

UNIVERZITET U BEOGRADU - MATEMATIČKI  
FAKULTET



EVOLUTIVNA KVANTITATIVNA  
GENOMIKA

*Seminarski rad*

*iz predmeta:*

LINEARNI STATISTIČKI MODELI

**PROFESOR:**

Dr Bojana Milošević

**ASISTENT:**

Luka Perović

**STUDENTI:**

Gala Posedi 064/2021

Ana Jevtić 279/2021

Sara Crnomarković 139/2021

Beograd, septembar, 2025.

# Sadržaj

<b>1 Sličnosti, disperzije i aditivne genetičke vrednosti</b>	<b>4</b>
1.1 Fišerova genetička dekompozicija . . . . .	4
1.2 Aditivne genetičke disperzije i kovarijacije . . . . .	5
<b>2 Regresije roditelj–potomak i odgovor na selekciju</b>	<b>6</b>
2.1 Jednoosobinske regresije roditelj–potomak . . . . .	6
2.2 Preciznost procene heritabilnosti $h^2$ . . . . .	7
2.2.1 Analiza braće i sestara . . . . .	8
2.3 Selektioni diferencijali i Brederova jednačina . . . . .	8
2.4 Predikcija genetske vrednosti ( <i>breeding value</i> ) . . . . .	8
2.5 Višeosobinske regresije roditelj–potomak . . . . .	9
2.6 Genetičke i fenotipske kovarijacione matrice . . . . .	9
2.7 Multivarijantna brederova jednačina . . . . .	9
<b>3 Infinitesimalni model</b>	<b>9</b>
3.1 Linearost regresije roditelj–potomak u infinitesimalnom modelu . . . . .	10
3.2 Promene učestalosti alela u infinitesimalnom modelu . . . . .	10
3.3 Promene disperzije . . . . .	10
3.4 Ravnotežna aditivna genetička disperzija . . . . .	11
<b>4 Procena <math>\sigma_A^2</math> i <math>G</math></b>	<b>12</b>
4.1 Animal Model . . . . .	12
4.1.1 Primer Animal Model-a . . . . .	13
<b>5 Fitness</b>	<b>15</b>
5.1 Epizode selekcije . . . . .	16
<b>6 Robertson–Price identitet i Teoreme selekcije</b>	<b>17</b>
<b>7 Mogućnost selekcije</b>	<b>19</b>
<b>8 Koeficijenti selekcije</b>	<b>20</b>
8.1 Mere selekcije srednje vrednosti . . . . .	20
8.2 Mere selekcije varijacije . . . . .	21
<b>9 Fitness funkcije i karakterizacija selekcije</b>	<b>22</b>
9.1 Individualne i prosečne fitness funkcije . . . . .	22
9.2 Gradijenti i lokalna geometrija fitness površi . . . . .	24
<b>10 Multivarijantna selekcija</b>	<b>25</b>
10.1 Kratkoročne promene u srednjim vrednostima: Multivarijantna Brederova jednačina . . . . .	25
10.2 Efekti genetičkih korelacija: direktni i korelisani odgovori . . . . .	25
10.3 "Path" analiza i prošireni gradijenti selekcije . . . . .	27
<b>11 Procena gradijenata selekcije</b>	<b>27</b>
11.1 Analiza metodom MNK . . . . .	27
11.2 Fleksibilna procena fitness funkcija sa pridruženim gradijentima selekcije . . . . .	28



# Uvod

Evolutivna kvantitativna genetika je naučna disciplina koja proučava kako se promene u frekvenciji alela za osobine koje zavise od više gena (kvantitativne osobine) dešavaju tokom vremena pod uticajem evolutivnih sila, kao što su selekcija, mutacija, genetski drift i migracija. Cilj joj je da objasni evoluciju ovih složenih osobina i da predviđa kako će populacije reagovati na promene u okruženju.

U ovom radu razmatramo osnovne principe evolutivne kvantitativne genetike. Prvo, obrađujemo koncept aditivne genetičke disperzije, koja je komponenta genetičke disperzije najrelevantnija za evolutivne odgovore na selekciju. Zatim razmatramo evolutivni odgovor na selekciju (tj. kako se transgeneracijski menja srednja vrednost fenotipske osobine) kao odgovor na unutar-generacijske promene u sredini. Naredno, razmatramo dinamiku osobina koje kontroliše veliki broj lokusa, kao i modeli koji opisuju kako srednja vrednost osobine i aditivna genetička disperzija reaguju na selekciju tokom više generacija. Na kraju, bavimo se praktičnjim aspektima evolutivne kvantitativne genetike, uključujući procenu veličina kao što su aditivne genetičke varijanse, definisanje fitnessa u prirodi i merenje prirodne selekcije u realnim populacijama.

Cilj ove oblasti je dublje razumevanje ne samo genetike i nasleđivanja složenih osobina u prirodi, već i prirode evolutivnih sila koje oblikuju disperzije osobina i njihove promene u prirodnim populacijama.

## Osnovni pojmovi genetike i genomike

**Dezoksiribonukleinska kiselina (DNK)** - molekul života koji sadrži uputstva za građu i funkcionalisanje organizma.

**Gen** - deo DNK koji nosi informaciju za jedan protein ili RNK molekul.

**Alel** - različita verzija istog gena.

**Hromozom** - upakovan molekul DNK. Ljudi imaju 46 hromozoma (23 para), po jedan iz svakog para dolazi od oca i majke.

**Genom** - kompletna DNK organizma, tj. svi hromozomi zajedno.

**Genotip i fenotip** - genotip je skup gena koje jedinka nosi, dok je fenotip način na koji se oni ispoljavaju u spoljašnjem svetu.

**Mutacija** - trajna promena u DNK sekvenci. Može biti korisna, neutralna ili štetna.

**Lokus** - tačno određena pozicija na hromozomu gde se nalazi određeni gen ili DNK sekvenca.

## 1 Sličnosti, disperzije i aditivne genetičke vrednosti

U ovom odeljku uvodimo osnovnu terminologiju i principe koji su neophodni pre razmatranja naprednijih i aktuelnih pitanja u nastavku poglavila.

### 1.1 Fišerova genetička dekompozicija

Začetnik moderne kvantitativne genetike bio je Ronald Fišer, koji je u svom radu iz 1918. godine dao niz ključnih uvida, posebno da:

- (i) kod polnog razmnožavanja samo određena komponenta genotipske vrednosti jedinke se prenosi na potomstvo;
- (ii) disperzije povezane sa ovom i drugim komponentama mogu se proceniti iz fenotipskih kovarijacija između odgovarajućih srodnika.

Fišer je dekomponovao posmatranu fenotipsku vrednost  $z$  jedinke na genotipsku  $G$  i ekološku  $E$  komponentu:

$$z = G + E. \quad (1.1)$$

Genotipska vrednost može se posmatrati kao prosečan fenotip (srednja vrednost), ako bi jedinka bila klonirana i replicirana u svakom mogućem realnom okruženju. Kako je Fišer primetio, roditelj ne prenosi celu svoju  $G$  vrednost na potomstvo, jer za svaki dati lokus (specifično mesto gena na hromozomu) roditelj prenosi samo jedan od svoja dva alela (različiti oblici istog gena). Zbog toga se genotipska vrednost može dalje razložiti na aditivnu komponentu  $A$  koja se prenosi na potomstvo i neaditivnu komponentu, koja uključuje dominaciju  $D$  i eventualne epistatičke efekte. U ovom radu ćemo zanemariti epistatičke efekte na gene. Dakle:

$$z = \mu + A + D + E, \quad (1.2)$$

gde je  $\mu$  populacioni prosek (po konstrukciji, srednje vrednosti  $A$ ,  $D$  i  $E$  jednake su nuli, tako da  $G = \mu + A + D$  u jednačini (1.1)).

Aditivne vrednosti  $A$  se često nazivaju *vrednostima za oplemenjivanje* (breeding values), jer je prosečan fenotip potomstva upravo prosečna vrednost za oplemenjivanje roditelja. Dakle, očekivana fenotipska vrednost potomstva je:

$$\mu + \frac{A_f + A_m}{2},$$

gde su  $A_f$  (očev) i  $A_m$  (majčin) vrednosti za oplemenjivanje.

Aditivne genetičke vrednosti su po definiciji nezavisne od svih ostalih efekata. Zbog njihove nezavisnosti i aditivnosti na nivou jedinki, disperzija ovih efekata je aditivna i na nivou populacije:

$$\sigma_z^2 = \sigma_A^2 + \sigma_D^2 + \sigma_E^2, \quad (1.3)$$

## 1.2 Aditivne genetičke disperzije i kovarijacije

Kada se fenotip može dekomponovati kao u jednačini (1.2), Fišer je pokazao da je doprinos aditivnih genetičkih efekata fenotipskoj kovarijaciji jednaka:

$$\sigma_A(z_p, z_o) = \frac{\sigma_A^2}{2}. \quad (1.4)$$

Gde su  $z_p, z_o$  redom fenotip roditelja i fenotip potomka. Koeficijent  $\frac{1}{2}$  pojavljuje se zato što roditelji i potomci dele polovinu svojih gena. Dakle, dvostruka fenotipska kovarijansa roditelj–potomak daje procenu  $\sigma_A^2$ , koja se naziva aditivna genetička varijansa. Na ovaj način možemo kvantifikovati koliko genetika utiče na disperziju fenotipa, koristeći srodnost roditelj–potomak.

Dodatno, jednačina (1.4) može se koristiti da opiše sličnosti između proizvoljne klase srodnika. Pod pretpostavkom nasumičnog ukrštanja:

$$\sigma_A(z_a, z_b) = \sigma_A^2 \cdot 2\Theta_{ab}, \quad (1.5)$$

gde je  $\Theta$  koeficijent koancestrijе (mera genetske sličnosti između dve jedinke).  $\Theta_{ab}$  je verovatnoćа da su dva alela, nasumično izabrana iz svake od dve jedinke  $a$  i  $b$ , identična po poreklu.

Genetska povezanost između osobina pokazuje kako višestruke osobine zajedno reaguju na selekciju. Zato je aditivna genetička kovarijacija  $\sigma_A(x, y)$  važna za razumevanje evolucije više osobina.

Ove aditivne genetičke kovarijanse mogu se proceniti i regresijama roditelj–potomak:

$$\sigma(x_p, y_o) = \sigma(y_p, x_o) = \frac{\sigma_A(x, y)}{2}. \quad (1.6)$$

Kovarijansa između osobina može nastati na dva načina: kroz neuravnotežena ukrštanja (*linkage disequilibrium*), gde su aleli različitih lokusa međusobno povezani i kratkoročno stvaraju korelaciju, koja vremenom nestaje rekombinacijom, i kroz *pleiotropiju*, gde jedan lokus utiče na više osobina, stvarajući stabilnu kovarijaciju tokom vremena.

## 2 Regresije roditelj–potomak i odgovor na selekciju

U ovom odeljku selekciju ćemo posmatrati jednostavno–kao bilo kakvu promenu u raspodeli fenotipa unutar generacije. Opštije, to je razlika u raspodeli fenotipa među reproduktivnim jedinkama, ponderisana plodnošću, i raspodeli među svim novonastalim jedinkama. Selekcija menja srednju vrednost i/ili raspodelu osobine u populaciji jer neke jedinke daju više potomaka od drugih.

### 2.1 Jednoosobinske regresije roditelj–potomak

Većina teorije o odgovoru na selekciju u kvantitativnoj genetici prepostavlja su regresije roditelj–potomak linearne i da imaju homoskedastične reziduale. Najjednostavnija verzija razmatra fenotipsku vrednost srednjeg roditelja,  $z_{mp} = (z_m + z_f)/2$ , pri čemu se fenotip potomka  $z_o$  predviđa kao:

$$z_o = \mu + b(z_{mp} - \mu), \quad (2.1)$$

gde je  $\mu$  populacioni prosek, a  $b$  koeficijent linearne regresije. Kako je ocena za  $b$ , dobijena metodom najmanjih kvadrata, u opštem slučaju jednaka:

$$b = \frac{\sigma(x, y)}{\sigma_x^2},$$

koeficijent uz prediktor kod regresije roditelj–potomak postaje

$$b = \frac{\sigma(z_o, z_{mp})}{\sigma^2(z_{mp})} = \frac{\sigma(z_o, z_m)/2 + \sigma(z_o, z_f)/2}{\sigma^2(z_m/2 + z_f/2)} = \frac{2\sigma_A^2/4}{2\sigma_z^2/4} = \frac{\sigma_A^2}{\sigma_z^2}. \quad (2.2)$$

Ovaj odnos aditivne genetičke i fenotipske disperzije označavamo sa  $h^2$  i naziva se uskougaona heritabilnost (*narrow-sense heritability*). Dakle, regresiona funkcija roditelj–potomak data je kao:

$$E[z_o | z_{mp}] = \mu + \frac{\sigma(z_o, z_{mp})}{\sigma^2(z_{mp})}(z_{mp} - \mu), \quad (2.3)$$

gde je:

- $z_o$  – fenotip potomka,
- $z_{mp} = \frac{z_m + z_f}{2}$  – srednji fenotip roditelja,
- $\mu$  – populacioni prosek,
- $\sigma(z_o, z_{mp})$  – kovarijansa između potomka i srednjeg roditelja,
- $\sigma^2(z_{mp})$  – varijansa srednjeg roditelja.

Korišćenjem aditivne genetičke disperzije  $\sigma_A^2$  i fenotipske disperzije  $\sigma_z^2$ , dobijamo pojednostavljenu formu:

$$E[z_o | z_{mp}] = \mu + h^2(z_{mp} - \mu), \quad h^2 = \frac{\sigma_A^2}{\sigma_z^2}, \quad (2.4)$$

Stvarna vrednost za konkretnog potomka data je kao:

$$z_o = \mu + h^2(z_{mp} - \mu) + e, \quad (2.5)$$

gde rezidual  $e$  regresije roditelj–potomak ima srednju vrednost nula i varijansu

$$\sigma_e^2 = (1 - h^4) \sigma_z^2. \quad (2.6)$$

## 2.2 Preciznost procene heritabilnosti $h^2$

U kvantitativnoj genetici, pored same vrednosti hereditabiliteta  $h^2$ , veoma je važna i njena preciznost, koju označavamo standardnom greškom  $SE(h^2)$ . Preciznost zavisi od disperzije – manja disperzija povlači veću preciznost.

Neka je  $X$  vrednost roditelja (ili srednja vrednost oba roditelja), a  $Y$  srednja vrednost potomaka. Za broj potomaka po porodici  $n$  i broj porodica  $N$ , ukupno izmereno  $T$  jedinki:

$$T = \begin{cases} N(n+1), & \text{ako je merena jedna roditeljska jedinka} \\ N(n+2), & \text{ako su merena oba roditelja} \end{cases}$$

Za procenu heritabilnosti:

$$SE(h^2) \approx \sqrt{2} SE(b_{\text{offspring-parent}}), \quad \text{za regresiju na jednog roditelja.}$$

$$SE(h^2) \approx SE(b_{\text{mid-parent}}), \quad \text{za regresiju na mid-parent.}$$

Ovde  $b_{\text{mid-parent}}$  i  $b_{\text{offspring-parent}}$  označavaju koeficijent b regresije iz jednačine (2.2).

### 2.2.1 Analiza braće i sestara

Za full-sib porodice:

$$h^2 = 2t, \quad \text{Var}(h^2) \approx \frac{16t(1-t)^2}{T}$$

Za half-sib porodice:

$$h^2 = 4t, \quad \text{Var}(h^2) \approx \frac{64t(1-t)^2}{T}$$

Optimalna veličina porodica zavisi od očekivane heritabilnosti; greška je veća ako je porodica manja od optimalne.

## 2.3 Selepcioni diferencijali i Brederova jednačina

Regresija roditelj–potomak omogućava nam da predvidimo odgovor na selekciju. Pretpostavimo da se srednja fenotipska vrednost roditelja koji se reprodukuju,  $\mu^*$ , razlikuje od populacionog proseka pre selekcije,  $\mu$ . Definišemo direkcioni selepcioni diferencijal kao  $S = \mu^* - \mu$  ( $S$  se često jednostavno naziva selepcioni diferencijal).

Iz jednačine (2.4), razlika  $R$  između prosečne vrednosti u potomstvu tih roditelja i originalnog proseka populacije pre selekcije data je sa:

$$R = h^2 S. \quad (2.7)$$

Ovo je *brederova jednačina* (*breeder's equation*), koja transformiše unutar-generacijsku promenu u sredini ( $S$ ) u među-generacijsku promenu u sredini (odgovor na selekciju,  $R$ ).

Čak i ako selekcija jako favorizuje određene jedinke ( $S$ ), evolucijski odgovor ( $R$ ) može biti mali ili nikakav ako osobina slabo nasledna (niska heritabilnost, tj mala vrednost  $h$ ). Dakle, selekcija ne garantuje promenu u populaciji.

## 2.4 Predikcija genetske vrednosti (*breeding value*)

Predikcija genetske vrednosti pojedinca na osnovu njegovog fenotipskog zapisu je praktična primena prethodnog. Kada postoji samo jedan fenotipski zapis po jedinku (*single record*), procenjena genetska vrednost  $\hat{a}_i$  za jedinku  $i$  se računa kao:

$$\hat{a}_i = h^2(y_i - m),$$

gde je  $y_i$  fenotipski zapis jedinke,  $m$  prosečna vrednost u istoj upravljačkoj grupi, a  $h^2$  mera heritabilnost. Ova procena predstavlja korelaciju između selepcionog kriterijuma i stvarne genetske vrednosti, poznatu kao tačnost predikcije ( $r_{ay}$ ):

$$r_{ay} = \sqrt{h^2}.$$

Kombinacija selepcionih diferencijala i predikcije genetskih vrednosti omogućava precizno planiranje selekcije. Brederova jednačina povezuje promenu u proseku populacije sa heritabilnošću, dok predikcija breeding value pojedinca omogućava procenu genetskog potencijala na osnovu dostupnih fenotipskih zapisa.

## 2.5 Višeosobinske regresije roditelj–potomak

Regresija roditelj–potomak za jednu osobinu može se generalizovati na vektor od  $n$  osobina,  $z = (z_1, \dots, z_n)^T$ . Neka je  $z_o$  vektor vrednosti osobina kod potomstva,  $z_{mp}$  vektor srednjih vrednosti roditelja (tj.  $i$ -ti element je  $(z_{f,i} + z_{m,i})/2$ ), a  $\mu$  vektor populacionih sredina. Tada je regresiona funkcija roditelj–potomak za više osobina jednaka:

$$\mathbb{E}[z_o] = \mu + H(z_{mp} - \mu), \quad (2.8)$$

gde je  $H$  multivarijantna generalizacija heritabilnosti  $h^2$ .

## 2.6 Genetičke i fenotipske kovarijacione matrice

Da bismo dalje dekomponovali  $H$ , potrebno je definisati fenotipsku kovarijacionu matricu  $P$  čiji je element  $ij$  fenotipska kovarijacija između osobina  $i$  i  $j$ .

Na sličan način, možemo definisati simetričnu matricu  $G$ , čiji je element  $ij$  aditivna genetička kovarijacija između osobina  $i$  i  $j$ . Analogno kao kod jednačine (2.2) (koeficijent uz prediktor regresije za jednu osobinu), može se pokazati da je:

$$H = GP^{-1},$$

pa je regresiona funkcija jednaka:

$$\mathbb{E}[z_o] = \mu + GP^{-1}(z_{mp} - \mu). \quad (2.9)$$

## 2.7 Multivarijantna brederova jednačina

Neka  $R$  predstavlja vektor evolutivnog odgovora (tako da  $i$ -ti element predstavlja među-generacijsku promenu srednje vrednosti fenotipske osobine  $i$ ), a  $S$  vektor selekcionih diferencijala. Jednačina (2.9) nam pruža generalizaciju Brederove jednačine na više osobina:

$$R = HS = GP^{-1}S. \quad (2.10)$$

Ova jednačina pokazuje kako genetska kovarijacija i genetske veze između osobina određuju koliko će promene izazvane selekcijom u jednoj generaciji (kod roditelja) zapravo uticati na promene u sledećoj generaciji (kod potomaka).

$H$  predstavlja višedimenzionalni heritabilitet, kao analog za jednodimenzionalni koeficijent  $h^2$ .

## 3 Infinitesimalni model

Brederova jednačina predviđa promenu sredine nakon jedne generacije selekcije iz neselektovane osnovne populacije. Glavna prepostavka potrebna je da je regresija roditelj–potomak linearna.

Postavlja se pitanje kada ta linearost važi. Takođe, Brederova jednačina se fokusira samo

na promenu sredine, ostavljajući otvorenim pitanje šta se dešava sa disperzijom (fenotipskom i genetičkom). Prethodno je od posebnog značaja, jer se smatra da je stabilizujuća selekcija česta u prirodnim populacijama.

Infinitesimalni model, uveden od strane Fišera, 1918. godine.

### 3.1 Linearost regresije roditelj–potomak u infinitesimalnom modelu

U infinitesimalnom modelu prepostavljamo da mnogi (beskonačan broj) lokusa doprinose genetičkoj varijaciji za datu osobinu. Dakle, prema centralnoj graničnoj teoremi, genotipske vrednosti ( $G$ ) i vrednosti za oplemenjivanje normalno su raspodeljene pre selekcije. Ako su i ekološke vrednosti ( $E$ ) normalne, tada je i fenotip  $z$  normalan (kao linearna kombinacija).

U ovom slučaju, regresija fenotipske vrednosti potomka  $z_o$  na fenotipove roditelja ( $z_f$  i  $z_m$  za oca i majku) je linearna i homoskedastična.

### 3.2 Promene učestalosti alela u infinitesimalnom modelu

U infinitesimalnom modelu osobina je određena velikim brojem gena sa vrlo malim efektom. Iako selekcija praktično ne menja učestalost pojedinačnih alela, zbirni efekti na mnogim lokusima mogu izazvati velike promene prosečne vrednosti osobine, dok disperzija populacije ostaje gotovo nepromenjena.

U ovom modelu, fenotipska osobina je određena sa  $n$  aditivnih lokusa, svaki sa malim efektom  $a = O(n^{-1/2})$ . Ako se učestalost alela promeni za  $\Delta p$ , srednja vrednost se menja:

$$\Delta\mu = 2na \Delta p = O(n^{1/2} \Delta p).$$

Aditivna disperzija je:

$$\sigma_A^2 = 2na^2 p(1-p),$$

$$\Delta\sigma_A^2 = 2na^2 \Delta p(1 - 2p - \Delta p) \approx O(a\Delta\mu) = O(n^{-1/2} \Delta\mu).$$

Kako je  $a = O(n^{-1/2})$ , dobijamo:

$$\Delta\sigma_A^2 \ll \Delta\mu \quad \text{i pri } n \rightarrow \infty \quad \Delta\sigma_A^2 \rightarrow 0, \text{ dok } \Delta\mu \neq 0.$$

Dakle, prosečna vrednost može se značajno promeniti bez promene aditivne varijanse.

### 3.3 Promene disperzije

Pod infinitesimalnim modelom, genetičke disprezije se praktično ne menjaju. Selekcija ipak je može menjati stvaranjem kovarijacija među alelima različitih lokusa (gametska neravnoteža), što utiče na genetičku disperziju i može narušiti normalnost i linearost regresije roditelj–potomak.

Ovde se fokusiramo na posledice gametske neravnoteže za aditivnu genetičku varijansu. Aditivna genetička varijansa u trenutku  $t$ ,  $\sigma_{A,t}^2$ , u prisustvu zavisnosti alela na različitim lokusima može se zapisati kao:

$$\sigma_{A,t}^2 = \sigma_a^2 + d_t, \quad (3.1)$$

gde je  $\sigma_a^2$  genička varijansa (aditivna varijansa u odsustvu neravnoteže), a  $d_t$  doprinos neravnoteže u trenutku  $t$ .

Pretpostavljajući nevezane lokuse, Bulmer je pokazao da je promena neravnoteže data sa:

$$d_{t+1} = \frac{d_t}{2} + \frac{h_t^4}{2} (\sigma_{z^*,t}^2 - \sigma_{z,t}^2), \quad (3.2)$$

gde je  $\sigma_{z^*,t}^2 - \sigma_{z,t}^2$  unutar-generacijska promena fenotipske disperzije. Prvi član predstavlja uklanjanje neravnoteže rekombinacijom, dok drugi predstavlja generisanje neravnoteže selekcijom. Prethodna jednačina je analog Brederove jednačine za disperziju, povezujući među-generacijske ( $d_{t+1} - d_t$ ) i unutar-generacijske promene ( $\sigma_{z^*,t}^2 - \sigma_{z,t}^2$ ).

Selekcija može menjati aditivnu genetičku varijansu i bez promene srednje vrednosti fenotipa. Stabilizujuća selekcija smanjuje disperziju uklanjanjem ekstremnih jedinki, dok disruptivna selekcija povećava disperziju uklanjanjem jedinki sa srednjim vrednostima. Nakon prestanka selekcije, aditivna varijansa se obično brzo vraća na svoju početnu vrednost.

Za modeliranje se često pretpostavlja da se unutar-generacijska promena fenotipske disperzije može zapisati kao:

$$\sigma_{z^*,t}^2 = (1 - \kappa) \sigma_{z,t}^2, \quad (3.3)$$

gde je  $\kappa$  konstanta. Ako se ovo ubaci u (3.2), dobija se rezultat Bulmer-a:

$$d_{t+1} = \frac{d_t}{2} - \frac{\kappa}{2} h_t^2 \sigma_{A,t}^2 = \frac{d_t}{2} - \frac{\kappa}{2} \frac{(\sigma_a^2 + d_t)^2}{\sigma_z^2 + d_t}. \quad (3.4)$$

### 3.4 Ravnotežna aditivna genetička disperzija

Ravnotežne disperzije, i time asimptotska stopa odgovora pod direkcionom selekcijom, mogu se lako dobiti.

U ravnoteži, iz (3.4) sledi:

$$\tilde{d} = -\kappa \tilde{h}^2 \tilde{\sigma}_A^2 = -\kappa \frac{(\sigma_a^2 + \tilde{d})^2}{\sigma_z^2 + \tilde{d}}.$$

Rešavanjem za  $\tilde{d}$  dobija se ravnotežna aditivna genetička disperzija:

$$\tilde{\sigma}_A^2 = \sigma_z^2 \theta, \quad \text{gde} \quad \theta = \frac{2h^2 - 1 + \sqrt{1 + 4h^2(1 - h^2)\kappa}}{2(1 + \kappa)}.$$

Rezultujuća heritabilnost u ravnoteži je:

$$\tilde{h}^2 = \frac{\tilde{\sigma}_A^2}{\tilde{\sigma}_z^2} = \frac{\theta}{1 + \theta - h^2}. \quad (3.5)$$

## 4 Procena $\sigma_A^2$ i $G$

Pošto nisu direktno opažljive, ali su ključni determinanti evolucije kao odgovora na selekciju, empirijska ocena genetičkih disperzija i kovarijacije je važna tema. Oko vremena tog izdanja, *animal model* (Henderson, 1975), sistem zasnovan na mešovitim modelima koji koristi rodoslovno izvedenu matricu srodstva da bi omogućio direktnu inferenciju kvantitativnih genetičkih parametara. Postao je dominantan pristup za procenu parametara kao što su  $\sigma_A^2$  i  $G$  u prirodi i mnogim eksperimentima.

### 4.1 Animal Model

Animal model je linearni model koji se koristi u kvantitativnoj genetici za procenu genetskih i rezidualnih disperzija. Osnovni model je:

$$y = Xb + Za + e$$

gde su:

- $y$  – vektor posmatranih fenotipskih vrednosti,
- $X$  – matrica dizajna za fiksne efekte  $b$ ,
- $Z$  – matrica dizajna za genetske efekte  $a$  (*breeding values*),
- $e$  – vektor reziduala.

Varijanse su definisane kao:

$$\text{Var}(a) = A\sigma_a^2, \quad \text{Var}(e) = I\sigma_e^2, \quad \text{Cov}(a, e) = 0$$

gde je  $A$ , matrica srodstva između životinja, a cilj je procena  $\sigma_a^2$  (genetska disperzija) i  $\sigma_e^2$  (disperzija reziduala).

### REML – Residual Maximum Likelihood

Za procenu disperzija često se koristi **REML**. Log-likelihood funkcija je:

$$L = -\frac{1}{2} \left[ \log \det(V) + \log \det(X'V^{-1}X) + (y - Xb)'V^{-1}(y - Xb) \right]$$

gde je:

$$b = (X'V^{-1}X)^{-1}X'V^{-1}y, \quad V = ZAZ'\sigma_a^2 + I\sigma_e^2$$

U log-likelihood funkciji postoje tri osnovna člana: kvadratna suma reziduala, član koji zavisi od varijacione matrice i član koji zavisi od fiksnih efekata. Kvadratna suma ponderisanih reziduala  $(y - Xb)'V^{-1}(y - Xb)$ , meri koliko dobro model prati posmatrane podatke. Član koji zavisi od varijacione matrice  $V$  ( $\log \det(V)$ ) kontroliše stabilnost procene disperzije i sprečava *overfitting*. Poslednji član zavisi od fiksnih efekata ( $\log \det(X'V^{-1}X)$ ) i on se može smatrati kaznom za procenu fiksnih efekata i doprinosi nepristrasnosti procene disperzije.

## Mixed Model Equations (MME)

Jednačine mešanog modela su:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1} \frac{\sigma_e^2}{\sigma_a^2} \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

Ove jednačine se koriste za predikciju  $a$  (*breeding values*).

## Prediction Error Variance (PEV)

Predikciona greška za  $a$  se definiše kao:

$$PEV = \text{Var}(\hat{a} - a) = C_{22}\sigma_e^2$$

gde je  $C_{22}$  deo inverzne matrice mešanog modela.

## Procena disperzije - iterativni pristup

Prvi izvodi log-likelihood funkcije po  $\sigma_a^2$  i  $\sigma_e^2$  se postavljaju na nulu:

$$\frac{\partial L}{\partial \sigma_q^2} = 0$$

Rešenje se često dobija iterativno:

- **EM algoritam:** koristi sumu kvadrata reziduala i ponderisanu sumu kvadrata prediktovanih vrednosti za ažuriranje  $\sigma_a^2$  i  $\sigma_e^2$ .
- **AI-REML (Average Information REML):** koristi prosečnu informaciju između očekivane i posmatrane druge diferencijale da bi ažurirao varijanse.

Asimptotska varijansa procena  $\sigma_a^2$  i  $\sigma_e^2$  se dobija iz inverzne matrice informacija  $E^{-1}$  ili  $A^{-1}$ .

### 4.1.1 Primer Animal Model-a

Prikazaćemo na primeru intuitivno objašnjenje kako se matrica srodstva koristi u mešovitom modelu da bi se dobila analiza poznata kao *animal model*.

Razmotrimo analizu polubraće po ocu. Ako eksperiment generiše grupe od tri potomka, pri čemu svaka grupa deli oca kao roditelja, a svaki potomak ima različitu majku, svaka individua unutar sib-grupe (*sibling groups*) delila bi četvrtinu svojih gena sa ostalim članovima grupe i nijedan gen identičan poreklom sa jedinkama iz drugih grupa.

Potomci prve tri sib-grupe imali bi matricu aditivne genetičke srodnosti:

$$A = \begin{bmatrix} 1 & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & 1 & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & 1 & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & 1 & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & 1 \end{bmatrix}.$$

Neka su sada sada disperzije  $\sigma_A^2 = 8, \sigma_E^2 = 12$ . Kovarijacija polubraće po ocu tada bi bila  $\frac{1}{4}\sigma_A^2 = 2$ , jer se očekuje da dele četvrtinu gena identičnih poreklom.

Ako je okruženje pod kontrolom, onda je svaka zajednička sličnost između braće dokaz genetičke povezanosti, a ne posledica spoljašnjih faktora.

Model svih fenotipova istovremeno mogao bi se zapisati kao:

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \\ z_5 \\ z_6 \\ z_7 \\ z_8 \\ z_9 \end{bmatrix} \sim N(X\beta, A\sigma_A^2 + I\sigma_E^2).$$

Eksplisitno:

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \\ z_5 \\ z_6 \\ z_7 \\ z_8 \\ z_9 \end{bmatrix} \sim N(X\beta, 8A + 12I),$$

što daje kovarijacionu matricu:

$$\begin{bmatrix} 20 & 2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 20 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 2 & 20 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 20 & 2 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 20 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 & 20 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 20 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 20 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 2 & 20 \end{bmatrix}.$$

Ovo je jedan *animal model*. U praksi, vrednosti disperzija ( $\sigma_A^2$  i  $\sigma_E^2$ ) nisu poznate, pa se procenjuju na osnovu posmatranih kovarijacija i očekivanja datih  $A$  matricom (koja opisuje srodstvo). Prednost ovog modela je što istovremeno koristi informacije iz različitih klasa srodnika, što omogućava preciznije procene.

## 5 Fitness

**Fitness** (prilagođenost) u evolutivnoj biologiji meri sposobnost jedinke da preživi i ostavi potomstvo u odnosu na druge jedinke u populaciji. Meri koliko se posmatrani genotip favorizuje prirodnom selekcijom i koliko efikasno se prenosi dalje. Apsolutni fitnes je ukupan broj potomaka (ili kopija gena) koje jedinka ostavlja, dok je relativni fitnes je odnos apsolutnog fitnesa jedinke prema prosečnom fitnesu populacije (tj. koliko je uspešna u odnosu na druge).

Proučavanjem fitnesa na određenoj populaciji kroz vreme u određenoj okolini se može utvrditi koje osobine/genotipi rade u korist života. On obuhvata sposobnost jedinke da preživi, pronađe partnera, proizvede potomstvo i na kraju prenese svoje gene na sledeću generaciju. Koristi se za predviđanje kako će se genetski sastav populacije menjati tokom vremena pod različitim selepcionim pritiscima i kao takav predstavlja osnovu proučavanja i prognoziranja dešavanja u prirodi.

Demografski doprinos jedinke populaciji može biti posredno oblikovan njenim uticajem na fitness sopstvenih potomaka. Ova činjenica se u praksi obično rešava tako što se fitness meri u fazi kada se potomci smatraju suštinski nezavisnim.

Pošto je jako teško izvesti merenje fitness-a na terenu/u prirodi ona se retko i sprovode, sistemi za merenje fitnesa su posebno razvijeni za laboratorijske uslove. Umesto toga, pažnja se usmerava na određene epizode selekcije/pojedinačne faze životnih ciklusa. Komponente prilagođenosti za svaku epizodu selekcije definišu se kao multiplikativne (umnožavaju se, jer komponente prilagođenosti u različitim epizodama selekcije se tretiraju kao nezavisni faktori koji zajedno određuju ukupni fitnes jedinke).

Primeri fitness-a:

- **Broj potomaka** – koliko je jedinki određeni organizam ostavio u sledećoj generaciji.
- **Stopa preživljavanja** – verovatnoća da jedinka preživi do reproduktivnog perioda.
- **Reproduktivni uspeh** – broj parenja, broj jaja/polena/semena, broj mlađih koji se uspešno izlegu.
- **Relativni doprinos genetskog materijala** – koliko genetskog materijala određena jedinka prenosi dalje u odnosu na druge.
- **Proksi-mere (zamenske mere)** – npr. telesna masa, visina, brzina rasta, jer su povezane sa šansama za reprodukciju.

## 5.1 Epizode selekcije

*Jedinke se često mere u različitim epizodama selekcije (različitim životnim dobima).*

Neka je  $W_j(r)$  prosečna vrednost fitness-a za r-tu jedinku (sa određenom fenotipskom osobinom koju posmatramo) nakon j-te epizode selekcije. **Relativne komponente prilagođenosti** se definišu kao

$$w_j(r) = \frac{W_j(r)}{W_j} \quad (5.1)$$

Na početku studije, učestalost (frekvencija) svake jedinke je  $1/n$  što za prvu (posmatranu) epizodu selekcije daje:

$$W_1 = \frac{1}{n} \sum_{r=1}^n W_1(r)$$

Ova formula kaže da je prosečna vrednost fitness-a u prvoj epizodi selekcije  $W_1$  prosek svih pojedinačnih vrednosti  $W_1(r)$ , jer svaka jedinka ima jednaku početnu frekvenciju  $1/n$ , gde je  $n$  obim posmatrane populacije.

Nakon prve epizode selekcije, nova frekvencija jedinke r, ponderisana prema fitness-u, biće:

$$\frac{w_1(r)}{n}$$

Što implicira da je prosečan fitness u drugoj epizodi selekcije:

$$W_2 = \sum_{r=1}^n W_2(r) \cdot w_1(r) \cdot \left(\frac{1}{n}\right)$$

Ova formula pokazuje kako se prosečna vrednost fitness-a u drugoj epizodi selekcije  $W_2$  računa tako što se pojedinačni fitness  $W_2(r)$  množi sa prethodnom relativnom vrednošću  $w_1(r)$ , tj. sa time koliko je selekcija već „favorizovala“ tu jedinku nakon prve epizode.

Za j-tu epizodu selekcije:

$$W_j = \sum_{r=1}^n W_j(r) \cdot w_{j-1}(r) \cdot w_{j-2}(r) \cdots w_1(r) \cdot \left(\frac{1}{n}\right) \quad (5.2)$$

Neka je  $p_j(r)$  frekvencija jedinke r ponderisana njenom fitnes vrednošću nakon j kru-gova selekcije, onda važi da je:

$$p_0 = \frac{1}{n}$$

$$p_j(r) = w_j(r) \cdot p_{j-1}(r) = \frac{1}{n} \prod_{i=1}^j w_i(r).$$

Odakle:

$$W_j = \sum_{r=1}^n W_j(r) \cdot p_{j-1}(r) \quad (5.3)$$

**Srednja vrednost određene osobine nakon j-te epizode selekcije** računa se kao:

$$\mu_z(j) = \sum_{r=1}^n z(r) \cdot p_j(r)$$

gde je  $z(r)$  vrednost osobine kod jedinke  $r$ . Ova formula kaže da prosečna vrednost osobine u nekoj generaciji zavisi od toga koje jedinke prežive selekciju, koliko su im izražene osobine i koliko svaka od njih doprinosi populaciji.

**Diferencijal usmerene selekcije**  $S_j$  je mera jačine selekcije koja deluje na određenu osobinu u populaciji. Računa se kao razlika između srednjih vrednosti osobine ponderisanih prema fitnesu pre i posle jedne epizode selekcije. Diferencijal selekcije za  $j$ -tu epizodu je:

$$S_j = \mu_z(j) - \mu_z(j-1) \quad (5.4)$$

Diferencijali selekcije su aditivni kroz epizode, tako da je **ukupan diferencijal S** nakon  $k$  epizoda selekcije:

$$\begin{aligned} S &= \mu_z(k) - \mu_z(0) \\ &= (\mu_z(k) - \mu_z(k-1)) + (\mu_z(k-1) - \mu_z(k-2)) + \cdots + (\mu_z(1) - \mu_z(0)) \\ &= S_k + S_{k-1} + \cdots + S_1 = \sum_{i=1}^k S_i \end{aligned}$$

## 6 Robertson–Price identitet i Teoreme selekcije

Neka je  $z_i$  vrednost osobine jedinke  $i$ ,  $w_i$  njena absolutna prilagođenost (fitness) i  $\bar{w}$  prosečan fitnes populacije koji se računa:

$$\bar{w} = \frac{1}{n} \sum_{i=1}^n w_i \quad (6.1)$$

Neka je  $z'_i$  srednja vrednost osobine potomaka od roditelja  $i$ . Onda je prosečna vrednost osobine u sledećoj generaciji

$$\bar{z}' = \frac{1}{\sum_{i=1}^n w_i} \sum_{i=1}^n w_i z'_i = \frac{1}{\bar{w}} \mathbb{E}[w z'] \quad (6.2)$$

**Tvrđenje (Price):**

$$\Delta \bar{z} \equiv \bar{z}' - \bar{z} = \frac{\text{Cov}(w, z)}{\bar{w}} + \frac{\mathbb{E}[w \Delta z]}{\bar{w}}, \quad \text{gde je } \Delta z_i := z'_i - z_i.$$

**Dokaz:**

$$\begin{aligned} \Delta \bar{z} &= \frac{1}{\bar{w}} \mathbb{E}[w z'] - \mathbb{E}[z] \\ &= \frac{1}{\bar{w}} \mathbb{E}[w(z + \Delta z)] - \mathbb{E}[z] \\ &= \frac{1}{\bar{w}} (\mathbb{E}[wz] - \bar{w} \mathbb{E}[z]) + \frac{1}{\bar{w}} \mathbb{E}[w \Delta z] \\ &= \frac{\text{Cov}(w, z)}{\bar{w}} + \frac{\mathbb{E}[w \Delta z]}{\bar{w}} \end{aligned}$$

Sada, razložimo fenotip na aditivnu genetsku (breeding) vrednost  $A_i$  i ostatak  $E$ :  
 $z_i = A_i + E_i$  uz standardne pretpostavke kvantitativne genetike:

$$\mathbb{E}[E | A] = 0 \quad (\text{okruženje deluje nasumično})$$

$$\text{Cov}(A, E) = 0 \quad (\text{genetski i nenasledni deo su nekorelirani})$$

$$\mathbb{E}[\Delta A] = 0 \quad (\text{nema sistematskog pomaka u naslednom delu osobine})$$

$\Delta A := A' - A$ , tj. promena aditivne genetske vrednosti pri prenosu sa roditelja na potomka.

**Tvrdjenje (Robertson–Price / sekundarna teorema):**

$$\Delta \bar{A} = \text{Cov}(\tilde{w}, A) = \text{Cov}_A(z, \tilde{w}),$$

tj. promena srednje aditivne genetske vrednosti jednaka je aditivnoj genetskoj kovarijaciji između osobine i relativnog fitnesa.

**Dokaz.** Primenjujemo Price-ovu jednačinu na  $A$ :

$$\Delta \bar{A} = \text{Cov}(\tilde{w}, A) + \mathbb{E}[\tilde{w} \Delta A].$$

Ovde smo koristili relativni fitnes, pa nemamo deljenje sa  $\bar{w}$ . Po pretpostavci transmisije (standardna u kvantitativnoj genetici i znači da nema sistematskog (prosečnog) pomaka u aditivnim genetskim vrednostima),  $\mathbb{E}[\tilde{w} \Delta A] = 0$ , pa sledi

$$\Delta \bar{A} = \text{Cov}(\tilde{w}, A).$$

Kako je  $\text{Cov}_A(z, \tilde{w}) := \text{Cov}(A, \tilde{w})$  (jer  $z = A + E$  i  $\text{Cov}(E, \tilde{w}) = 0$ ), dobijamo i drugu jednakost.

**Tvrdjenje (Fisherova fundamentalna teorema prirodne selekcije):**

Stopa porasta prosečne prilagođenosti (fitness-a) populacije u datom trenutku jednaka je aditivnoj genetskoj varijansi u fitnesu u tom trenutku:

$$\frac{d\bar{w}}{dt} = \frac{V_A(w)}{\bar{w}}.$$

gde je  $V_A(w)$  aditivna genetska varijansa u fitnesu, a  $\bar{w}$  prosečan fitnes populacije.

**Skica dokaza.**

Primenjujemo Price-ovu jednačinu na osobinu  $z = w$ :

$$\Delta \bar{w} = \frac{\text{Cov}(w, w)}{\bar{w}} + \frac{\mathbb{E}[w \Delta w]}{\bar{w}}.$$

Prvi sabirak daje

$$\frac{\text{Cov}(w, w)}{\bar{w}} = \frac{\text{Var}(w)}{\bar{w}}.$$

Ako izdvojimo aditivnu genetsku komponentu fitnesa, tj. naslediv deo varijanse, dobijamo

$$\Delta \bar{w}_{\text{genetski}} = \frac{V_A(w)}{\bar{w}}.$$

U izvornoj formulaciji Fisher prepostavlja da nema promena okoline, mutacija, frekvencijske zavisnosti ili drugih transmisijskih efekata, pa drugi sabirak nestaje. Ostaje:

$$\Delta \bar{w} = \frac{V_A(w)}{\bar{w}}.$$

Fisherova teorema važi samo pod restriktivnim uslovima: konstantna sredina, bez promena u genetskim efektima, bez interakcija gena i okruženja. Zato se danas FTPS često posmatra kao “idealizovan” slučaj, a ne univerzalni zakon.

## 7 Mogućnost selekcije

Kako se poredi količina selekcije koja deluje u različitim epizodama ili populacijama? **Standardizovani selekcioni diferencijal (ili intenzitet selekcije):**

$$i = \frac{S}{\sigma_z} \quad (7.1)$$

što je samo pravolinijski selekcioni diferencijal podeljen standardnom devijacijom osobine z u populaciji i on omogućava poređenje jačine selekcije između različitih osobina i populacija, bez obzira na merne jedinice. Dobro je za poređenje selekcije na istu osobinu u različitim uslovima, ali nije prikladan za poređenje ukupne selekcije kada istovremeno deluje selekcija na više osobina. U prirodi, selekcioni pritisci često nisu usmereni samo na jednu osobinu već na kombinaciju osobina (osobine mogu biti genetski povezane i selekcija na jednu osobinu može indirektno uticati i na druge).

U tom smislu, mnogo „čistija“ mera, nezavisna od osobina koje su pod selekcijom, jeste:

$$I = \sigma_w^2 = \frac{\sigma_W^2}{\bar{W}^2} \quad (7.2)$$

gde je  **$I$  prilika za selekciju**, tj. disperzija relativnog fitnesa. Ovde je  $\sigma_W^2$  disperzija apsolutnog fitnesa, a  $\bar{W}$  prosečan apsolutni fitnes. Ovu meru je uveo Crow (1958.), koji ju je nazvao **indeks ukupne selekcije**. Crow je pokazao da, ukoliko je fitness savršeno nasledan (heredabilnost fitness-a:  $h^2 = 1$ ), tada  $I$  direktno opisuje potencijalnu snagu selekcije.

$$I = \Delta w \quad (7.3)$$

Što znači da  $I$  meri maksimalnu moguću promenu prosečnog fitnesa unutar generacije usled selekcije. Veća vrednost  $I$  znači da postoji veća razlika u fitness-u među jedinkama — dakle, ima više potencijala za selekciju, jer jedinke sa boljim genetskim osobinama mogu imati znatno veći broj potomaka. Mala vrednost  $I$  znači da su jedinke manje različite po fitnesu i da selekcioni potencijal može biti manji.

Jedna važna osobina  $I$  jeste da ograničava maksimalni intenzitet selekcije i za bilo koju osobinu. Na osnovu Robertson–Price identiteta:

$$S = \bar{z}^* - \bar{z} = \frac{\text{Cov}(w, z)}{\bar{w}} = \text{Cov}(\tilde{w}, z),$$

gde je  $\bar{z}^*$  prosečna vrednost osobine posle selekcije, a  $\tilde{w} = \frac{w}{\bar{w}}$  relativni fitnes. Ako je  $\rho_{z,w}$  koeficijent korelacije:

$$|\rho_{z,w}| = \frac{|\text{Cov}(\tilde{w}, z)|}{\sigma_z \sigma_w} = \frac{|S|}{\sigma_z \sqrt{I}} = \frac{|i|}{\sqrt{I}} \leq 1$$

što implicira:

$$|i| \leq \sqrt{I}$$

Dakle, najviše što srednja vrednost bilo koje osobine može da se pomeri tokom jedne generacije jeste  $\sqrt{I}$ . Dakle, I postavlja teorijsku gornju granicu jačine selekcije na bilo koju osobinu.

## 8 Koeficijenti selekcije

Kvantitativno-genetički pristup omogućava da se veza između osobina i fitnesa izrazi preciznim matematičkim merama, čime se dobijaju jasne i uporedive predikcije o delovanju selekcije. Selekcija se otkriva poređenjem raspodele osobine pre i posle nje, uz uvažavanje uticaja razvoja, migracije, okruženja i koreliranih osobina.

### 8.1 Mere selekcije srednje vrednosti

Dve mere promene fenotipske srednje vrednosti unutar jedne generacije su već predstavljene: diferencijal pravca selekcije  $S$  i intenzitet selekcije  $i$ . Treća mera je **gradijent pravca selekcije**:

$$\beta = \frac{S}{\sigma_z^2} \quad (8.1)$$

$\beta$  predstavlja nagib linearne regresije fitnesa na fenotip. On meri koliko se fitnes menja kada se osobina promeni za jednu standardnu devijaciju, odnosno meri direktni uticaj osobine na fitnes, kada smo kontrolisali (matematički izolovali) uticaje drugih povezanih osobina.

Kada posmatramo selekciju na jednu osobinu, tri mere selekcije  $S$ ,  $i$  i  $\beta$  daju slične (zamenljive) kvalitativne informacije i čak kvantitativno iste rezultate ako su osobine standardizovane. Međutim, kada imamo više osobina (multivarijantni slučaj), te mere se ponašaju drugačije:  $\beta$  je najbolji za razumevanje direktne selekcije na konkretnu osobinu, jer „izdvaja“ njen uticaj, dok  $S$  i  $i$  uključuju i direktne efekte selekcije i indirektne efekte koji dolaze od povezanosti (korelacije) sa drugim osobinama.

Drugi način izražavanja odgovora na selekciju ( $R$ ) – koliko se osobina promeni u nadrednoj generaciji:

$$R = h^2 \cdot \sigma_z \cdot i = \sigma_A \cdot h \cdot i \quad (8.3)$$

gde je:

- $i$  – intenzitet selekcije
- $\sigma_z$  – standardna devijacija osobine
- $h$  – koren heritabilnosti

Alternativno, koristi se i **gradijent selekcije**  $\beta$ , pa se onda odgovor piše kao:

$$R = \sigma_A^2 \cdot \beta \quad (8.4)$$

što znači da je promena u osobini jednaka proizvodu aditivne genetske disperzije i direktnog uticaja selekcije na osobinu.

## 8.2 Mere selekcije varijacije

Slične mere mogu se definisati za kvantifikovanje promene fenotipske varijanse. Na prvi pogled, ova promena izgleda da se najbolje opisuje kao

$$\sigma_z^{*2} - \sigma_z^2$$

gde je  $\sigma_z^{*2}$  fenotipska disperzija nakon selekcije. Međutim, pravolinijska selekcija — selekcija koja vuče srednju vrednost osobine npr. ka većim vrednostima — smanjuje disperziju, i to ne zato što selekcija deluje na širinu raspodele, već zato što pomera prosek. To može dovesti do zabune: izgleda kao da je disperzija opala jer je osobina postala „ujednačenija“, ali to je delimično samo matematički efekat pomeranja proseka.

Lande i Arnold (1983.) su pokazali da:

$$\sigma_z^{*2} - \sigma_z^2 = Cov(w, (z - \mu_z)^2) - S^2 \quad (8.5)$$

Gde je:  $Cov(w, (z - \mu_z)^2)$  kovarijacija između fitnesa i kvadrata odstupanja od prosečne vrednosti osobine (stvarni efekat selekcije na širinu raspodele), a  $S^2$  kvadrat diferencijala selekcije (deo koji „umetno“ smanjuje varijansu zbog pomeranja proseka). To implicira da pravolinijska selekcija smanjuje fenotipsku varijansu za  $S^2$ . Imajući ovo u vidu, Lande i Arnold predlažu korigovanu meru, **diferencijal stabilizujuće selekcije**:

$$C = \sigma_z^{*2} - \sigma_z^2 + S^2 \quad (8.6)$$

koja opisuje selekciju koja direktno deluje na varijansu.

Analogno tome što je  $S$  jednako kovarijaciji između  $z$  i relativnog fitness-a, poslednja jednačina implicira da je  $C$  kovarijacija između relativnog fitnesa i kvadrata odstupanja osobine od njenog srednjeg vrednosti:

$$C = Cov(w, (z - \mu)^2) \quad (8.7)$$

Dakle,  $C$  govori koliko selekcija menja raznolikost (širinu) osobine u populaciji, a  $S$  govori koliko menja prosečnu vrednost.

- Ako je  $C > 0$ : jedinke koje su dalje od proseka imaju veći fitnes → selekcija povećava varijansu (npr. favorizuje ekstremne vrednosti)
- Ako je  $C < 0$ : jedinke bliže proseku imaju veći fitnes → selekcija smanjuje varijansu (stabilizujuća selekcija).
- Ako je  $C = 0$ : selekcija ne utiče na raspon osobine → nema promene u varijansi.

Kvadratni analog gradijenta selekcije  $\beta$ , tj. **kvadratni (ili stabilizujući) gradijent selekcije  $\gamma$** :

$$\gamma = \frac{Cov(w, (z - \mu)^2)}{\sigma_z^4} = \frac{C}{\sigma_z^4} \quad (8.8)$$

Gradijenti selekcije  $\beta$  i  $\gamma$  omogućavaju da se pri proučavanju više osobina istovremeno razdvoje pravi efekti selekcije od uticaja fenotipskih korelacija, čime daju precizniju sliku o delovanju selekcije i predstavljaju najčešće korišćenu meru u kvantitativnoj genetici.

## 9 Fitness funkcije i karakterizacija selekcije

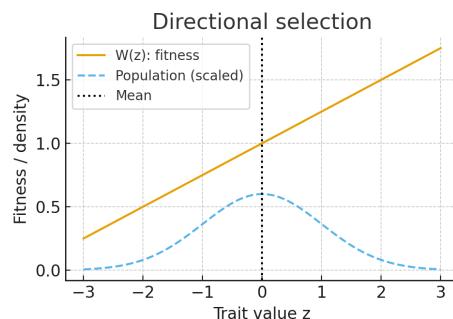
U ovom odeljku dalje istražujemo mehanizme koji povezuju fitness funkcije, odnosno odnose između individualnih vrednosti osobina i očekivanog fitness-a jedinke, sa koeficijentima selekcije na nivou populacije koji su razmatrani u ranije. Ovi mehanizmi će na kraju predstavljati osnovu za empirijsku procenu selekcije, o čemu će biti reči kasnije.

### 9.1 Individualne i prosečne fitness funkcije

$W(z)$  je funkcija koja svakoj vrednosti osobine  $z$  dodeljuje očekivani fitness (na taj način određuje fitness funkciju). Relativna fitness površ  $w(z) = \frac{W(z)}{W}$  je samo još dodatno podeljena i sa prosečnim fitness-om populacije i često je pogodnija za rad.

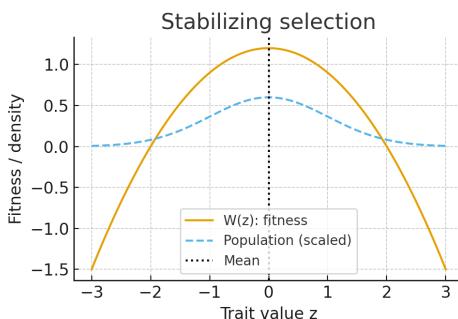
Da bismo razumeli koji oblik selekcije deluje na posmatranu osobinu, potrebno je da sagledamo oblik funkcije fitnessa  $W(z)$  u odnosu na srednju vrednost fenotipa u populaciji ( $\mu$ ). Populacija ima raspodelu vrednosti osobine  $z$ , a srednja vrednost  $\mu$  predstavlja njen centar. Upravo nagib i zakrivljenost funkcije  $W(z)$  u tački  $\mu$  određuju tip selekcije. Objasnimo to na primeru:

- **Direkciona selekcija**



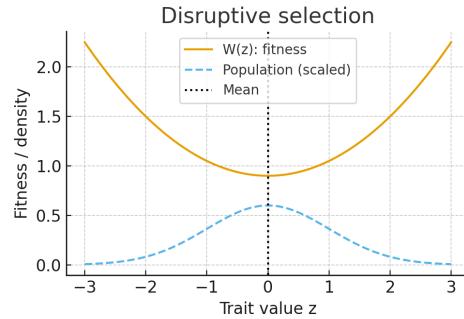
Kada je nagib funkcije  $W(z)$  pozitivan (ili negativan) u okolini srednje vrednosti fenotipa  $\mu$ , jedinke sa većim (ili manjim) vrednostima osobine imaju veći fitness. Populacija se zato pomera naviše ili naniže, što dovodi do promene prosečne vrednosti osobine.

- **Stabilizujuća selekcija**



Ako funkcija  $W(z)$  ima lokalni maksimum kod prosečne vrednosti  $\mu$ , jedinke koje su blizu prosečne vrednosti imaju najveći fitness, dok ekstremne vrednosti imaju manji fitness. Posledica je smanjenje disperzije u populaciji, jer se favorizuju srednje vrednosti.

- **Disruptivna selekcija**



Ako funkcija  $W(z)$  ima lokalni minimum kod  $\mu$ , jedinke sa ekstremnim vrednostima osobine imaju veći fitness od onih sa prosečnim vrednostima. Ovo dovodi do povećanja disperzije i populacija se „razdvaja“ ka ekstremima.

$W(z)$  može da varira u zavisnosti od genotipske i ekološke pozadine. U nekim situacijama očekivani fitness jedinke sa datom vrednošću osobine zavisi od raspodele osobina u populaciji. U tom slučaju, kaže se da je fitness zavisan od učestalosti. Npr. ako većina miševa na polju ima braon krvno, sive miševe je lakše uočiti dakle imaju niži fitness.

Prosečan fitness populacije  $W$  je takođe funkcija, ili površ, koja opisuje očekivani fitness populacije kao funkciju raspodele  $p(z)$  fenotipova u toj populaciji:

$$\bar{W} = \int W(z) p(z) dz. \quad (9.1)$$

Prosečan fitness može se posmatrati kao funkcija parametara fenotipske raspodele. Na primer, ako je  $z$  normalno raspodeljen, prosečan fitness je funkcija srednje vrednosti  $\mu_z$  i disperzije  $\sigma_z^2$  za tu populaciju. Funkcije koje opisuju prosečan fitness u zavisnosti od svojstava populacije često se nazivaju *fitness pejzaži*.

Fitness pejzaži prosečnog fitnessa obično su *glatkiji* od individualnih fitness funkcija. Drugim rečima, raspodela fenotipa u populaciji „izglađuje“ neravnine koje postoje na individualnom nivou. Ovo znači da oblik raspodele u populaciji može značajno da utiče na evolutivni proces. Selekcija uvek favorizuje promene u pravcu povećanja prosečnog fitnessa populacije.

Na primer, razmotrimo dva fitness pejzaža dobijena za istu individualnu fitness funkciju. Ako je disperzija fenotipa veća ( $\sigma_z^2 = 2$ ), prosečan fitness se menja monotono u zavisnosti od  $\mu_z$ . Ako je disperzija mala ( $\sigma_z^2 = 0.1$ ), tada prosečan fitness pejzaž ima lokalni maksimum, što implicira stabilizujuću selekciju (favorizovanje srednjih vrednosti).

Iako se u individualnoj fitness funkciji može pojaviti maksimum ili minimum, to ne znači nužno da se u populaciji javlja stabilizujuća ili disruptivna selekcija - ishod zavisi i od disperzije fenotipa.

## 9.2 Gradijenti i lokalna geometrija fitness površi

Podsetimo se da smo ranije uveli koeficijente  $\beta$ ,  $\gamma$  i  $C$  koji kvantifikuju pravac i oblik selekcije ( $\beta$  meri nagib, a  $\gamma$  meri zakrivljenost). Važno je naglasiti da vrednosti  $\gamma$  i  $C$  ne treba direktno poistovećivati sa stabilizujućom ili disruptivnom selekcijom, jer promene disperzije mogu biti posledica i jake direkcione selekcije. Zbog toga se  $C$  često preciznije naziva *kvadratni selepcioni diferencijal*, pri čemu važi da  $C < 0$  obično ukazuje na smanjenje varijanse (stabilizacija), a  $C > 0$  na povećanje varijanse (disrupcija).

Kada  $z$  ima normalnu raspodelu i kada fitness jedinki ne zavisi od učestalosti,  $\beta$  se može izraziti u terminima geometrije pejzaža prosečnog fitnessa:

$$\beta = \frac{\partial \ln \bar{W}}{\partial \mu_z} = \frac{1}{\bar{W}} \frac{\partial \bar{W}}{\partial \mu_z}, \quad (9.1)$$

pa je  $\beta$  proporcionalan nagibu površi  $W$  u odnosu na srednju vrednost populacije.

$\beta$  se takođe može izraziti kao funkcija individualne fitness površi (takođe pod pretpostavkom da  $z$  ima normalnu raspodelu).

$$\beta = \int \frac{\partial W(z)}{\partial z} p(z) dz, \quad (9.3)$$

što implicira da  $\beta$  predstavlja prosečan nagib individualne fitness površi, gde se prosek uzima preko populacije koja se proučava.

Na sličan način, ako je  $z$  normalno raspodeljen:

$$\gamma = \int \frac{\partial^2 W(z)}{\partial z^2} p(z) dz, \quad (9.4)$$

što predstavlja prosečnu zakrivljenost individualne fitness površi.

Dakle,  $\beta$  i  $\gamma$  daju meru geometrije individualne fitness površi, prosečene preko populacije koja se razmatra.

Dodatna prednost  $\beta$  i  $\gamma$  jeste to što se oni pojavljuju kao mere fenotipske selekcije u jednačinama koje opisuju evolutivni odgovor na selekciju. Podsetimo se (jednačina (8.4)) da, pod prepostavkama koje vode ka Breeder-ovoj jednačini, važi:

$$R = \Delta \mu = \sigma_A^2 \beta.$$

Sledeća jednačina nam govori da na genetičku disperziju utiču zakrivljenost i kvadrat direkcije

$$\Delta \sigma_A^2 = \frac{\sigma_A^4}{2} (\gamma - \beta^2), \quad (9.5)$$

Ovo predstavlja doprinos selekcije komponenti neravnoteže aditivne genetičke disperzije ( $d_t$ , jednačina (3.1)). Zapisano u terminima gradijenata selekcije, Bulmer-ova iteracija (jednačina (3.2)) glasi:

$$d_{t+1} = \frac{d_t}{2} + \frac{\sigma_A^4}{2} (\gamma - \beta^2). \quad (9.6)$$

## 10 Multivariantna selekcija

Nakon što smo istražili teoriju o odgovoru na selekciju za samo jednu osobinu, sada prelazimo na opis posledica selekcije koja deluje na više osobina istovremeno. Razmatramo slučaj selekcije koja deluje na jednu osobinu i kako se evolutivni odgovori mogu javiti i kod drugih osobina koje su genetički korelisane.

### 10.1 Kratkoročne promene u srednjim vrednostima: Multivariantna Brederova jednačina

Prijetimo se multivariantne Brederove jednačine (2.10):

$$R = GP^{-1}S.$$

Ako posmatramo jednu osobinu  $z$  imamo:

$$\beta = \frac{S}{\sigma_z^2}, \quad (10.1)$$

i tada jednačina (2.10) ima sledeći oblik:

$$R = GP^{-1}S = G\beta, \quad (10.2)$$

gde je  $G$  matrica aditivnih genetičkih kovarijacija (poznato od ranije).

Ako bi posmatrali samo sirove promene u populaciji  $S$  može izgledati kao da sve osobine zajedno odgovaraju na selekciju, ali to može biti samo iluzija korelacije (osobina A može izgledati važna za fitness samo zato što je povezana sa osobinom B, koja je zapravo ta koja direktno utiče). Zato nam  $\beta$  definisano na pomenuti način rešava problem jer predstavlja parcijalne efekte.

Dakle multivariantna selekcija funkcioniše identično kao višestruka regresija:

- $S$  su sirovi koeficijenti (kao u prostim regresijama),
- $\beta$  su parcijalni koeficijenti (kao u višestrukoj regresiji),
- $R = G\beta$  predstavlja predikciju stvarnog evolutivnog odgovora.

### 10.2 Efekti genetičkih korelacija: direktni i korelisani odgovori

Za pojedinačnu osobinu  $i$ , imamo:

$$R_i = g_{ii}\beta_i + \sum_{j \neq i} g_{ij}\beta_j. \quad (10.3)$$

Ovo možemo protumačiti kao:

- **Direktan odgovor** ( $g_{ii}\beta_i$ ): promena u osobini  $i$  koja nastaje jer osobina  $i$  sama direktno utiče na fitness.
- **Korelisan odgovor** ( $\sum_{j \neq i} g_{ij}\beta_j$ ): promena u osobini  $i$  koja nastaje isključivo zato što je genetički povezana sa drugim osobinama koje su pod selekcijom.

Ovo je potpuno analogno višestrukoj linearnej regresiji:

$$y = X\beta + \varepsilon,$$

gde su  $X$  prediktori,  $\beta$  parcijalni regresioni koeficijenti, a  $y$  zavisna promenljiva. Kada su prediktori međusobno korelisani, parcijalni efekti ( $\beta$ ) omogućavaju da razdvojimo:

- **uzročni (direktni) efekat** pojedinačne osobine na fitness,
- **indirektne (korelisane) efekte** koji potiču od povezanosti osobina.

U multivariantnoj selekciji:

- $S$  (seleksioni diferencijal) odgovara *naivnoj regresiji* koja ne kontroliše korelacije,
- $\beta = P^{-1}S$  odgovara *višestrukoj regresiji*, gde  $P$  uklanja fenotipske korelacije i daje „ispravljene“ direktne efekte,
- $G$  funkcioniše kao „težinski faktor“ koji prevodi efekte sa nivoa fenotipa na nivo genetičkih odgovora.

## Primer

Zamislimo ptice kod kojih posmatramo dve osobine:

- $z_1$  = dužina krila,
- $z_2$  = dužina repa.

Selekcija direktno favorizuje duža krila (pozitivno  $\beta_1$ ), dok rep nije direktno pod selekcijom ( $\beta_2 = 0$ ). Ako je  $g_{12} \neq 0$  (npr.  $g_{12} = -0.8$ ), promena u krilima povlači i promenu u repu, jer su osobine genetički povezane. Formalno:

$$R_1 = g_{11}\beta_1, \quad R_2 = g_{21}\beta_1.$$

Dakle:

- **krilo** se menja direktno jer utiče na fitness,
- **rep** se menja korelisano, iako nije bio pod direktnom selekcijom.

Ovaj formalizam pokazuje zašto je multivariantna analiza (tj. višestruka regresija u linearnim modelima) nužna: bez nje bismo pogrešno zaključili da rep direktno utiče na fitness, iako je njegov efekat posledica korelacije sa krilima. Dakle, upotreba  $\beta$  uklanja fenotipske korelacije, dok  $G$  pokazuje kako genetičke korelacije generišu korelisan odgovor na selekciju.

### 10.3 "Path" analiza i prošireni gradijenti selekcije

Pored gradijenata selekcije  $\beta$ , u literaturi se razvila i *path* analiza, koja omogućava da se računaju i indirektni efekti osobina na fitness. Ukupni efekti dobijeni na ovaj način često se označavaju sa  $\eta$  i nazivaju *prošireni gradijenti selekcije*.

Formalno, ako  $\Phi$  opisuje međusobne odnose osobina (putem path koeficijenata), tada važi:

$$\eta = \Phi^T \beta, \quad (10.4)$$

gde  $\eta$  uključuje i direktne i indirektne efekte osobina na fitness.

Iako path analiza koristi sličan formalizam kao linearne modeli (oslanja se na sistem regresionih jednačina), njen cilj je više ekološka interpretacija odnosa između osobina i fitnessa, a ne predikcija evolutivnog odgovora. Zbog toga se u ovom radu fokusiramo na standardne gradijente selekcije  $\beta$  i Brederove jednačine.

## 11 Procena gradijenata selekcije

Procena oblika i jačine prirodne selekcije predstavlja jedan od najvažnijih empirijskih zadataka u evolutivnoj kvantitativnoj genetici. Gradijenti selekcije se mogu **poistovetiti sa parametrima linearног regresionог modelа**, pa se njihova procena svodi na metode koje već poznajemo iz teorije linearnih statističkih modela – pre svega na metodu *običnih najmanjih kvadrata* (MNK).

### 11.1 Analiza metodom MNK

Kada su fenotipi višedimenzionalno normalno raspodeljeni, gradijent selekcije  $\beta$  se procenjuje kao koeficijent regresije relativnog fitnessa na fenotip. Za jednu osobinu ( $z$  – fenotipska vrednost osobine za pojedinca) dobijamo model:

$$w_i = 1 + \beta(z_i - \bar{z}) + e_i, \quad (11.1)$$

gde je  $w_i = W_i/\bar{W}$  relativni fitness,  $z_i$  vrednost osobine kod pojedinca  $i$ , a  $e_i$  rezidual. Dakle,  $\beta$  je regresioni koeficijent dobijen metodom OLS:

$$\hat{\beta} = \arg \min_{\beta} \sum_i (w_i - \hat{w}_i)^2.$$

Model se lako proširuje na više osobina primenom višestruke regresije:

$$w_i = 1 + \sum_{j=1}^k \beta_j (z_{ij} - \bar{z}_j) + e_i, \quad (11.2)$$

gde  $z_{ij}$  označava vrednost  $j$ -te osobine kod  $i$ -tog pojedinca. Vektor  $\beta$  tada sadrži *parcijalne regresione koeficijente*, koji mere efekat svake osobine na fitness kada se ostale osobine drže konstantnim – potpuno analogno interpretaciji parametara u višestrukoj regresiji.

## Kvadratni gradijenti selekcije

Da bismo opisali zakriviljenost funkcije fitnessa, u model uključujemo i kvadratne i interakcione članove:

$$w_i = \alpha + \sum_{j=1}^k \beta_j (z_{ij} - \bar{z}_j) + \sum_{j=1}^k \frac{1}{2} \gamma_j (z_{ij} - \bar{z}_j)^2 + \sum_{j=1}^k \sum_{l=j+1}^k \gamma_{jl} (z_{ij} - \bar{z}_j)(z_{il} - \bar{z}_l) + e_i. \quad (11.3)$$

Ovde su  $\gamma_j$  i  $\gamma_{jl}$  **kvadratni gradijenti selekcije**, odnosno koeficijenti kvadratnih i interakcionih članova u regresiji. Oni kvantifikuju da li selekcija favorizuje *srednje vrednosti* (stabilizujuća selekcija,  $\gamma_j < 0$ ) ili *ekstreme* (disruptivna selekcija,  $\gamma_j > 0$ ).

Kao i u standardnim linearnim modelima, interpretacija parametara zavisi od toga kako je model postavljen:  $\beta$  meren parcijalno (analog višestrukoj regresiji), dok  $\gamma$  pokazuje zakriviljenost.

### Primer

Zamislimo ptice kod kojih merimo dužinu krila ( $z$ ) i računamo njihov relativni fitness ( $w$ ). Model (11.1) daje:

$$w_i = 1 + \beta(z_i - \bar{z}) + e_i.$$

Ako procenimo  $\hat{\beta} = 0.3$ , to znači da jedinke sa dužim krilima u proseku imaju veći fitness. Ako dodamo kvadratni član i dobijemo  $\hat{\gamma} < 0$ , to bi značilo da su ptice sa *srednjim* dužinama krila u najpovoljnijoj poziciji (stabilizujuća selekcija).

Na ovaj način vidimo da su gradijenti selekcije  $\beta$  i  $\gamma$  zapravo **parametri linearног modelа**, koje procenjujemo standardnim statističkim metodama (MNK), i da se njihova biološka interpretacija naslanja na klasičnu regresionu analizu.

## 11.2 Fleksibilna procena fitness funkcija sa pridruženim gradijentima selekcije

Klasična MNK analiza daje linearnu aproksimaciju odnosa između fenotipa i relativnog fitnessa. Međutim, ako je funkcija fitnessa zakriviljena ili složenog oblika, ovakav pristup može biti ograničen. Zbog toga se koriste fleksibilniji statistički modeli, kao što su splajn funkcije i generalizovani linearni modeli (GLM).

**Splajnovi** omogućavaju da se oblik funkcije  $\hat{W}(z)$  proceni bez nametanja unapred definisanog oblika. Na taj način se model „prilagođava“ podacima, a gradijenti selekcije se zatim računaju kao prosečni parcijalni izvodi te procenjene funkcije. Formalno, za uzorak od  $n$  jedinki:

$$\hat{\beta}_j = \frac{1}{n} \sum_{i=1}^n \frac{\partial \hat{W}(z_i)}{\partial z_j} \left( \frac{1}{n} \sum_{i=1}^n \hat{W}(z_i) \right)^{-1}, \quad (11.4)$$

$$\hat{\gamma}_{jl} = \frac{1}{n} \sum_{i=1}^n \frac{\partial^2 \hat{W}(z_i)}{\partial z_j^2} \left( \frac{1}{n} \sum_{i=1}^n \hat{W}(z_i) \right)^{-1}. \quad (11.5)$$

Ovde  $\hat{W}(z)$  može poticati iz različitih statističkih modela:

- OLS model (linearni odnos fenotipa i fitnessa),
- logistička regresija (kada je fitness izražen kao preživljavanje, Janzen i Stern, 1998),
- splajn regresija ili GLM (Morrissey i Sakrejda, 2013).

Na ovaj način,  $\beta$  i  $\gamma$  i dalje imaju interpretaciju regresionih parametara (kao kod linearnih modela), ali sada izvedenih iz opštije procenjene funkcije.

**Zaključak:** ovaj pristup zadržava vezu sa linearnim modelima (jer su gradijenti parcijalni regresioni koeficijenti), ali ih širi ka fleksibilnijim okvirima, što omogućava preciznije hvatanje oblika fitness funkcije i samim tim preciznije razumevanje prirodne selekcije.

## Normalnost i gradijenti selekcije

Analiza gradijenata selekcije nameće neuobičajene prepostavke o normalnosti u regresionim analizama. Ovo pitanje je izvor značajne konfuzije. Zaključujemo ovo poglavlje obrađivanjem ključnih izvora potencijalne konfuzije u ovoj oblasti.

### Normalnost reziduala

Često se smatra da MNK regresija zahteva da reziduali budu normalno raspodeljeni. Međutim, prema Gauss-Markov teoremi, MNK procene su nepristrasne i imaju minimalnu varijansu među linearnim procenama ako su reziduali nezavisni, homoskedastični i imaju očekivanje nula. Normalnost je važna samo za hipotezno testiranje u malim uzorcima, dok se u velikim uzorcima može osloniti na centralnu graničnu teoremu.

### Normalnost fenotipa i genetičkih vrednosti

Kod teorije gradijenata selekcije normalnost fenotipa ima važniju ulogu nego normalnost reziduala. Ako je fenotip snažno asimetričan, otežano je razdvajanje direkcionih i kvadratnih komponenata u kvadratnim regresionim modelima. Još važnije, nenormalnost fenotipa može dovesti do toga da nelinearni aspekti selekcije utiču na srednju vrednost osobine kroz generacije, čime se narušava veza između statističkih parametara i evolutivne genetike.

Jedan od matematičkih rezultata koji povezuje ove ideje je **Stein-ova lema**: ako je slučajni vektor  $x$  multivarijantno normalan, tada važi

$$\Sigma(x, y) = \Sigma(x) \mathbb{E} \left[ \frac{\partial y}{\partial x} \right]. \quad (11.6)$$

Ovo omogućava da se prosečan izvod relativnog fitnessa (tj. gradijent selekcije  $\beta$ ) ispravno interpretira, pod uslovom da su aditivne genetičke vrednosti normalne i nezavisne od negenetičkih efekata.

**Zaključak:** U analizi prirodne selekcije, prepostavke linearnih modela treba pažljivo tumačiti. Za razliku od standardne regresione prakse, gde se akcenat stavlja na normalnost reziduala, ovde je ključna normalnost fenotipa i genetičkih komponenti. To osigurava da regresioni koeficijenti ( $\beta, \gamma$ ) zadrže svoje teorijsko evolutivno značenje.

## 12 Rad sa bazom podataka

Za potrebe ovog rada koristićemo dataset **GaltonFamilies** koji se nalazi u R paketu **HistData**. Ovaj skup podataka potiče iz klasičnih istraživanja Francisa Galtona o nasleđivanju telesnih karakteristika u porodicama.

Prvo smo odredili teorijski koeficijent heratibilnosti, koji je iznosio oko 0.32 i poklapao se sa ocenjenom vrednošću. Simuliranjem selekcije dobija se da je R približno 1.62. Grafik populacionih proseka nam prikazuje da postoji rastuća veza između visine roditelja i potomka. Kasnije se vidi da to važi nezavisno od pola, implementacijom Brederove jednačine i Animal Model-a za višeosobinsku regresiju.

Aproksimirali smo fitness brojem dece i izračunali promenu prosečne osobine  $\Delta z$  prema Price-ovoj jednačini, dobijajući vrednost  $-0.087$ , što implicira selekcioni pritisak u suprotnom smeru od prosečne visine. Kroz Robertson–Price pristup, predviđeli smo aditivnu genetsku komponentu visine korišćenjem linearne regresije visine deteta na `midparentHeight` i izračunali  $\Delta z_A = -0.056$ , što pokazuje genetski doprinos selekciji. Analiza po polu pokazuje da je selekcioni pritisak značajno jači kod žena ( $\Delta z = -0.152$ ,  $\Delta z_A = -0.097$ ) nego kod muškaraca ( $\Delta z = -0.025$ ,  $\Delta z_A = -0.016$ ). Rezultate vizualizujemo bar plotom koji upoređuje  $\Delta z$  i  $\Delta z_A$  po polu, jasno ilustrujući razliku između ukupne i aditivne komponente selekcije.

Dalje smo uveli relativnu fitness funkciju  $W/\bar{W}$  i posmatrali zavisnost broja dece od fenotipa roditelja (srednje visine). Centriranjem i standardizovanjem osobine izračunali smo selekcione koeficijente  $\beta$  i  $\gamma$ . Dobijena vrednost  $\beta$  bila je veoma mala i statistički beznačajna, što ukazuje da prosečna visina roditelja nema linearan efekat na fitness, dok empirijska procena funkcije  $W(z)$  pokazuje maksimum u blizini srednje vrednosti osobine, što upućuje na postojanje stabilizujuće selekcije. Grafikon linearног modela jasno potvrđuje da pravolinijski efekat ne postoji, ali je model konstruisan radi provere teorijskih zaključaka i konzistentnosti sa biološkom interpretacijom.

## Literatura

- [1] Balding, D.J., Moltke, I., & Marioni, J. (2015). *Handbook of Statistical Genomics* (4th ed.). Chichester: Wiley.
- [2] Fisher, R.A. (1930). *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press.
- [3] Falconer, D.S. (1981). *Introduction to Quantitative Genetics*. Edinburgh: Oliver and Boyd.
- [4] Hartl, D.L., & Jones, E.W. (2007). *Genetics: Principles and Analysis* (4th ed.). Boston: Jones and Bartlett Publishers.
- [5] Mrode, R.A. (2014). *Linear Models for the Prediction of Animal Breeding Values* (2nd ed.). Wallingford: CABI.
- [6] Walsh, B., & Rosa, G. (2017). *Principles of Quantitative Genetics*. Course syllabus, SISG, Seattle, 17–19 July.