Hardy-Weinbergov model ravnoteže

NENAD ŠUVAK*

Sažetak. *Ovim člankom, koji je gradivom i pristupom prilagođen četvrtim razredima srednjih škola koje u nastavi matematike imaju teoriju vjerojatnosti, pokušat ćemo kroz zanimljive genetičke činjenice predstaviti upotrebu osnovnog matematičkog aparata teorije vjerojatnosti. S matematičke točke gledišta obradit ćemo model ravnoteže u populaciji do kojeg su neovisno došli britanski matematičar G. H. Hardy (1877. – 1947.) i njemački fizičar W. Weinberg (1862. – 1937.), te na nekoliko primjera ilustrirati upotrebu ovog modela.*

Ključne riječi: *alel, genom, genotip, matematički model, populacija, ravnoteža, udio, vjerojatnost.*

Hardy - Weinberg equilibrium

Abstract. *This article, which is by its content and approach adapted to students of senior classes of high school, represents usage of some important terms of probability theory through interesting facts in genetics. A population equilibrium model, which is the result of independent research of British mathematician G. H. Hardy (1877 – 1947) and German physician W. Weinberg (1862 – 1937), is going to be presented from mathematical point of view . The usage of this model is going to be illustrated by a few examples.*

Key words: *allele, genome, genotype, mathematical model, population, equilibrium, share, probability.*

Uvod

Evolucija[1] je vrlo dugotrajan proces koji se očituje u promjeni fizičkih svojstava, te načina ponašanja i reagiranja jedinki. Iako se evolucijske promjene očituju tek na razini populacije, njihov uzrok treba tražiti na mikrorazini (tj. na razini gena). Genetička struktura populacije određena je brojem, odnosno udjelom, alela[2] u ukupnom genetičkom materijalu promatrane populacije. Promjene broja alela (odnosno promjene udjela alela u ukupnom genetičkom materijalu populacije) uzrokuju promjenu genetičke strukture i time postaju glavni pokretač evolucije. Taj vid evolucije

*Odjel za matematiku, Gajev trg 6, HR-31 000 Osijek, e-mail: nsuvak@mathos.hr

[1] Proces razvoja i prilagodbe živih bića uvjetima okoliša u kojem obitavaju.

[2] Aleli su različite varijante jednog gena.

nazivamo mikroevolucijom i upravo ćemo njezine matematičke aspekte obraditi u ovom članku.

Što bi se dogodilo s populacijom kada bi bio zadovoljen određeni skup uvjeta koji osiguravaju konstantnost udjela određenih tipova alela u njoj? Naravno, pod tim bi uvjetima evolucija izostala i rekli bismo da se populacija nalazi u stanju Hardy - Weinbergove ravnoteže. Uvjeti koji populaciju dovode u takvu ravnotežu su sljedeći:

- Populacija je vrlo velika, te nema fenotipskog[3] preferiranja pri izboru partnera za stvaranje potomstva (engl. *random mating*).
- Nema mutacija, tj. promjena u strukturi DNA.
- Nema migracija, tj. nema izlazaka pripadnih ili ulazaka novih jedinki u populaciju.
- Nema genetičkog drifta, odnosno slučajnih promjena u genetičkoj strukturi populacije izazvanih vanjskim utjecajima.
- Nema prirodne selekcije, tj. sve jedinke u potpuno jednakoj mjeri sudjeluju u stvaranju potomstva.

Da bi populacija evoluirala dovoljan je izostanak samo jednog od prethodno navedenih uvjeta. U stvarnosti je gotovo nemoguće postići Hardy - Weinbergovu ravnotežu, odnosno zbog nedostatka bilo jednog ili više prethodno navedenih uvjeta, udjeli određenih alela stalno su podložni promjenama.

Prirodno se nameće pitanje: *"Kakva je korist od Hardy - Weinbergova modela ravnoteže?"*. Odgovor na to pitanje postat će jasan nakon izgradnje jednostavnog matematičkog modela za ovaj problem.

Matematički model

U postupku izgradnje matematičkog modela koji opisuje Hardy - Weinbergovu ravnotežu u populaciji krenut ćemo od promatranja jedinstveno određenog lokusa[4] u ljudskom genomu[5]. Radi jednostavnosti pretpostavit ćemo da na tom lokusu mogu postojati samo dvije varijante gena (dva alela) koji je odgovoran za točno određeno svojstvo (jedan alel naslijeđen je od majke, a drugi od oca). Označimo ih A i a. Također, pretpostavimo da A označava dominantan alel (alel koji određuje fenotip jedinke), a a recesivan alel (alel čije djelovanje nije očito u fenotipu jedinke). Kao posljedica ovih pretpostavki na tom se lokusu može realizirati samo jedan od sljedećih genotipa[6]:

- AA - od oba roditelja je naslijeđen dominantni alel A,

[3] Fenotip je skup svih genetski određenih svojstava jedinke.
[4] Dio molekule DNA na kojem se nalazi određeni gen.
[5] Cjelokupni genetički materijal neke vrste.
[6] Skup svih gena koji određuju točno određeno svojstvo nekog organizma.

- $Aa = aA$ - od jednog je roditelja naslijeđen dominantni alel A, a od drugog recesivni alel a,
- aa - od oba roditelja je naslijeđen recesivni alel a.

Kako ćemo pri izgradnji ovog modela koristiti elementarni matematički aparat iz područja teorije vjerojatnosti, bitno je uspostaviti svojevrsnu vezu između genetike i te grane matematike. Tu vezu počet ćemo graditi definiranjem nekih od osnovnih objekata vjerojatnosne teorije:

1) $\Omega = \{AA, Aa, aa\}$ - skup svih mogućih ishoda nekog slučajnog pokusa. U našem slučaju skup svih mogućih genotipa na promatranom lokusu.
2) $\mathcal{F} = \{\emptyset, \{AA\}, \{Aa\}, \{aa\}, \{AA, Aa\}, \{AA, aa\}, \{Aa, aa\}, \Omega\}$ - pripadna algebra događaja, odnosno familija svih podskupova skupa Ω.
3) $\mathbb{P} : \mathcal{F} \to [0, 1]$ - vjerojatnost događaja.

Uređenu trojku $(\Omega, \mathcal{F}, \mathbb{P})$ nazivamo vjerojatnosnim prostorom. Svaki element skupa Ω naziva se elementarni događaj, a svaki element familije $\mathcal{F}$ događaj. Za potrebe ovog članka dovoljno je poznavati samo vjerojatnosti elementarnih događaja $\{AA\}$, $\{Aa\}$ i $\{aa\}$. Njih možemo računati npr. prema klasičnoj definiciji vjerojatnosti, a bitno svojstvo tih vjerojatnosti je da se upravo oko njih grupiraju udjeli pripadnih događaja, što je ilustrirano u sljedećem jednostavnom primjeru:

Primjer 1:

Pretpostavimo da na raspolaganju imamo uzorak koji se sastoji od četiri osobe: jedna osoba na promatranom lokusu ima genotip AA, druge dvije Aa, a četvrta aa.

(a) Slučajan pokus sastoji se od izbora proizvoljne osobe iz zadanog uzorka. Budući se uzorak sastoji od po jedne osobe s genotipom AA, odnosno aa, i dvije osobe s genotipom Aa, prirodno je zaključiti da je vjerojatnost odabira osobe koja na promatranom lokusu ima genotip AA jednaka $1/4$, osobe koja na promatranom lokusu ima genotip Aa $1/2$, a osobe koja na promatranom lokusu ima genotip aa $1/4$.

(b) Pretpostavimo da prethodno definiran slučajni pokus ponovimo n puta, tako da svaku osobu iz zadanog uzorka možemo odabrati više puta. Iskustvo pokazuje da je opravdano očekivati da je broj odabira osobe sa genotipom AA, Aa, odnosno aa sljedeći:

$$n_{AA} \approx \frac{n}{4}; \quad n_{Aa} \approx \frac{n}{2}; \quad n_{aa} \approx \frac{n}{4},$$

gdje n_{AA} označava broj odabranih osoba iz uzorka s genotipom AA, n_{Aa} broj odabranih osoba iz uzorka s genotipom Aa, a n_{aa} broj odabranih osoba iz uzorka s genotipom aa i vrijedi $n_{AA} + n_{Aa} + n_{aa} = n$. Odavde se lako vidi da su pripadni udjeli pojedinih genotipa sljedeći:

$$\frac{n_{AA}}{n} \approx \frac{1}{4}; \quad \frac{n_{Aa}}{n} \approx \frac{1}{2}; \quad \frac{n_{aa}}{n} \approx \frac{1}{4}.$$

Stoga je razumno 1/4 uzeti kao vjerojatnost da se pri jednom provođenju pokusa odabere osoba koja na promatranom lokusu ima genotip AA, odnosno aa, a 1/2 kao vjerojatnost da se pri jednom provođenju pokusa odabere osoba koja na promatranom lokusu ima genotip Aa. Također, iskustvo pokazuje da su, što je broj ponavljanja pokusa veći, udjeli pojedinih genotipa grupirani sve bliže prethodno navedenim vjerojatnostima.

◇

Općenito, vjerojatnosni model za razdiobu genotipa u populaciji je sljedeći:

$$p_{AA} = \mathbb{P}(\{AA\}),$$

$$p_{Aa} = \mathbb{P}(\{Aa\}),$$

$$p_{aa} = \mathbb{P}(\{aa\}),$$

pri čemu vrijedi:

$$p_{AA}, p_{Aa}, p_{aa} > 0, \quad p_{AA} + p_{Aa} + p_{aa} = 1.$$

Pri tome vjerojatnosti p_{AA}, p_{Aa} i p_{aa} ne moraju nužno biti jednake.

Nadalje, zanima nas kako utvrditi točne iznose vjerojatnosti p_{AA}, p_{Aa} i p_{aa}. Taj problem neposredno je vezan uz model razdiobe alela A i a u populaciji. Do tog modela lako dolazimo ako genotip na promatranom lokusu prestanemo gledati kao cjelinu i shvatimo ga kao par alela od kojih je jedan naslijeđen od oca, a drugi od majke. Budući da ne znamo koji je alel naslijeđen od kojeg roditelja, zaključujemo da su vjerojatnosti nasljeđivanja alela A, odnosno alela a, od istog roditelja jednake. Iz toga slijedi da vjerojatnosni model koji opisuje razdiobu alela u populaciji izgleda ovako:

$$p_A = \mathbb{P}(\{A\}),$$

$$p_a = \mathbb{P}(\{a\}),$$

pri čemu standardno vrijedi da su $p_A, p_a > 0$ i $p_A + p_a = 1$. Analogno kao kod modela razdiobe genotipa, vjerojatnosti p_A i p_a ne moraju nužno biti jednake, što je ilustrirano sljedećim primjerom:

Primjer 2:

Promatrani uzorak iz primjera 1 sastojao se od četiri osobe: genotipi AA i aa pripadali su po jednoj osobi, a genotip Aa dvjema osobama. Jednostavnim prebrojavanjem alela, utvrdili smo da se aleli A i a u genetičkoj strukturi uzorka pojavljuju svaki po četiri puta. Budući promatramo ukupno 8 alela, prema klasičnoj definiciji vjerojatnosti slijedi da su vjerojatnosti iz modela razdiobe alela sljedeće:

$$p_A = p_a = \frac{1}{2}.$$

No, kad bi se uzorak sastojao od četiri osobe od kojih bi tri osobe na promatranom lokusu imale genotip AA, a samo jedna osoba genotip Aa, tada bismo na isti način zaključili da su vjerojatnosti iz modela razdiobe alela sljedeće:

$$p_A = \frac{7}{8}; \quad p_a = \frac{1}{8}.$$

Dakle, vjerojatnosti p_A i p_a ne moraju nužno biti jednake.

$\diamond$

Uz poznate vjerojatnosti p_A i p_a iz modela razdiobe alela, zadovoljene specifične uvjete koji osiguravaju konstantnost udjela pojedinih genotipa u populaciji (a koji su navedeni na početku članka) i nezavisnost alela naslijeđenih od pojedinog roditelja, primjenom osnovnih svojstava iz definicije vjerojatnosti ([2], str. 107) lako slijedi:

$$\begin{aligned}
1)\ p_{AA} &= \mathbb{P}(\{A\},\{A\}) = \mathbb{P}(\{A\}) \cdot \mathbb{P}(\{A\}) = p_A^2 \\
2)\ p_{Aa} &= \mathbb{P}(\{\{A\},\{a\}\} \cup \{\{a\},\{A\}\}) = \mathbb{P}(\{A\},\{a\}) + \mathbb{P}(\{a\},\{A\}) = \\
&= \mathbb{P}(\{A\}) \cdot \mathbb{P}(\{a\}) + \mathbb{P}(\{a\}) \cdot \mathbb{P}(\{A\}) = p_A \cdot p_a + p_a \cdot p_A = 2p_A p_a \\
3)\ p_{aa} &= \mathbb{P}(\{a\},\{a\}) = \mathbb{P}(\{a\}) \cdot \mathbb{P}(\{a\}) = p_a^2.
\end{aligned}$$

Iz dobivenog prikaza modela razdiobe genotipa preko vjerojatnosti p_A i p_a, primjenom poznate formule $p_{AA} + p_{Aa} + p_{aa} = 1$ slijedi:

$$p_A^2 + 2p_A p_a + p_a^2 = 1,$$

pa prema formuli za kvadrat binoma imamo:

$$(p_A + p_a)^2 = 1.$$

Prethodna formula koristi se za izračunavanje udjela pojedinih alela u populaciji ako znamo da se ona nalazi u stanju Hardy - Weinbergove ravnoteže ili za provjeru uravnoteženosti populacije prema Hardy - Weinbergovom modelu ukoliko su nam poznati udjeli pojedinih alela. Upotreba formule ilustrirana je sljedećim primjerom:

Primjer 3:

Recimo da proučavamo boju krzna u populaciji miševa koja se nalazi u stanju Hardy - Weinbergove ravnoteže. Poznati su nam udjeli pojedinih alela koji određuju sivu (dominantno svojstvo određeno alelom A), odnosno bijelu (recesivno svojstvo određeno alelom a) boju krzna:

$$p_A = 0.8; \quad p_a = 0.2.$$

Pod uvjetima Hardy - Weinbergove ravnoteže sada je vrlo jednostavno odrediti udjele pojedinih genotipa u populaciji:

$$p_{AA} = p_A^2 = 0.64$$

$$p_{Aa} = 2p_A p_a = 0.32$$
$$p_{aa} = p_a^2 = 0.04$$

pri čemu je očito da je $p_{AA}, p_{Aa}, p_{aa} > 0$ i $p_{AA} + p_{Aa} + p_{aa} = 0.64 + 0.32 + 0.04 = 1$.

Budući je siva boja krzna dominantno svojstvo određeno alelom A iz prethodnog proračuna slijedi da će 96% $(0.64 + 0.32 = 0.96)$ miševa imati krzno sive, a samo 4% bijele boje. Postavlja se pitanje: "*Hoće li miševi bijele boje s vremenom nestati?*".

Odgovor je negativan, a razlog leži u razdiobi alela A i alela a u spolnim stanicama ili gametama neophodnim za razmnožavanje miševa. Naime, gamete koje prenose dominantno svojstvo sadrže alel A, dok gamete koje prenose recesivno svojstvo sadrže alel a. Alel A bit će sadržan u svim gametama miševa sa genotipom AA i polovici gameta miševa sa genotipom Aa, dakle u ukupno $64\% + 32\%/2 = 80\%$ gameta. Analogno, alel a bit će sadržan u svim gametama miševa sa genotipom aa i polovici gameta miševa sa genotipom Aa, dakle u ukupno $4\% + 32\%/2 = 20\%$ gameta. Kako je alel A dominantan, svi potomci sa genotipima AA i Aa imat će krzno sive boje (njih 80%). Ostatak potomstva na promatranom lokusu ima genotip aa, odnosno ima krzno bijele boje (takvih potomaka je 20%).

Dakle, vidimo da se pod uvjetima Hardy - Weinbergove ravnoteže udjeli pojedinih alela u populaciji ne mijenjaju iz generacije u generaciju.

◇

Također, na vrlo jednostavan način možemo dobiti prikaz vjerojatnosti iz modela razdiobe alela pomoću vjerojatnosti iz modela razdiobe genotipa. U tu svrhu pretpostavimo da smo uzeli slučajan uzorak iz populacije koji se sastoji od n osoba. Budući se bavimo samo genotipima tih osoba na točno određenom lokusu, promatrat ćemo slučajan uzorak koji se sastoji od točno n genotipa. Prema klasičnoj definiciji vjerojatnosti slijedi da su vjerojatnosti iz modela razdiobe genotipa u promatranom uzorku sljedeće:

$$p_{AA} = \frac{n_{AA}}{n}; \quad p_{Aa} = \frac{n_{Aa}}{n}; \quad p_{aa} = \frac{n_{aa}}{n},$$

gdje n_{AA} označava broj osoba iz uzorka s genotipom AA, n_{Aa} broj osoba iz uzorka s genotipom Aa, a n_{aa} broj osoba iz uzorka s genotipom aa.

Na isti način određujemo i vjerojatnosti iz modela razdiobe alela u promatranom uzorku:

$$p_A = \frac{2 \cdot n_{AA} + n_{Aa}}{2n}; \quad p_a = \frac{2 \cdot n_{aa} + n_{Aa}}{2n}.$$

Iz prethodno određenih vjerojatnosti koje tvore modele razdiobe genotipa i alela u populaciji slijedi:

$$p_A = \frac{2 \cdot n_{AA} + n_{Aa}}{2n} = \frac{2 \cdot n_{AA}}{2n} + \frac{n_{Aa}}{2n} = \frac{n_{AA}}{n} + \frac{1}{2} \cdot \frac{n_{Aa}}{n} = p_{AA} + \frac{1}{2} \cdot p_{Aa}.$$

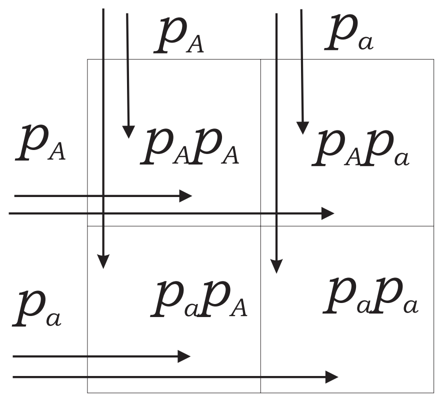
Na analogan način dobivamo:

$$p_a = p_{aa} + \frac{1}{2} \cdot p_{Aa}.$$

Iz ovog je zapisa očito da su vjerojatnosti iz modela razdiobe alela u potpunosti određene vjerojatnostima iz modela razdiobe genotipa.

Punnettov kvadrat

Punnettov[7] kvadrat, koji se vrlo često koristi u genetici, svojevrsno je mnemotehničko pravilo za vizualizaciju određivanja udjela pojedinih genotipa iz poznatih udjela alela u populaciji, a proizlazi iz geometrijskog dokaza formule za kvadrat binoma[8]. On izgleda ovako:



Slika 1. Punnettov kvadrat

Ilustrirajmo upotrebu Punnettovog kvadrata na primjeru 2, u kojem su nam poznati točni udjeli pojedinih alela u populaciji:

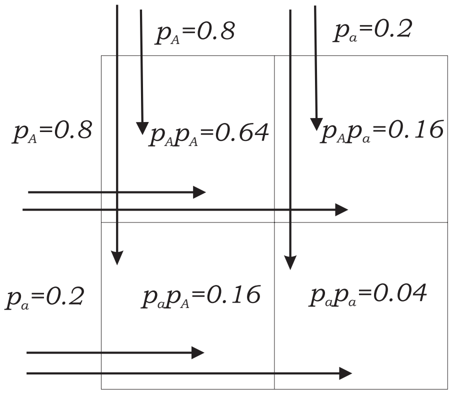
Primjer 4:

Poznate su nam vjerojatnosti iz modela razdiobe alela:

$$p_A = 0.8, \quad p_a = 0.2$$

Pomoću Punnettovog kvadrata izračunajmo vjerojatnosti iz modela razdiobe genotipa:

[7] Reginald Punnett (1875. - 1967.), britanski genetičar.
[8] http://www.leonelearningsystems.com/BinomialSquareExplained.pdf



Odavde lako slijedi:

$$\begin{aligned} p_{AA} &= p_A^2 = 0.64, \\ p_{Aa} &= 2p_A p_a = 2 \cdot 0.16 = 0.32, \\ p_{aa} &= p_a^2 = 0.04. \end{aligned}$$

◇

Zaključak

Budući se svi uvjeti koji populaciju dovode u stanje Hardy - Weinbergove ravnoteže gotovo nikada ne poklope, prethodno provedene kalkulacije koriste se za donošenje različitih procjena udjela određenih genotipa u populaciji. Te procjene daju vrlo dobre i primjenjive rezultate kad se radi o velikim populacijama u kojima su fenotipska preferencija pri izboru partnera i utjecaj okoliša na genetičku strukturu svedeni na minimum, dok su rezultati vezani za male populacije loši. Iako je Hardy - Weinbergov model ravnoteže u populaciji idealizacija stvarne situacije, vrlo je primjenjiv i koristan u istraživanjima.

Zadaci za vježbu

Zadatak 1:
Boju leptirovih krila određuju kodominantni[9] aleli B (crvena boja) i b (plava boja). U uzorku od 1612 leptira njih 1469 ima genotip BB (krila crvene boje), 138 genotip Bb (krila ljubičaste boje), a samo 5 genotip bb (krila plave boje). Odredite razdiobe alela i genotipa u ovom uzorku leptira.

Rješenje: $p_B = 0.954$, $p_b = 0.046$, $p_{BB} = 0.91$, $p_{Bb} = 0.087$, $p_{bb} = 0.002$.

Zadatak 2:
Jednog lijepog sunčanog dana grupa od 20 prijatelja odlučila je izgraditi splav,

[9] Za alele kažemo da su kodominantni kada je djelovanje oba alela očito u fenotipu jedinke.

otploviti na pusti otok i stvoriti novu populaciju potpuno izoliranu od vanjskog svijeta. Ni jedan član grupe ne boluje od albinizma, ali dva člana ove grupe nosioci su recesivnog alela x koji uzrokuje albinizam kod osoba sa genotipom xx. Pod pretpostavkom da se nova populacija nalazi u stanju Hardy - Weinbergove ravnoteže, izračunajte koliki će udio nove populacije bolovati od albinizma?

Rješenje: $p_{xx} = 0.0025$.

Literatura

[1] P. Almgren, P. Bendhal, H. Bengtsson, O. Hössjer, R. Perfekt, *Statistics in Genetics - lecture notes*, Lund Institute of Technology - Centre for Mathematical Sciences, on-line izdanje, 2002.

[2] N. Elezović, *Matematika 4 - udžbenik za četvrti razred gimnazije*, Element, Zagreb, 1997.

[3] I. Iachine, *Statistical Methods in Genetic Epidemiology*, University of Southern Denmark, on-line izdanje

[4] S. Jelinić, M. Kerovac, I. Ternjej, Z. Mihaljević, *Biologija 4 - ekologija, evolucija, genetika*, Profil, Zagreb, 2004.

[5] `http://science.nhmccd.edu/biol/hwe.html`

[6] `http://www.bookrags.com/sciences/genetics/hardy-weinberg-equilibrium-wog.html`

[7] `http://www.monarchwatch.org/biology/theory.htm`