}{2}}$  próbaintervallum alkalmazása az eljárás pontosságát nem javítja, emellett az esetek jelentős részében nem eredményez intervallumbecslést.
7. A magyarországi szarvasmarha-állományok BHV-1 szeropozitivitásából megbecsültük a BHV-1 fertőzöttség valódi prevalenciáját. Egzakt konfidenciaintervallumokkal kimutattuk, hogy a valódi prevalencia minden korcsoportban és összességében is kisebb a mentesítés utáni ellenőrzéskor a mentesítést megelőző felméréshez képest. Megállapítottuk, hogy a valódi

prevalenciák felméréskor nagyobbak, ellenőrzéskor pedig kisebbek, mint a szeropozitivitások, tehát a vizsgált periódusban (2002 és 2006 között) a mentesítés valójában eredményesebb volt, mint azt a szeropozitivitási arányok mutatják.

8. A prevalenciák megyék közötti és megyéken belüli ingadozását vizsgálva megállapítottuk, hogy a telepek között, megyétől függetlenül, nagy a heterogenitás mind a felmérő, mind az ellenőrző vizsgálatok során.
9. Az országot három földrajzilag összefüggő homogén megyecsoportra osztva megvizsgáltuk a prevalencia átlagát és csoporton belüli szórását. Kimutattuk, hogy felméréskor nincs lényeges különbség a csoportok átlaga és szórása között. Mentés után a Dunántúlon és az északi megyékben valamivel jobban sikerült csökkenteni a fertőzés prevalenciáját, mint az Alföldön. A heterogenitás is csökkent, leginkább az északi megyékben.
10. Az állománylétszám és a prevalencia kapcsolatát szimultán konfidenciaintervallumokkal vizsgálva, Bonferroni módszerét követve kimutattuk, hogy felméréskor a nagyobb létszámú telepek csoportjához növekvő tendenciájú BHV-1 prevalencia tartozik, 5%-os szimultán tévedési valószínűség mellett. Megállapítottuk, hogy a mentés utáni ellenőrzésnél fordított a trend, a nagy létszámú telepeken 5%-os szinten szignifikánsan kisebb a prevalencia.
11. Korcsoportonként vizsgálva, felméréskor a borjak és tehének esetében állapítottunk meg a telep állatszámra és a prevalencia között 5%-os szinten szignifikáns pozitív kapcsolatot. A növendék és vemhes üszőknél nem találtunk szignifikáns összefüggést. A mentést követő ellenőrzéskor nem mutattunk ki korcsoportonként szignifikáns kapcsolatot a telepek állatszámra és a prevalencia között.
12. Kimutattuk, hogy a zsúfoltsági index felméréskor 867, ellenőrzéskor 817. Az átlagosan lehetséges kontaktusszám 50–80%-kal nagyobb, mint a telepek átlagos létszáma.
13. Fogalmi analógiát tártam fel a zsúfoltság és a fajok biológiai diverzitása között. A biológiai diverzitást Patil és Taille (1982) mint a fajok skálázott ritkaságának egyedekre vett átlagát értelmezték. Kimutattam, hogy a skálázó függvény megadásával a zsúfoltság és a diverzitás egymással kifejezhetők. Az egymásnak ilyen módon megfeleltetett legfontosabb zsúfoltsági és diverzitási indexek lineáris skálafüggvényre a Lloyd- és Reiczigel-féle lineáris zsúfoltság és a Simpson-féle diverzitás, logaritmus függvényre a logaritmusos zsúfoltság és a Shannon-féle

diverzitási index, hiperbolikus függvényre a hiperbolikus zsúfoltság és a fajgazdagság (a fajok száma mínusz 1).

14. A diverzitások elméletében értelmezett effektív fajszám megfelelőjeként a zsúfoltság fogalomkörében tetszőleges monoton növény skálafüggvény mellett bevezettem az effektív csoportszám fogalmát. Ez olyan egyenlő egyedszámú csoportok számát jelenti, ahol a csoportok egyedeinek zsúfoltsága a vizsgált közösség zsúfoltságával egyezik meg. Erre a fogalomra építve, a tényleges csoportszám és az effektív csoportszám hányadosaként tetszőleges monoton növény skálafüggvényre általánosítottam a Bez-féle aggregáltsági indexet.
15. Általános feltételek mellett bebizonyítottam, hogy az aggregáltsági index  $\geq 1$ , és értéke pontosan akkor 1, ha az egyedek számának csoportok közötti megoszlása teljesen egyenletes.
16. Kimutattam, hogy a zsúfoltság és a diverzitás között közvetlen, nem csak a fogalmi analógia szintjén megnyilvánuló kapcsolat is van olyankor, ha a vizsgált közösség csoportjaiban az egyedek több fajhoz tartoznak. A fajok együttesére vett zsúfoltság felbontható a fajenkénti átlagos zsúfoltság és az átlagos csoporton belüli, szubjektív diverzitás összegére.
17. Ha a vizsgált közösség csoportjaiban az egyedek több fajhoz tartoznak, akkor ismeretes, hogy a közösség együttes, ún. gamma valódi diverzitása (azaz effektív fajszáma) a csoportokon belüli átlagos valódi diverzitás (ún. alfa diverzitás) és a csoportok közötti változatosságot kifejező béta diverzitás szorzata. A zsúfoltság–diverzitás fogalmi analógia mentén a valódi diverzitásnak az aggregáltsági indexet feleltettem meg. A teljes közösség aggregáltságát gamma aggregáltságnak, a fajenkénti átlagos aggregáltságot alfa aggregáltságnak neveztem el. A béta aggregáltságot az alfa és gamma aggregáltságok hányadosaként értelmeztem, ez az aggregáltságnak a fajok közötti változatosságot kifejező tényezője.
18. A zsúfoltságra és a diverzitásra kapott elméleti eredményeket felhasználtam vetési varjak (*Corvus frugilegus*) és dolmányos varjak (*Corvus cornix*) tetűparazitáltságának összehasonlítására. A paraziták zsúfoltsága vetési varjakon nagyobb, emellett a zsúfoltság szubjektív diverzitásos összetevőjének is nagyobb a százalékaránya. A varjófajok közötti eltérésre többek között azt a magyarázatot adtam, hogy a fertőzött dolmányos varjak körében nagyobb arányú az egyetlen tetűfajjal fertőzöttek, így 0 diverzitásúak aránya (50%), mint a fertőzött vetési varjakon (21%).



19. A valódi diverzitások és aggregáltságok multiplikatív felbontásával kimutattam, hogy a tetvek valódi alfa és gamma diverzitásai a vetési varjak esetében valamivel nagyobbak, mint a dolmányos varjaknál. A béta diverzitások a két varjófajnál nagyjából megegyeznek, értékük 1.5-2.5 között van. A tetűfajok átlagos, azaz alfa aggregáltsága a dolmányos varjak körében nagyobb, mint a vetési varjaknál. Ez azt jelenti, hogy a dolmányos varjak tetűfajai átlagosan aggregáltabban oszlanak meg a gazdaállatok között, mint a vetési varjak tetűfajai.
20. A béta aggregáltságra mindkét varjófajnál 1 és 2 közötti értékeket kaptam. Ebből kiindulva blokk klaszterezéssel és korrespondanciaelemzéssel kimutattam, hogy a vetési varjak génuszai két, gazdákon kevéssé átfedő csoportra, a *Brueelia*, illetve a *Myrsidea* és *Menacanthus* együttesére bonthatók. A dolmányos varjakon ugyanezekkel a statisztikai módszerekkel azt találtam, hogy a két domináns génusz, a *Myrsidea* és a *Menacanthus* alkot a gazdákon csak kismértékben átfedő csoportot. Érdekeség, hogy a vetési varjakon – a dolmányos varjakkal ellentétben – a *Myrsidea* és *Menacanthus* génuszok nem különülnek el, gyakran előfordulnak közös gazdán.
21. Általános valószínűségelméleti modellt vezettem be csoportokba szerveződő egyedek jellemzőinek egyedszemponitú és csoportszintű eloszlásaira. Formulákat adtam a két szemléletmód szerinti eloszlások közötti alapvető transzformációkra.
22. A bevezetett modellen belül valószínűségelméleti jelentést rendeltem az általános skálafüggvényű zsúfoltságához.
23. Elégséges és szükséges és elégséges feltételeket adtam arra, hogy egy általános regressziós modell milyen feltételekkel egyezik meg az egyedek és a csoportok szempontjából. Ezek egyike az, hogy a csoportok egyedszáma szerepeljen a regressziós modell magyarázó változói között. A többi ennek az alkalmazási feltételek és körülmények szerint kissé módosított változata.
24. Az említett valószínűségelméleti modellt és a hozzá kapcsolódó elméleti eredményeket alkalmaztam gímszarvasokat (*Cervus elaphus*) parazitáló szarvaskullancslegyek (*Lipoptena cervi*) ivararánya és a parazitáltság mértéke közötti összefüggés vizsgálatára. Az adatokra kevert hatásos logisztikus regressziót illesztettem úgy, hogy a gímszarvasokon talált paraziták száma a modell egyik magyarázó változója. Kimutattam, hogy a hím szarvaskullancslegyek részaránya szignifikánsan kisebb az erősebben parazitált gímszarvasokon ( $p = 0.0001$ ). Az eredmény érvényes mind a gímszarvasokra, mint egymástól független entitásokra, megfigyelési egységekre, mind pedig az egyes kullancslegyek saját gazdáján élő

mikropopulációkra. A regressziós modell eredményéből levezettem, hogy míg egy tipikus gímszarvason az ivararány nem különbözik szignifikánsan az 1:1 aránytól, egy tipikus kullancslégy gazdakörnyezetén az ivararány a nőstények irányába tolódik el.

## 9 Irodalomjegyzék

Agresti A.: *Categorical Data Analysis*, 2nd Edition. Wiley, New York. 2002.

Agresti A., Caffo B.: Simple and Effective Confidence Intervals for Proportions and Differences of Proportions Result from Adding Two Successes and Two Failures. *Am. Stat.* 54, 280-288. 2000.

Agresti A., Coull B.A.: Approximate is better than 'exact' for interval estimation of binomial proportions. *Am. Stat.* 52, 119–126. 1998.

Akaike H.: A new look at the statistical model identification. *Automatic Control, IEEE Transactions on* 19, 6, 716–723. 1974.

Anderson N.L., Grahn R.A. Van Hoosear K., Bondurant R.H.: Studies of trichomonad protozoa in free ranging songbirds: Prevalence of *Trichomonas gallinae* in house finches (*Carpodacus mexicanus*) and corvids and a novel trichomonad in mockingbirds (*Mimus polyglottos*). *Vet. Parasitol.* 161, 178–186. 2009.

Bates D., Maechler M., Bolker B., Walker S.: lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-5. <http://CRAN.R-project.org/package=lme4>. 2013.

Belgorodski N., Greiner M., Engelhardt A.: Package "rriskBayes". <http://cran.r-project.org/web/packages/rriskBayes/rriskBayes.pdf>, (accessed 5 July 2013). 2012.

Berger W.H., Parker F.L.: Diversity of Planktonic Foraminifera in Deep-Sea Sediments. *Science* 168, 1345–1347. 1970.

Bez N.: On the use of Lloyd's index of patchiness. *Fisheries Oceanography* 9, 372–376. 2000.

Bhatia P., Iovleff S., Goavert G., Brault V., Biernacki C., Celeux G.: blockcluster: Co-Clustering package for binary, contingency, continuous and categorical data-sets. R package version 3.0.1. <http://CRAN.R-project.org/package=blockcluster>. 2014.

Blaker H.: Confidence curves and improved exact confidence intervals for discrete distributions. *Can. J. Stat.* 28, 783–798. 2000.

Boadella M., Vicente J., Ruiz-Fons F., de la Fuente J., Gortazar C.: Effects of culling Eurasian wild boar on the prevalence of *Mycobacterium bovis* and Aujeszky's disease virus. *Prev. Vet. Med.* 107, 214–221. 2012.

Boelaert F., Biront P., Soumare B., Dispas M., Vanopdenbosch E., Vermeersch J.P., Raskin A., Dufey J., Berkvens D., Kerkhofs P.: Prevalence of bovine herpesvirus-1 in the Belgian cattle population. *Prev. Vet. Med.* 45, 285–295. 2000.

Bolla M.: Korrespondenciaanalízis. *Alkalmazott Matematikai Lapok* 13, 207–230. 1987.

Canty A., Ripley B.: *boot: Bootstrap R (S-Plus) Functions*. R package version 1.3-9. 2013.

Clopper C.J., Pearson E.S.: The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* 26, 404–413. 1934.

Davison A.C., Hinkley D.V.: *Bootstrap Methods and Their Application*. Cambridge University Press. 1997.

de Wit J.J., Hage J.J., Brinkhof J., Westenbrink F.: A comparative study of serological tests for use in the bovine herpesvirus-1 eradication programme in The Netherlands. *Vet. Microbiol.* 61, 153–163. 1998.

Enøe C., Georgiadis M., Johnson W.: Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease state is unknown. *Prev. Vet. Med.* 45, 61–81. 2000.

Farnham M.W., Norby B., Goldsmith T.J., Wells S.J.: Meta-analysis of field studies on bovine tuberculosis skin tests in United States cattle herds. *Prev. Vet. Med.* 103, 234–242. 2012.

Faye B., Castel V., Lesnoff M., Rutabinda D., Dhalwa J.: Tuberculosis and brucellosis prevalence survey on dairy cattle in Mbarara milk basin (Uganda). *Prev. Vet. Med.* 67, 267–281. 2005.

Greiner M., Gardner I.A.: Application of diagnostic tests in veterinary epidemiologic studies. *Prev. Vet. Med.* 45, 43–59. 2000a.

Greiner M., Smid J., Havelaar A.H., Müller-Graf C.: Graphical models and Bayesian domains in risk modelling: Application in microbiological risk assessment. *Prev. Vet. Med.* 110, 4–11. 2013.

Greiner M., Gardner I.A.: Epidemiologic issues in the validation of veterinary diagnostic tests. *Prev. Vet. Med.* 45, 3-22. 2000b.

Govaert G., Nadif M.: Latent block model for contingency table. *Communications in Statistics - Theory and Methods* 39, 416–425. 2010.

Greenacre M.: *Correspondence Analysis in Practice*. 2nd edition. Chapman & Hall/CRC, Boca Raton. 2007.

Harbord R.M., Whiting P., Sterne J., Egger M., Deeks J., Shang A., Bachmann L.: An empirical comparison of methods for meta-analysis of diagnostic accuracy showed hierarchical models are necessary. *J. Clin. Epidemiol.* 61, 1095–1103. 2008.

Heesterbeek J.A.P.: A brief history of  $R_0$  and a recipe for its calculation. *Acta Biotheor.* 50, 189–204. 2002.

Hill M.O.: Diversity and Evenness: A Unifying Notation and Its Consequences. *Ecology* 54, 427–432. 1973.

Hui S.L., Walter S.D.: Estimating the error rates of diagnostic tests. *Biometrics* 36, 167–71. 1980.

Jarman P.J.: The social organisation of antelope in relation to their ecology. *Behaviour* 48, 215–268. 1974.

Jost L.: Entropy and diversity. *Oikos* 113, 363–375. 2006.

Jost L.: Partitioning diversity into independent alpha and beta components. *Ecology* 88, 2427–2439. 2007.

Kramps J.A., Banks M., Beer M., Kerkhofs P., Perrine M., Wellenberg G.J., Van Oirschot J.T.: Evaluation of tests for antibodies against bovine herpesvirus 1 performed in national reference laboratories in Europe. *Vet. Microbiol.* 102, 169–181. 2004.

Kramps J.A., Magdalena J., Quak J., Weerdmeester K., Kaashoek M.J., Maris-Veldhuis M.A., Rijsewijk F.A.M., Keil G., Van Oirschot J.T.: A simple, specific, and highly sensitive blocking enzyme-linked immunosorbent assay for detection of antibodies to bovine herpesvirus-1. *J. Clin. Microbiol.* 32, 2175–2181. 1994.

Lande R.: Statistics and partitioning of species diversity, and similarity among multiple communities. *Oikos* 76, 5–13. 1996.

- Lang Zs., Földi J., Ózsvári L., Reiczigel J.: Szeropozitivitás és prevalencia összefüggései hazai BHV-1-mentesítés adatainak járványtani elemzése alapján. *Magy. Állatorv. Lapja* 135, 525–534. 2013.
- Lang Z., Reiczigel J.: Confidence limits for prevalence of disease adjusted for estimated sensitivity and specificity. *Preventive Veterinary Medicine* 113, 13–22. 2014.
- Leeuw J., Mair P.: Simple and Canonical Correspondence Analysis Using the R Package anacor. *Journal of Statistical Software* 31(5), 1–18. URL <http://www.jstatsoft.org/v31/i05/>. 2009.
- Lesaffre E., Speybroeck N., Berkvens D.: Bayes and diagnostic testing. *Vet. Parasitol.* 148, 58–61. 2007.
- Lloyd M.: Mean crowding. *Journal of Animal Ecology* 36, 1–30. 1967.
- MacArthur R.H.: Patterns of species diversity. *Biol. Rev.* 40, 510–533. 1965.
- Magurran A.E.: *Measuring Biological Diversity*. Blackwell Science Ltd. 108–109. 2004.
- Marshall A.G.: The sex ratio in ectoparasitic insects. *Ecological Entomology* 6, 155–174. 1981.
- Messam L.L., Branscum A.J., Collins M.T., Gardner I.A.: Frequentist and Bayesian approaches to prevalence estimation using examples from Johne's disease. *Anim. Health Res. Rev.*, 9, 1–23. 2008.
- Nenadic O., Greenacre M.: Correspondence Analysis in R, with Two- and Three-dimensional Graphics: The ca Package. *Journal of Statistical Software* 20(3), 1–13. 2007.
- Neuhäuser M., Kotzmann J., Walier M., Poulin R.: The comparison of mean crowding between two groups. *Journal of Parasitology* 96, 477–481. 2010.
- Olszewski T.D.: A unified mathematical framework for the measurement of richness and evenness within and among multiple communities. *Oikos* 104, 377–387. 2004.
- Patil G. P., Taillie C.: Diversity as a Concept and its Measurement. *J Am Stat Assoc.* 77, 548–561. 1982.
- Pálfi V., Ózsvári L., Földi J., Hankó-Faragó E.: Élő IBR-markervakcinával folytatott mentesítés nagyüzemi tapasztalatai Magyarországon. *Magy. Állatorv. Lapja* 129, 323–331. 2007.

Pielou E.C.: The Use of Information Theory in the Study of the Diversity of Biological Populations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability (Vol. IV), 163–177. 1967.

Pinho L., Thompson G., Machado M., Carvalheira J.: Management practices associated with the bulk tank milk prevalence of *Mycoplasma* spp. in dairy herds in Northwestern Portugal. *Prev. Vet. Med.* 108, 21–27. 2013.

Price R.D., Hellenthal R.A., Palma R.L., Johnson K.P., Clayton D.H. (eds): The chewing lice: World checklist and biological overview. Illinois Natural History Survey, Champaign, Illinois, USA. 2003.

Quirin R., Rasolofo V., Andriambololona R., Ramboasolo A., Rasolonavalona T., Raharisolo T., Rakotoaritnahina H., Chanteau S., Boisier P.: Validity of intradermal tuberculin testing for the screening of bovine tuberculosis in Madagascar. *Onderstepoort Journal of Veterinary Research*, 68, 231–238. 2001.

R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>. 2014.

Reiczigel J.: Válogatott fejezetek a biostatistikából. SZIE ÁOTK, 2005. <http://www.univet.hu/users/jreiczig/valfej/val-fej-jegyzet-2005-02-05.pdf>

Reiczigel J.: Optimális skálázás: módszer és program. Magyar Tudományos Akadémia Szociológiai Kutatóintézet, Budapest, 1988.

Reiczigel J., Földi J., Ózsvári L.: Exact confidence limits for prevalence of a disease with an imperfect diagnostic test. *Epidemiol. Infect.* 138, 1674–1678. 2010.

Reiczigel J., Harnos A., Solymosi N.: Biostatistika nem statisztikusoknak. Pars Kft., Nagykovácsi, 182–183. 2007.

Reiczigel J., Lang Z., Rózsa L., Tóthmérész B.: Properties of crowding indices and statistical tools to analyse crowding data. *Journal of Parasitology* 91, 245–252. 2005.

Reiczigel J., Lang Z., Rózsa L., Tóthmérész B.: Measures of sociality: Two different views of group size. *Animal Behaviour* 75: 715–721. 2008.

Rékási J.: Mallophaga of birds in Hungary. I. (in Hungarian). *Parasitologia Hungarica* 6, 215–238. 1973.

Rékási J.: Die Federling-Sammlung des Ungarischen Naturwissenschaftlichen Museums. I. *Parasitologia Hungarica* 11, 107–112. 1978.

Rékási J.: The Mallophaga collection of the Matra Museum I. (in Hungarian). *Folia Historio-Naturalia Musei Matraensis*, Gyöngyös, 9, 23–47. 1984.

Rékási J.: Feather lice of Hungarian birds (Mallophaga). II. (in Hungarian). *Parasitologia Hungarica* 19, 119–126. 1986.

Rékási J.: The lice fauna of the Bátorliget Nature Reserves (Mallophaga). The Bátorliget Nature Reserves After Forty Years (1990) (ed. S. Mahunka). Hungarian Natural History Museum, Budapest. pp. 323–331. 1991.

Rékási J.: (1993) Data on the Mallophaga fauna of the Bükk National Park. The Fauna of the Bükk National Park (ed. S. Mahunka). Hungarian Natural History Museum, Budapest. pp. 61–63. 1993.

Rékási J., Kiss J.B.: Beiträge zur Kenntnis der Federlinge (Mallophaga) der Vögel Nord-Dobrudschas (Rumänien). *Parasitologia Hungarica* 10, 97–116. 1977.

Rékási J., Kiss J.B.: Weitere Beiträge zur Kenntnis der Federlinge (Mallophaga) von Vögel der Nord-Dobrudscha. *Parasitologia Hungarica* 13, 67–93. 1980.

Rékási J., Kiss J.B.: (1984) Weitere Angaben zur Kenntnis der Federlinge (Mallophaga) der Vögel Nord-Dobrudschas (Rumänien). II. *Parasitologia Hungarica* 17, 97–117. 1984.

Rékási J., Rózsa L., Kiss J.B.: Patterns in the distribution of avian lice (Phthiraptera: Amblycera, Ischnocera). *J. Avian. Biol.* 28: 150–156. 1997.

Rényi A.: On Measures of Entropy and Information. *Proc. Fourth Berkeley Symp. Math. Stat. and Probability*, Vol. 1. Berkeley, CA. University of California Press. 547–561. 1961.

Rigó E.: *Lipoptena cervi* - fertőzés előfordulása a Hanság szarvasállományában. Egyetemi szakdolgozat. Nyugat-magyarországi Egyetem. Mosonmagyaróvár. 2011.

Rogan W.J., Gladen B.: Estimating prevalence from the results of a screening test, *Am. J. Epidemiol.* 107, 71–76. 1978.

Rózsa L.: Patterns in the abundance of avian lice (Phthiraptera: Amblycera, Ischnocera). *J. Avian Biol.* 28: 249–254. 1997.



- Rózsa L., Rékási J., Reiczigel J.: Relationship of host coloniality to the population ecology of avian lice (Insecta: Phthiraptera). *J. Anim. Ecol.* 65, 242–248. 1996.
- Sandvik H., Hunskaar S., Vanvik A., Bratt H., Seim A., Hermstad R.: Diagnostic classification of female urinary incontinence: an epidemiological survey corrected for validity. *J. Clin. Epidemiol.* 48, 339–343. 1995.
- Sarrazin S., Veldhuis A., Méroc E., Vangeel I., Laureyns J., Dewulf J., Caij A.B., Piepers S., Hooyberghs J., Ribbens S., Van Der Stede I.: Serological and virological BVDV prevalence and risk factor analysis for herds to be BVDV seropositive in Belgian cattle herds. *Prev. Vet. Med.* 108, 28–37. 2013.
- Shannon C.E., Weaver W.: *The mathematical theory of communication*. University of Illinois Press, Urbana, IL. 1963.
- Simpson E.H.: Measurement of diversity. *Nature* 163, 688. 1949.
- Soós T., Tuboly S.: *Vakcinológia: a fertőző állatbetegségek immunprophylaxisa. A/3 Nyip. Kft. Budapest*, 59–60, 141–146. 2009.
- Tang M., Qiu S., Poon W.: Confidence interval construction for disease prevalence based on partial validation series. *Comput. Stat. Data An.* 56, 1200–1220. 2012.
- Thrusfield M.: *Veterinary Epidemiology*, 3rd ed. Blackwell Science Ltd. UK, 53–57. 2005.
- Toft N., Jørgensen E., Højsgaard S.: Diagnosing diagnostic tests: evaluating the assumptions underlying the estimation of sensitivity and specificity in the absence of a gold standard. *Preventive Veterinary Medicine*, 79, 244–256. 2005.
- Tóthmérész B.: Comparison of different methods for diversity ordering. *Journal of Vegetation Science*, 6: 283–290. 1995.
- Tóthmérész B.: *Diverzitási rendezések*. Scientia Kiadó, Budapest. 1997.
- Tuomisto H.: A diversity of beta diversities: straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity. *Ecography* 33, 2–22. 2010a.
- Tuomisto H.: A diversity of beta diversities: straightening up a concept gone awry. Part 2. Quantifying beta diversity and related phenomena. *Ecography* 33, 23–45. 2010b.

Vágó E., Lang Zs., Kemény S.: Overdispersion at the Binomial and Multinomial Distribution. *Periodica Polytechnica* 55, 17–20. 2011.

Veech J.A., Summerville K.S., Crist, T.O., Gering J.C.: The additive partitioning of species diversity: recent revival of an old idea. *Oikos* 99, 3–9. 2002.

Whittaker R.H.: Vegetation of the Siskiyou Mountains, Oregon and California. *Ecol. Monogr.* 30, 279–338. 1960.

Whittaker R.H.: Evolution and measurement of species diversity. *Taxon* 21, 213–251. 1972.

Youden D.: Index for rating diagnostic tests. *Cancer* 3, 32–35. 1950.

## 10 A doktori kutatás eredményeinek közlései

Lang Zs., Földi J., Ózsvári L., Reiczigel J.: Szeropozitivitás és prevalencia összefüggései hazai BHV-1-mentesítés adatainak járványtani elemzése alapján. *Magy. Állatorv. Lapja* 135, 525–534. 2013. IF<sub>2012</sub>: 0.146

Lang Z., Reiczigel J.: Confidence limits for prevalence of disease adjusted for estimated sensitivity and specificity. *Preventive Veterinary Medicine* 113, 13–22. 2014. IF<sub>2012</sub>: 2.389

Reiczigel J., Lang Z., Rózsa L., Tóthmérész B.: Properties of crowding indices and statistical tools to analyse crowding data. *Journal of Parasitology* 91, 245–252. 2005. IF<sub>2011</sub>: 1.320

Reiczigel J., Lang Z., Rózsa L., Tóthmérész B.: Measures of sociality: Two different views of group size. *Animal Behaviour* 75: 715–721. 2008. IF<sub>2012</sub>: 3.068

Vágó E., Lang Zs., Kemény S.: Overdispersion at the Binomial and Multinomial Distribution. *Periodica Polytechnica* 55, 17–20. 2011. IF<sub>2012</sub>: 0.269

## 11 A doktori kutatáshoz nem kapcsolódó közlemények

Antal J., Boda K., Gachályi B., Hajtman B., Lakner G., Lang Zs., Reiczigel J., Singer J., Vargha P., Vereczkey L., Lakner G. Gachályi B., Singer J. (szerk.): Klinikai farmakológiai kislexikon biostatisztikai fogalomtárral. SpringMed Publishing Ltd., Budapest, 424 p. 2005.

Bacskai M., Lang Zs., Lazányi I.: Betegút alapú modellek gyakorlati alkalmazásainak lehetőségei. Informatika és menedzsment az egészségügyben: az egészségügyi vezetők szaklapja VIII:(6), 37–41. 2009.

Bányai J., Fischer J., Lang Z.: Tolerances Based on Poisson Distributions. Seed Science And Technology 16:(1) 321–329. 1988.

Bányai J., Fischer J., Lang Z.: Improved Range Homogeneity Test For Checking Seed Lots. Seed Science And Technology 18:(2) 239–253. 1990.

Bitter I., Katona L., Zámboi J., Takács P., Fehér L., Diels J., Bacskai M., Lang Z., Gyáni G., Czobor P.: Comparative effectiveness of depot and oral second generation antipsychotic drugs in schizophrenia: A nationwide study in Hungary. European Neuropsychopharmacology 23,1383–1390. 2013.

Eory A., Gonda X., Lang Zs., Torzsa P., Kalman J., Kalabay L., Rihmer Z.: Personality and cardiovascular risk: Association between hypertension and affective temperaments—a cross-sectional observational study in primary care settings. European Journal Of General Practice Early online:(1) 1–6. 2014.

Erdelyi B., Csakany B., Rödönyi G., Soumelidis A., Lang Z., Nemeth J.: Dynamics of ocular surface topography in healthy subjects. Ophthalmic And Physiological Optics 26:(4) 419–425. 2006.

Hagyo K., Csakany B., Lang Z., Nemeth J.: Variability of Higher Order Wavefront Aberrations After Blinks. Journal Of Refractive Surgery 25:(1) 59–68. 2009.

Jacso O., Fok E., Kiss G., Kokeny G., Lang Z.: Preliminary Findings on the Efficacy of Selamectin in the Treatment of Dogs Naturally Infected with *Dirofilaria Repens*.

Acta Veterinaria Hungarica 58:(4) 405–412. 2010.

Németh J., Fodor E., Lang Zs., Kosina-Hagyó K., Berta A., Komár T., Petricek I., Higazy M., Prost M., Grupcheva C., Evren Kemer O., Schollmayer P., Samaha A., Hlavackova K.: Lid-parallel conjunctival folds (LIPCOF) and dry eye: a multicentre study. *British Journal of Ophthalmology* 96, 1380–1385. 2012.

Kosina-Hagyó K., Veres A., Fodor E., Lang Z., Csakany B., Nemeth J.: Kinetic analysis of topographical parameters and interference pattern of tear lipid layer in normal subjects. *Ophthalmic and Physiological Optics* 30:(6) 790–799. 2010.

Lakatos P., Tóth E., Lang Zs., Nagy B., Szekeres L., Takács I.: Adherencia a magyar posztmenopauzális osteoporotikus női betegek körében. *Lege Artis Medicinae Kid* 2:(4) 5–15. 2012.

Lang Zs.: Horocycles of a Dynamical System on the Plane. *Studia Scientiarum Mathematicarum Hungarica* 19, 131–139. 1984.

Lang Zs., Bacskai M., Kasza G.: Biztosítási termékek díjkalkulációjának lehetőségei a magánfinanszírozásban. *Informatika és menedzsment az egészségügyben: az egészségügyi vezetők szaklapja VII*: p. 14. 2008.

Lang Zs., Rakonczai P., Lazányi I., Lohn E., Bacskai M.: Eseményt követő kiadások eloszlásának és várható értékének modellezése Cox-féle arányos hazard modellel. *Informatika és menedzsment az egészségügyben: az egészségügyi vezetők szaklapja XI*:(1) 34–39. 2012.

Lang Zs., Bacskai M., Rakonczai P., Csiha J.: Terápiás hatások mérése betegforgalmi adatokon. *Informatika és menedzsment az egészségügyben: az egészségügyi vezetők szaklapja XII*:(különszám) 5–9. 2013.

Nagy B., Lang Zs., Hegyi R.: Az analóg bázisinzulinok terápiás hatásának összevetése 2-es típusú cukorbeteg napi egyszeri kezelésében. *Orvostovábbképző Szemle XVI*:(1) 86–94. 2009.

Németh J., Erdélyi B., Csákány B., Gáspár P., Soumelidis A., Kahlesz F., Lang Zs.: High-speed videotopographic measurement of tear-film build-up time. *Investigative Ophthalmology & Visual Science* 43:(6) 1783–1790. 2002.

Posgay M., Varró M., Ungváry Gy., Lang Zs.: Eternityvár 500 m-es körzetén belül, valamint azon kívül élő lakosság egészségi állapotának felmérése. *Egészségtudomány* 48, 27–36. 2004.

Posgay M., Varró M., Lang Zs.: Felnőttek egészségi állapotának felmérése és annak környezeti tényezőkkel való összefüggésének vizsgálata Komáromban kérdőíves felmérés alapján. Magyar epidemiológia: a Magyar Molekuláris és Prediktív Epidemiológiai Társaság hivatalos, negyedévenként megjelenő folyóirata 3:(3) 179–191. 2006.

Posgay M., Varró M., Szentmihályi R., Lang Zs.: Légzőszervi környezet-epidemiológiai vizsgálat Nyergesújfalun és Komáromban: Összehasonlító tanulmány. Orvosi Hetilap 151:(10) 378–384. 2010.

Pribenszky C., Losonczi E., Molnar M., Lang Z., Matyas S., Rajczy K., Molnar K., Kovacs P., Nagy P., Conceicao J., Vajta G.: Prediction of in-vitro developmental competence of early cleavage-stage mouse embryos with compact time-lapse equipment. Reproductive Biomedicine Online 20:(3) 371–379. 2010.

Rakonczai P., Lang Zs., Csiha J., Balázs T., Bacskai M.: Költség eloszlások statisztikai modellezése cenzorált adatokon. Informatika és menedzsment az egészségügyben: az egészségügyi vezetők szaklapja XII:(különszám) 16–20. 2013.

Reiczigel J., Brugger K., Rubel F., Solymosi N., Lang Z.: Bayesian analysis of a dynamical model for the spread of the Usutu virus. Stochastic Environmental Research And Risk Assessment 24:(3) 455–462. 2010.

Tóth E., Hegyi R., Csiha J., Lang Zs.: Az onkológiai terápiás betegektől eredményességének értékelése OEP ellátási adatokon. Informatika és menedzsment az egészségügyben: az egészségügyi vezetők szaklapja XII:(5) 42–46. 2013.

Varró M., Virágh Z., Rudnai P., Lang Zs., Gombkötő Gy., Szeremi M., Mácsik A., Szabó E.: Associations of agricultural airborne pesticide sprays reaching inhabited areas with respiratory and other symptoms among 7-11 year old children in Heves county, Hungary. European Epimarker 8:(2) 1–5. 2004.

## 12 Köszönetnyilvánítás

Szeretnék köszönetet mondani témavezetőmnek, Dr. Reiczigel Jenőnek és Dr. Rózsa Lajosnak, hogy kutatási témáikba bevontak és munkámat nagy hozzáértéssel segítették. Köszönetet mondok Dr. Földi Józsefnek és Dr. Ózsvári Lászlónak a közös kutató munkáért, lelkes támogatásukért. Köszönet illeti Dr. Kemény Sándort a kitartó buzdításért és Dr. Vágó Emesét hármunk közös kutató munkájáért. Köszönöm opponenseimnek, Dr. Harnos Andreának és Dr. Singer Júliának, hogy a dolgozatot elolvasták és észrevételeikkel annak színvonalát javították. Köszönöm Dr. Pásztory-Kovács Szilviának, hogy a fokozatszerzéssel összefüggő tapasztalatait megosztotta velem. Köszönöm Fehérvári Péternek és Dr. Vas Zoltánnak segítő megjegyzéseiket, javaslataikat.

Köszönöm feleségemnek, Annának és fiamnak, Péternek a szeretetet és türelmet, és szüleimnek, hogy mindig mellettem álltak.