UNIVERZITET U NOVOM SADU
PRIRODNO-MATEMATIČKI FAKULTET
DEPARTMAN ZA MATEMATIKU I
INFORMATIKU

Ivana Jančić

MEŠOVITI LINEARNI MODELI I NjIHOVA UPOTREBA U SELEKCIJI DOMAĆIH ŽIVOTINjA

- master rad -

Novi Sad, 2017.

Zahvaljujem se mentoru dr Zagorki Lozanov-Crveković na ukazanom poverenju. Posebnu zahvalnost na saradnji upućujem MSc Ljubi Štrbac kao i profesorima dr Snežani Trivunović i dr Lidiji Perić sa Poljoprivrednog fakulteta, Novi Sad.

Sadržaj

Uvod

U ovom radu bavićemo se linearnim mešovitim modelima koji imaju široku primenu u genetskom poboljšanju populacije, a samim tim i u selekciji životinja. Za razliku od linearnih modela, u kojima se posmatraju samo fiksni efekti (faktori), u ovakvim modelima su, pored njih, uključeni i slučajni efekti (faktori). Uključivanje slučajnih faktora u model je tipično za merenja koja su korelisana. U poljoprivredi, medicini, psihologiji merenja se često vrše na individuama tokom vremena što dovodi do povezanosti između merenja, posebno ako je razmak između istih relativno mali. Na primer, u oplemenjivanju životinja može biti povezan prinos mleka kod mlečnih krava koje potiču od istog oca, te se stoga uticaj oca uključuje u model kao slučajni faktor.

Pre nego što je čovek pripitomio, odnosno počeo da uzgaja životinje radi koristi ili pak razonode postojala je prirodna selekcija u kojoj su samo one najbolje prilagođene jedinke, odnosno populacije opstajale. Danas, selekcija podrazumeva *odabir*, pomoću nekog utvrđenog metoda, najbolje jedinke koja će se koristiti za dalji priplod. Drugim rečima, suština je da se odaberu najbolje životinje koje će svoj genetski potencijal preneti na potomstvo.

Glavni cilj svakog odgajivačkog programa je definisati metode kojima će se postići najveća ekonomska korist od same životinje, pre svega genetskim unapređenjem životinja na nivou populacije. Stoga je ove programe potrebno prilagođavati uslovima i potrebama tržišta. Savremene metode selekcije se nisu oduvek upotrebljavale, već se prvobitna selekcija oslanjala na iskustvo odgajivača. Dakle, biranje životinja za dalji priplod zavisilo je od znanja, veština i dugogodišnjeg iskustva odgajivača.

Treba uzeti u obzir da svaka jedinka nasleđuje određene osobine od svojih roditelja, koje se mogu ispoljiti ili ne. Svaka primetna, uočljiva karakteristika koja se javlja kod individue, a koja je posledica delovanja gena predstavlja *fenotipsku* osobinu. Takva osobina može biti kvalitativna (boja očiju, boja dlake, boja perja,...) ili kvantitativna koja se izražava količinska mera (količina mleka, mesnatost, brzina konja,...). Svaka fenotipska osobina može biti nasleđena, određena uticajem životne sredine ili se pak ispoljiti zbog zajedničkog delovanja oba faktora. Kvantitativna genetika se zasniva na pretpostavci da je fenotip kvantitativne osobine posledica delovanja gena i okoline. Takođe, bitno je napomenuti da na ispoljavanje neke osobine ne utiče jedan gen već grupa gena svaki sa svojim udelom. Jedan od osnovnih pojmova u kvantitativnoj genetici jeste varijansa. Ukupna fenotipska varijansa u populaciji proizilazi iz dva izvora: razlike u genima i uslova spoljašnje sredine. Ove komponente je potrebno proceniti jer ih nije moguće direktno izračunati. Polazi se od pretpostavke da različite vrste srodnika imaju različite stepene genetske povezanosti. Tako je na primer povezanost kod monozigotnih blizanaca 100%, kod dizigotnih, roditelja i dece, braće i sestara 50%. Pored toga na stepen povezanosti utiče i okolina pa se ukupna fenotipska varijansa može podeliti na genetsku i varijansu okoline. Stoga, kada je u pitanju oplemenjivanje životinja osnovni cilj je poboljšanje genetskog kapaciteta životinja, a za to nam umnogome mogu pomoći mešoviti linearni modeli.

U prvom delu rada biće dati neki od osnovnih kako statističkih tako i uopšte matematičkih pojmova koji će biti potrebni za dalju analizu. Potom ćemo definisati linearne mešovite modele i biće data njihova klasifikacija. Centralni deo rada odnosi se na procenu, pre svega, slučajnih efekata. Za njihovu procenu koristi se najbolji linearni nepristrasan prediktor - BLUP, koji podrazumeva minimizaciju srednje kvadratne greške - MSE. Ovaj način predviđanja ima dugu istoriju koju je verovatno započeo Charles Roy Henderson(1948) u svom istraživanju koje je direktno vezano za oplemenjivanje životinja (engl. *Animal breeding*). Pored ove ocene biće naveden i empirijski najbolji linearni nepristrasan prediktor - EBLUP, koji se primenjuje u slučaju kada nam varijanse nisu poznate. Formira se na taj način što se vrši ocenjivanje varijansi, a potom se one tako ocenjene uvrštavaju u BLUP ocenu. Jedna od osnovnih problematika u tom slučaju jeste ocena srednje kvadradne greške. Za ocenu MSE biće data ocena dobijena *jackknife metodom* koju je predložio Jiming Jiang [4]. Na kraju će kroz realne primere biti prikazana primena ovih modela kao i kratak osvrt na njihovu upotrebu u R. Srbiji.

Glava 1

Osnovni pojmovi

Kako bismo izveli zaključke o zakonitostima i međusobnim vezama koji se odnose na neku pojavu, potrebno je izvršiti analizu podataka dobijenih posmatranjem i merenjem na nekom uzorku. Ishod merenja smatramo zavisnom promenljivom dok nezavisne promenljive predstavljaju sve faktore koje utiču na dati ishod.

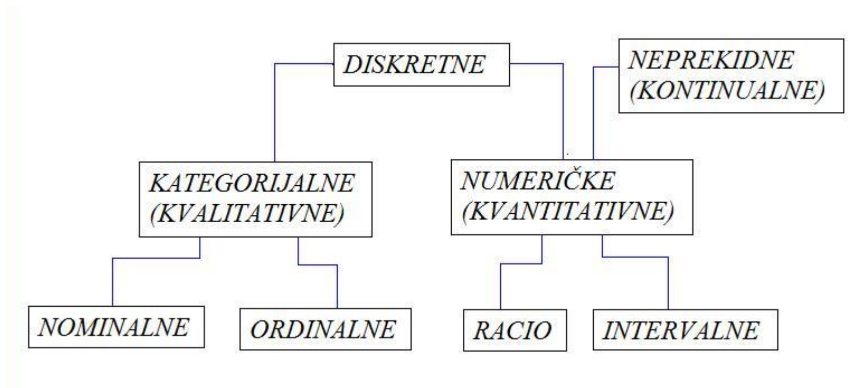
Pre svega navešćemo neke od podela promenljivih jer to utiče na statističke metode koje ćemo kasnije koristiti.

Prva i najčešća klasifikacija promenljivih jeste na osnovu njihovih vrednosti. Na taj način promenljive delimo u dve osnovne grupe, a to su *kategorijalne* i *numeričke* promenljive. Kategorijalne promenljijve sadrže podatke kategorisane u grupe koje mogu biti poređane po veličini. Ispitanici se razdvajaju u jasno razgraničene grupe po određenoj karakteristici. Kategorijalne promenljive se još nazivaju i kvalitativne. S druge strane, numerički ili kvantitativni podaci imaju za vrednosti brojeve. Osnovna razlika između ova dva tipa promenljivih je u tome što kategorijalne promenljive sadrže podatke koji nemaju pravo numeričko značenje pa se na njima ne mogu vršiti matematičke operacije, dok na numeričkim to ima smisla.

Prema vrednostima koje mogu primiti, numeričke promenljive se mogu klasifikovati na *diskretne* i *neprekidne* (kontinualne). Diskretne promenljive primaju vrednosti iz prebrojivog skupa podataka, dok neprekidne mogu primiti vrednosti iz neprebrojivog skupa podataka.

Prema nivou merenja ili skali merenja promenljive možemo podeliti na *nominalne*, *ordinalne*, *intervalne* i *racio* promenljive. Kada su u pitanju nominalne promenljive, podaci se grupišu u deskriptivne kategorije koje se ne mogu poređati po veličini. Podaci se svrstavaju po nekoj osobini ili karakteristici po kojoj se međusobno razlikuju. Dakle, pretpostavlja se da su svi pripadnici jedne kategorije isti po toj osobini zbog koje su svrstani u datu kategoriju. Kako nominalne promenljive grupišu podatke po njihovim karakteristikama, one spadaju u kategorijalni tip promenljivih. Za razliku od nominalnih, kod ordinalnih promenljivih se zna relativni polozaj svakog pojedinačnog slučaja u odnosu na ostale. Na primer, dijastolni krvni pritisak se može grupisati u kategorije: $\leqslant 70$, $71-90$, $91-110$, $110-130$ i $\geqslant 131$ $mmHg$. Kod intervalnih promenljivih jednake razlike u brojevima na intervalnoj skali predstavljaju jednake razlike u posmatranom svojstvu. Razlika između dve vrednosti se može iskazati brojem, ali on ne predstavlja njihovu relativnu razliku. Karakteristično za intervalne varijable je da se položj nule i merne jedninice određuje dogovorom. Sama nula ne znači nepostojanje posmatranog svojstva. S druge strane, kod racio promenljivih razliku između dve promenljive možemo iskazati brojem ali on predstavlja i stvarnu razliku između njih. Pri tome nula na skali merenja ukazuje na nepostojanje određenog svojsta.

Kategorijalne promenljive mogu biti nominalne i ordinalne i one su uvek diskretne, dok numeričke promenljive mogu biti intervalne i racio, a pored toga mogu biti i diskretne i neprekidne. Termin kvantitativan se često koristi za promenljive merene na neprekidnoj skali, dok se termin kvalitativan koristi za nominalne i ponekad ordinalne podatke.



Slika 1.1: Grafički prikaz podele slučajnih promenljivih.

Slučajna promenljiva je jedan od osnovnih pojmova u statistici. Pomoću nje možemo proučavati određene pojave. Skup svih elemenata na kojima se proučava određena pojava nazivamo statistički skup ili populacija, a osobine po kojima se elementi populacije razlikuju obeležja. Slučajna promenljiva predstavlja osnovni matematički model za obeležje. Osnovni zadatak statistike je određivanje raspodele ili funkcije raspodele posmatranog obeležja. Razlog za to je što su sve informacije vezane za slučajnu promenljivu sadržane u njenoj funkciji raspodele. Analiza podataka, ocenjivanje parametara, test statiske zavise od raspodele podataka. Međutim, u nekim slučajevima nam nije poznat ni tip raspodele, pa se njeno određivanje vrši na osnovu registrovanih vrednosti obeležja na svim elementima populacije. S druge strane, u praksi je vrlo često dovoljno ukazati samo na neke bitne parametre koji ilustruju posmatrano obeležje, odnosno bitne karakteristike raspodele slučajne promenljive. Najveći praktični značaj imaju dve grupe parametara. U prvoj grupi su parametri koji reprezentuju centar rasturanjavrednosti slučajne promenljive, a u drugoj grupi parametri koji mere to rasturanje oko centra rasturanja.[9] Dakle, za opisivanje obeležja koriste se neke bitne karakteristike i to najčeče matematičko očekivanje slučajne promenljive, koje spada u prvu grupu navedenih parametara i disperzija i standardna devijacija koja spada u drugu grupu parametara.

Najčešće se dešava da parametri obeležja nisu poznati, pa ih je potrebno oceniti. Pre svega navešćemo neke od bitnih karakteristika slučajne promenljive Y. Sa $E(Y)$ ćemo označavati očekivanu vrednost slučajne promenljive Y (matematičko očekivanje ili prosek populacije). Disperzija ili varijansa populacije $Var(Y)$ je brojna karakteristika koja predstavlja meru odstupanja od srednje vrednosti,

$$Var(Y) = \sigma^2 = E[(Y - E(Y))^2].$$

Kada je u pitanju uzorak od n subjekata imamo da je očekivanje uzorka, u oznaci $\bar{Y}$, ocena proseka populacije:

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i.$$

Kao ocena varijanse populacije koristi se:

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2.$$

Deljenje sa $n - 1$ se koristi kako bi ocena bila nepristrasna.
Kovarijansa predstavlja meru koja pokazuje kolika je povezanost između dve promenljive, gde pozitivna vrednost ukazuje da se velike vrednosti jedne promenljive poklapaju sa velikim vrednostima druge, a negativna da se velike vrednosti jedne poklapaju sa malim vrednostima druge promenljive. Dakle, ako su Y_i i Y_j dve slučajne promenljive onda je kovarijansa između njih

$$Cov(Y_i, Y_j) = E[(Y_i - E(Y_i))(Y_j - E(Y_j))].$$

Očigledno je da kovarijansa između dve iste promenljive predstavlja varijansu.
Neka sada y predstavlja vektor posmatranja:

$$y = \begin{bmatrix} y_1 & y_2 & \dots & y_n \end{bmatrix}',$$

a Y vektor slučajnih promenljivih:

$$Y = \begin{bmatrix} Y_1 & Y_2 & \dots & Y_n \end{bmatrix}'.$$

Dakle, pretpostavimo da imamo n nezavisnih slučajnih promenljivih $Y_1, Y_2, ..., Y_n$, takvih da važi $E(Y_i) = \mu_i$ i $Var(Y_i) = \sigma_i^2$ za $i = 1, 2, ..., n$. Neka je W slučajna promenljiva koja predstavlja njihovu linearnu kobinaciju, tj.

$$W = a_1Y_1 + a_2Y_2 + ... + a_nY_n,$$

gde su $a_i, i = 1, 2, ..., n$ konstante. Tada je očekivana vrednost slučajne promenljive W data sa

$$E(W) = a_1\mu_1 + a_2\mu_2 + ... + a_n\mu_n,$$

a njena varijansa

$$Var(W) = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + ... + a_n^2\sigma_n^2.$$

1.1 Normalna raspodela

Normalna raspodela predstavlja jedan od osnovnih uslova za sprovođenje parametarskih testova u statistici. Stoga su u ovom delu dati neki od osnovnih pojmova normalne raspodele.

Ako slučajna promenljiva Y ima normalnu raspodelu sa sredinom μ i varijansom σ^2, u oznaci

$$Y : N(\mu, \sigma^2),$$

tada je njena funkcija gustine raspodele[9] data sa

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}. \tag{1.1}$$

Normalna raspodela sa sredinom 0 i varijansom 1, $Y : N(0, 1)$, se naziva *standardna normalna raspodela.* Rezultate ponovljenih posmatranja na istoj jedinki potrebno je sagledati zajedno. Zbog toga nam je potreban slučajni vektor Y, naveden ranije. Svaki element vektora Y je slučajna promenljiva koja ima normalnu raspodelu sa očekivanjem μ_i i varijansom σ_i^2:

$$Y_i : N(\mu_i, \sigma_i^2),$$

gde $i = 1, 2, ..., n$.
Kovarijansa između dve slučajne promenljive Y_i i Y_j je data sa

$$Cov(Y_i, Y_j) = \rho_{ij}\sigma_i\sigma_j,$$

gde ρ_{ij} predstavlja koeficijent linearne korelacije između Y_i i Y_j.

Raspodelu jedne promenljive često nazivamo **univarijantna raspodela**, dok u ovom slučaju imamo zajedničku raspodelu za normalno distribuirane slučajne promenljive Y_i, odnosno **mulitivarijantnu normalnu raspodelu**. Kako posmatramo sve elemente Y zajedno, možemo zaključiti da će postojati veza između merenja na istoj jedinki u različitim vremenskim trenucima. Stoga, za slučajni vektor Y, očekivanu vrednost dobijamo kao očekivanje svakog elementa vektora Y i to predstavljamo

$$\mu = \begin{bmatrix} E(Y_1) & E(Y_2) & \dots & E(Y_n) \end{bmatrix}'.$$

Očekivanu vrednost definišemo kao

$$E(Y) = \mu.$$

Varijansno-kovarijansnu matricu V definišemo na sledeći način:

$$\begin{bmatrix} E[(Y_1-\mu_1)^2] & E[(Y_1-\mu_1)(Y_2-\mu_2)] & \cdots & E[(Y_1-\mu_1)(Y_n-\mu_n)] \\ E[(Y_2-\mu_2)(Y_1-\mu_1)] & E[(Y_2-\mu_2)^2] & \cdots & E[(Y_2-\mu_2)(Y_n-\mu_n)] \\ \vdots & \vdots & \ddots & \vdots \\ E[(Y_n-\mu_n)(Y_1-\mu_1)] & E[(Y_n-\mu_n)(Y_2-\mu_2)] & \cdots & E[(Y_n-\mu_n)^2] \end{bmatrix},$$

ili jednostavnije možemo zapisati

$$V = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{bmatrix}.$$

Matricu V nazivamo *kovarijansna* matrica. Kako važi da je $\sigma_{ij} = \sigma_{ji}$, za svako $i = 1, 2, ..., n$ i $j = 1, 2, ..., n$, matrica V je simetrična. Takođe, matrica V je kvadratnog oblika. Zapisujemo

$$Y : N(\mu, V),$$

gde je

$$Y = \begin{bmatrix} Y_1 & Y_2 & \ldots & Y_n \end{bmatrix}'.$$

Analogno se koeficijenti korelacije prikazuju u matričnom obliku, gde ćemo na dijagonali imati jedinice. Pretpostavimo da su $Y_1, Y_2, ..., Y_n$ nezavisne i normalno distribuirane i nezavisne slučajne promenljive $Y_i : N(\mu_i, \sigma_i^2)$ za $i = 1, 2, ..., n$. Tada je slučajna promenljiva W, koja predstavlja linearnu kombinaciju kao i ranije, takođe normalno distribuirana i važi:

$$W = \sum_{i=1}^{n} a_i Y_i : N(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2).$$

1.2 Uopštena inverzna matrica

Posmatramo jednostavan model:

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

gde je X n-dimenzionalni slučajni vektor, A matrica reda $n \times m$, gde je $m > n$, β m-dimenzionalni nepoznati parametar koji treba oceniti i ε n-dimenzionalni vektor slučajnih grešaka.
Kod rešavanja sistema linearnih jednačina neretko se javlja problem da matrica A nije punog ranga, a samim tim nije regularna matrica. To se dešava kada imamo manji broj jednačina od broja nepoznatih, tj. kada želimo da ocenimo m parametara, a imamo n opservacija, pri čemu je $n < m$. U toj situaciji koristimo uopštenu inverznu matricu.

Definicija 1.2.1 *Uopštena inverzna matrica matrice A reda $n \times m$ je bilo koja matrica G reda $m \times n$ za koju važi:*

$$AGA = A.$$

Za uopštenu inverznu matricu G matrice A koristićemo i oznaku A^-.

Teorema 1.2.1 *Neka je A regularna matrica. Tada je njena inverzna matrica A^{-1} i uopštena inverzna matrica i pri tome je jedinstvena.*

Dokaz. Kako je A^{-1} inverzna matrica matrice A važi:

$$AA^{-1}A = IA = A.$$

Neka je sada A^- bilo koja uopštena inverzna matrica matrice A. Tada

$$A^- = IA^-I = (A^{-1}A)A^-(AA^{-1}) = A^{-1}(AA^-A)A^{-1} = A^{-1}AA^{-1} = A^{-1}. \quad \blacksquare$$

Teorema 1.2.2 *Neka je $\mathbf{GI(A)}$ skup svih uopštenih inverznih matrica matrice A i neka $A^- \in \mathbf{GI(A)}$. Tada važi:*

1. *A^-A je idempotentna matrica, tj.*

$$(A^-A)^2 = A^-A;$$

2. *Opšte rešenje homogenog sistema jednačina $AX = 0$ je dato sa:*

$$X = (A^-A - I)z,$$

gde je z proizvoljan vektor;

3. *Opšte rešenje nehomogenog sistema jednačina $AX = Y$ je dato sa*

$$X = A^-Y + (A^-A - I)z,$$

gde je z proizvoljan vektor.

Dokaz.

1. $(A^-A)^2 = (A^-A)(A^-A) = A^-(AA^-A) = A^-A$;

2. $X = (A^-A - I)z$ *je zaista rešenje sistema* $AX = 0$*:*

$$AX = A(A^-A - I)z = A(A^-Az - Iz) = AA^-Az - Az = Az - Az = 0;$$

3. $A^-Y + (A^-A - I)z$ *je zaista rešenje sistema* $AX = Y$*:*

$$\begin{aligned} AX &= A(A^-Y + (A^-A - I)z) \\ &= A(A^-Y + A^-Az - Iz) \\ &= AA^-Y + AA^-Az - Az \\ &= AA^-Y + Az - Az \\ &= AA^-Y \end{aligned}$$

$$Y = AX = (AA^-A)X = (AA^-)(AX) = AA^-Y. \quad \blacksquare$$

1.3 Kronekerov produkt

Definicija 1.3.1 *Neka je $A = (a_{ij})$ matrica, gde $i = 1, 2, ..., m,\ j = 1, 2, ..., n$. Tada je za bilo koju matricu B Kronekerov produkt $A \otimes B$ blok matrica čiji su blokovi $(a_{ij}B)$ gde $i = 1, 2, ..., m,\ i\ j = 1, 2, ..., n$.*

Primetimo da ako je matrica A dimenzije $m \times n$, a matrica B dimenzije $p \times q$, onda je Kronekerov produkt $A \otimes B$ matrica dimenzije $mp \times nq$. Na primer, ako za matricu A uzmemo jediničnu, tj. $A = I_m$, a za matricu B matricu koja sadrži sve jedinice, tj. $B = 1_n$, onda kao Kronekerov produkt dobijamo:

$$A \otimes B = diag(1_n, ..., 1_n).$$

Neke od osobina[4] Kronekerovog produkta su sledeće:

- $(A_1 + A_2) \otimes B = A_1 \otimes B + A_2 \otimes B$.
- $A \otimes (B_1 + B_2) = A \otimes B_1 + A \otimes B_2$.
- $c \otimes A = A \otimes c = cA$, gde je c realan broj.
- $A \otimes (B \otimes C) = (A \otimes B) \otimes C$.
- $(A \otimes B)' = A' \otimes B'$.
- Ako je A blok matrica, tj.

  $$A = \begin{bmatrix} A_1 & A_2 \end{bmatrix},$$

  onda važi:

  $$\begin{bmatrix} A_1 & A_2 \end{bmatrix} \otimes B = \begin{bmatrix} A_1 \otimes B & A_2 \otimes B \end{bmatrix}.$$

  Međutim ako je B blok matrica, onda važi:

  $$A \otimes \begin{bmatrix} B_1 & B_2 \end{bmatrix} \neq \begin{bmatrix} A \otimes B_1 & A \otimes B_2 \end{bmatrix}.$$

- $(A_1 \otimes B_1)(A_2 \otimes B_2) = (A_1A_2) \otimes (B_1B_2)$.
- $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$, gde su A i B regularne matrice.
- $rank(A \otimes B) = rank(A)rank(B)$.
- $tr(A \otimes B) = tr(A)tr(B)$.
- Ako su A i B matrice dimenzije $m \times m$ i $n \times n$ respektivno, onda važi: $|A \otimes B| = |A|^m|B|^n$.
- Karakteristični koreni Kronekerovog produkta $A \otimes B$ su svi mogući proizvodi karakterističnih korena matrica A i B.

Glava 2

Mešoviti linearni modeli

Kako bi razumeli mešovite linearne modele potrebno je poznavanje linearnih regresijskih modela čiji je opšti oblik $y = X\beta + \varepsilon$, gde je y vektor posmatranja, X poznata kovarijansna matrica, β vektor nepoznatih koeficijenata i ε vektor slučajnih grešaka. U ovakvom modelu se smatra da su regresijski koeficijenti fiksni. Međutim, postavlja se pitanje šta se dešava kada su u model uključeni i slučajni efekti. U linearnim mešovitim modelima uključuju se, pored fiksnih i slučajni efekti, što nam omogućava kontrolu dodatnih izvora varijabilnosti. Pored toga, razmatranjem odnosa fiksnih i slučajnih efekata možemo detaljnije ispitati određenu pojavu. Uključivanje slučajnih faktora u model je tipično za merenja koja su korelisana. U poljoprivredi, medicini, psihologiji merenja se često vrše na individuama tokom vremena, što dovodi do povezanosti između samih merenja, posebno ako je razmak između tih merenja relativno mali. Kod uzgoja životinja može biti povezan npr. prinos mleka kod mlečnih krava koje potiču od istog oca, što je kasnije i razmatrano na jednom od primera.

Pretpostavimo sada da je svaka individua povezana sa nekim slučajnim efektom, na kojem nije vršeno posmatranje. Neka je sa y_{ij} označeno posmatranje i-te individue u vremenu j i neka je α_i slučajni efekat povezan sa i-tom individuom. Pretpostavimo da posmatranje vršimo na m individua u k vremenskih trenutaka. Kako bi pojednostavili smatraćemo da se merenja vrše istovremeno na svakoj individui, u istim vremenskim intervalima. Tada imamo model:

$$y_{ij} = x'_{ij}\beta + \alpha_i + \varepsilon_{ij},$$

gde je x_{ij} poznati vektor, β vektor nepoznatih regresijskih koeficijenata. Pretpostavljamo da su slučajni efekti α_i, i slučajne greške ε_{ij}, gde $i = 1, 2, ..., m$, $j = 1, 2, ..., k$, nezavisni i jednako raspoređeni sa sredinom nula i varijansama τ^2 i σ^2, respektivno. Pored toga smatramo i da su slučajni efekti i slučajne greške međusobno nezavisni. Odavde proizilazi da je korelacija između dva merenja na istoj individui $\tau^2/(\tau^2 + \sigma^2)$, pri čemu su posmatranja na različitim individuama nekorelisana. Ovaj model je specijalni slučaj longitudinalnog mešovitog linearnog modela o kome će biti reči kasnije. Dakle, za sada želimo opštiju klasu modela.[4]

Generalno se mešoviti modeli mogu predstaviti:

$$y = X\beta + Z\alpha + \varepsilon \tag{2.1}$$

gde su:
y vektor posmatranja;
β vektor nepoznatih regresijskih koeficijenata koje još nazivamo i fiksni efekti;
α vektor slučajnih efekata;
ε vektor slučajnih grešaka;
X, Z su poznate martice, koje povezuju elemente vektora β i α sa elementima vektora y.

Elementi vektora β predstavljaju fiksne, dok elementi α predstavljaju slučajne efekte populacije, gde nam je poznata varijansno - kovarijansna struktura. Osnovne pretpostavke statističkog modela (2.1) su da slučajni efekti i greške imaju očekivanje nula i konačnu varijansu. Varijansu ovih promenljivih označimo sa:

$$G = Var(\alpha),$$

$$R = Var(\varepsilon).$$

Takođe pretpostavljamo da su α i ε nekorelisane.
Dakle, očekivanja slučajnih promeljivih su:

$$\begin{aligned} E(\alpha) &= 0 \\ E(\varepsilon) &= 0 \\ E(y) &= E(X\beta + Z\alpha + \varepsilon) \\ &= E(X\beta) + E(Z\alpha) + E(\varepsilon) \\ &= XE(\beta) + ZE(\alpha) + E(\varepsilon) \\ &= X\beta \end{aligned}$$

Varijansno-kovarijansnu strukturu obično predstavljamo kao:

$$Var \begin{bmatrix} \alpha \\ \varepsilon \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix},$$

gde su G i R poznate, pozitivno definitne matrice.

Dakle,

$$\begin{aligned}
Var(y) &= Var(X\beta + Z\alpha + \varepsilon) \\
&= Var(Z\alpha + \varepsilon) \\
&= ZVar(\alpha)Z' + Var(\varepsilon) + ZCov(\alpha, \varepsilon) + Cov(\varepsilon, \alpha)Z' \\
&= ZGZ' + R \\
Cov(y, \alpha) &= ZG \\
Cov(y, \varepsilon) &= R
\end{aligned}$$

2.1 Klasifikacija mešovitih linearnih modela

Jedan od načina klasifikacije mešovitih linearnih modela jeste na osnovu pretpostavke o raspodeli podataka. Pretpostavka o normalnoj raspodeli slučajnih efekata i grešaka nam uglavnom omogućava više fleksibilnosti u modeliranju.

2.1.1 Gausovi mešoviti linearni modeli

Kod Gausovih mešovitih linearnih modela osnovna pretpostavka je da slučajni efekti i greške imaju normalnu raspodelu. Navešćemo neke od osnovnih modela ovog tipa.

1. ANOVA je jedan od najranijih Gausovih mešovitih modela koji se spominju u literaturi. ANOVA podrazumeva analizu varijanse.

   - (*One way random effects model*) Ovaj model se naziva model slučajnih efekata iz razloga što nam je nepoznata sredina populacije.[4]
     Pretpostavimo da imamo merenja y_{ij}, gde $i = 1, 2, ..., m$ i $j = 1, 2, ..., k_i$. Takođe, neka merenja y_{ij}, za svako i i j, zadovoljavaju model:

     $$y_{ij} = \mu + \alpha_i + \varepsilon_{ij},$$

     gde je μ nepoznata sredina, α_i, $i = 1, 2, ..., m$ slučajni efekti normalno distribuirani, tj. $\mu : N(0, \tau^2)$ i ε_{ij} greške,takođe sa normalnom raspodelom $\varepsilon_{ij} : N(0, \sigma^2)$. Pri tome smatramo da slučajni efekti i greške nezavisni. Uobičajeno je da su nam njihove varijanse nepoznate. Cilj nam je sve ovo prikazati pomoću modela (2.1). Dalje, neka je y_{ij} gde $j = 1, 2, ..., k_i$ vektor kolona posmatranja za i-tu grupu. Analogno je i $\varepsilon_i = (\varepsilon_{ij})_j$, gde $j = 1, 2, ..., k_i$. Sada možemo zapisati

     $$y = \begin{bmatrix} y'_1 & y'_2 & \dots & y'_m \end{bmatrix},$$

     $$\alpha = (\alpha_i), i = 1, 2, ..., m,$$

     $$\varepsilon = \begin{bmatrix} \varepsilon'_1 & \varepsilon'_2 & \dots & \varepsilon'_m \end{bmatrix}.$$

     Lako se može pokazati da se za odgovarajuće X i Z i $\beta = \mu$ model može iskazati pomoću (2.1) pri čemu

     $$\alpha : N(0, \tau^2 I_m),$$

     $$\varepsilon : N(0, \sigma^2 I_n),$$

     gde je $n = \Sigma k_i$. Specijalno, ako važi da je $k_i = k$ za svako i onda govorimo o balansiranom modelu.[4] U tom slučaju se matrice X i Z mogu prikazati kao:

     $$X = 1_m \otimes 1_k = 1_{mk},$$

     $$Z = I_m \otimes 1_k,$$

     gde je $\otimes$ označava Kronekerov produkt, objašnjen ranije.

- (*Two way random effects model*) U ovom slučaju pretpostavljamo da merenja y_{ij}, gde $i = 1, 2, ..., m$ i $j = 1, 2, ..., k$ zadovoljavaju model:

$$y_{ij} = \mu + \xi_i + \eta_j + \varepsilon_{ij}$$

za sve i i j, gde je μ ponovo sredina populacije, ξ_i, $i = 1, 2, ..., m$ i $\eta_j, j = 1, 2, ..., k$ su nezavisni slučajni efekti, a ε_{ij} su nezavisne slučajne greške. Sada imamo sledeće raspodele:

$$\begin{aligned} \xi_i &: N(0, \tau_1^2), \\ \eta_i &: N(0, \tau_2^2), \\ \varepsilon_{ij} &: N(0, \sigma^2). \end{aligned}$$

Takođe, ponovo pretpostavljamo da su slučajni efekti i greške nezavisni. Generalno se ANOVA model pomoću (2.1) definiše tako da

$$Z\alpha = Z_1\alpha_1 + Z_2\alpha_2 + ... + Z_s\alpha_s,$$

gde su $Z_1, Z_2, ..., Z_s$ poznate matrice, $\alpha_1, \alpha_2, ..., \alpha_s$ vektori slučajnih efekata. Pretpostavka je da su α_i, za svako i nezavisni sa raspodelama $N(0, \tau_i^2)$. Komponete ε su nezavisne sa raspodelom $N(0, \sigma^2)$. Takođe, $\alpha_1, \alpha_2, ..., \alpha_s, \varepsilon$ su nezavisni.
Skup τ_i^2 i σ^2 u ANOVA modelima nazivamo varijanse komponenti. ANOVA model je generalno balansiran ako se matrice X i Z_i, $i = 1, 2, ..., s$, mogu predstaviti kao $X = \bigotimes_{l=1}^{\omega+1} 1_{n_l}^{a_l}$ i $Z_i = \bigotimes_{l=1}^{\omega+1} 1_{n_l}^{b_{i,l}}$, gde $(a_1, ..., a_l) \in S_{\omega+1} = \{0, 1\}^{\omega+1}$, $(b_{i,1}, ..., b_{i,\omega+1}) \in S \subset S_{\omega+1}$. n_l predstavlja broj broj nivoa faktora l, $l = 1, ..., \omega$, pri čemu imamo da je $a_{s+1} = 1$ i da je $b_{i,s+1} = 1$ za svako i.
Za prvi primer imamo vrednosti $\omega = 1$, $n_1 = m$, $n_2 = k$ i $S = \{(0, 1)\}$ dok u drugom navedenom primeru imamo da je $\omega = 2$, $n_1 = m$, $n_2 = k$, $n_3 = 1$, dok je skup S u tom slučaju $\{(0, 1, 1), (1, 0, 1)\}$.[4]

2. *Longitudinalni model.* Naziv ovog modela potiče od toga što se najčešće koristi za analizu longitudinalnih podataka, tj. takvih podataka gde je neka karakteristika ili osobina merena na istom subjektu više puta. Analizu takvih podataka nazivamo longitudinalne studije. Takve studije mogu biti *eksperimentalne* (kontrolisani uslovi) i *posmatračke* (nekontrolisani uslovi). Kada su u pitanju ovakvi podaci, posmatračke studije imaju širu upotrebu. Vreme prikupljanja podataka može biti izraženo različitim jedinicama, godinama, semestrima, mesecima, danima, itd. Periodičnost prikupljanja podataka može biti ista ili različita, te stoga razlikujemo fiksni i fleksibilan raspored prikupljanja podataka. Treba uzeti u obzir da su merenja na istom subjektu u korelaciji.[4] Generalno se longitudinalni mešoviti model može predstaviti:

$$y_i = X_i\beta + Z_i\alpha_i + \varepsilon_i, i = 1, 2, ...m$$

gde su y_i vektor posmatranja na i-toj individui, X_i i Z_i poznate matrice, β nepoznati vektor regresijskih koeficijenata, α_i vektor slučajnih efekata i ε

vektor slučajnih grešaka. Pretpostavlja se da su α_i i ε_i nezavisni za svako $i = 1, 2, ...m$, pri čemu važi:

$$\alpha_i : N(0, G_i),$$

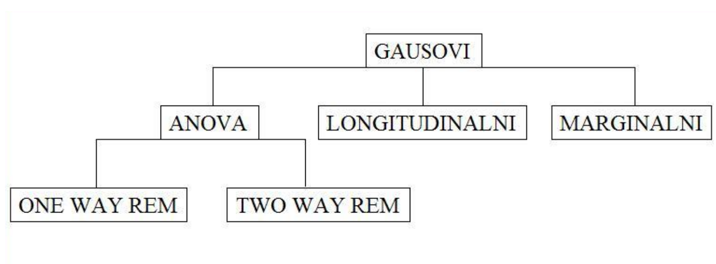
$$\varepsilon_i : N(0, R_i),$$

gde su kovarijansne matrice G_i i R_i poznate do na vektor varijanse komponenti θ.

3. *Marginalni model.* Gausov mešoviti model se može prikazati pomoću marginalne raspodele. Uzimajući u obzir model (2.1) i normalnost dobijamo:

$$y : N(X\beta, V), \qquad V = R + ZGZ' \tag{2.2}$$

Dakle, ovo je jednostavniji prikaz mešovitog modela. Za ANOVA model imamo da je $V = \sigma^2 I_n + \sum_{i=1}^{s} \sigma_i^2 Z_i Z_i'$. Nedostatak ovog modela jeste što slučajni efekti nisu eksplicitno definisani.[4] Neretko slučajni efekti imaju praktično značenje, pa zaključci o njima mogu biti od interesa.



Slika 2.1: Klasifikacija Gausovih mešovitih linearnih modela.

2.1.2 Ne-Gausovi mešoviti linearni modeli

Osnovna karakteristika ovih modela je da slučajni efekti i greške nemaju normalnu raspodelu. Pretpostavka o njihovoj nezavisnosti ostaje nepromenjena.

1. ANOVA. Ne-Gausov mešoviti linearni model se takođe definiše sa (2.1) gde su α_i slučajni efekti sa sredinom 0 i varijansom τ_i^2, $i = 1, 2, ..., s$. Komponente vektora ε takođe imaju sredinu 0, a njihova varijansa je σ^2. I u ovom slučaju važi pretpostavka da su $\alpha_1, \alpha_2, ..., \alpha_s, \varepsilon$ nezavisni. Sve ostale pretpostavke su iste kao i u Gausovom mešovitom modelu. Neka je sa $F_i, i = 1, 2, ..., s$ označena zajednička distribucija komponenti α_i, a sa G komponenti ε. Tada ako parametarske forme $F_1, F_2, ..., F_s, G$ nisu poznate, onda distribucija slučajne promenljive y nije određena do na skup parametara. Sve dok nije ispunjena pretpostavka o normalnosti nije moguće distribuciju od y analitički predstaviti.[4]

2. *Longitudinalni model.* Gausov longitudinalni mešoviti model se takođe može proširiti na ne-Gausov, ukidanjem pretpostavke o normalnosti. U slučaju ovog modela se pretpostavlja da su $y_1, y_2, ..., y_m$ nezavisne slučajne promenljive i da su α_i i ε_i nekorelisane za svako i. Pri tome imamo da:

$$E(\alpha_i) = 0 \qquad Var(\alpha_i) = G_i,$$
$$E(\varepsilon_i) = 0 \qquad Var(\varepsilon_i) = R_i.$$

   Ponovo imamo da se distribucija slučajne promenljive y ne može u potpunosti navesti do na skup parametara.[4]

3. *Marginalni model.* Ovo je možda najopštiji model od dosad navedenih tipova. U ovom modelu se pretpostavlja da slučajna promenljiva y zadovoljava:

$$E(y) = X\beta \qquad Var(y) = V,$$

   gde nam je matrica V poznata do na vektor θ. Ovaj model je nastao od Gausovog modela izbacujući pretpostavku o normalnosti. I dalje je glavni nedostatak što slučajni efekti nisu reprezentovani. Stoga, manjak pretpostavki u ovom modelu, tj. njegova opštost znatno otežava procenu osobina ocenjivača.[4]

Slika 2.2: Klasifikacija Ne-Gausovih mešovitih linearnih modela.

2.2 Metodi ocenjivanja za Gausove modele

Kroz istoriju su postojali mnogi metodi za ocenjivanje kada su u pitanju Gausovi linearni mešoviti modeli. U ovom poglavlju ćemo pažnju usmeriti na dva standardna metoda ocenjivanja, a to su su metod maksimalne verodostojnosti (*Maximum Likelihood* - **ML** *method*) i limitiran metod maksimalne verodostojnosti (*Restricted Maximum Likelihood* - **REML** *method*).

1. **Metod maksimalne verodostojnosti** je definisao i razvio engleski statističar Ronald Fisher još početkom dvadesetog veka.[4] On je predložio da se za ocenu parametara uzima ocena koja daje najveću verovatnoću realizacije uzoračkih podataka. Ipak, za mešovite linearne modele ga koriste tek Hartley i Rao(1967). Glavni razlog za to je što ocenjivanje varijansi u mešovitom modelu nije bilo jednostavno.

   Znamo od ranije da za Gausove linearne mešovite modele važi:

   $$y : N(X\beta, V), \qquad V = R + ZGZ'.$$

   Tada je zajednička funkcija gustine raspodele data sa

   $$f(y) = \frac{1}{(2\pi)^{\frac{n}{2}}|V|^{\frac{1}{2}}} e^{-\frac{1}{2}(y-X\beta)'V^{-1}(y-X\beta)},$$

   gde je n dimenzija slučajnog vektora y. Dalje je logaritam funkcije:

   $$l(\beta,\theta) = c - \frac{1}{2}log(|V|) - \frac{1}{2}(y - X\beta)'V^{-1}(y - X\beta),$$

   gde θ predstavlja vektor varijansi komponenti, a c neku konstantu.

   Diferenciranjem funkcije $l(\beta,\theta)$ po parametrima dobijamo sledeće:

   $$\frac{\partial l}{\partial \beta} = X'V^{-1}y - X'V^{-1}X\beta, \tag{2.3}$$

   $$\frac{\partial l}{\partial \theta_r} = \frac{1}{2}[(y - X\beta)'V^{-1}\frac{\partial V}{\partial \theta_r}V^{-1}(y - X\beta) - tr(V^{-1}\frac{\partial V}{\partial \theta_r})], \tag{2.4}$$

   gde je θ_r r-ta komponenta vektora θ, $r = 1, 2, ..., q$.

   Standardna procedura pronalaženja ML ocene podrazumeva rešavanje sistema jednačina

   $$\frac{\partial l}{\partial \beta} = 0,$$

   $$\frac{\partial l}{\partial \theta_r} = 0, \qquad r = 1, 2, ..., q.$$

Međutim takva rešenja ne moraju biti nužno ocena dobijena metodom maksimalne verodostojnosti, što ćemo videti kasnije.

Neka je sada β vektor dimenzije p. Radi jednostavnosti pretpostavimo da je matrica X punog ranga (po kolonama), tj. $rank(X) = p$. Neka je $(\hat{\beta}, \hat{\theta})$ MLE, tj. ocena dobijena metodom maksimalne verodostojnosti. Tada iz (2.3) dobijamo:

$$\hat{\beta} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}y,$$

gde je $\hat{V} = V(\hat{\theta})$, odnosno u V su komponente varijansi zamenjene njihovom ocenom.
ML ocena za β dobija se ubacivanjem ocene za θ u izraz (2.4).

Lako je pokazati da ocena metodom maksimalne verodostojnosti za θ zadovoljava

$$y'P\frac{\partial V}{\partial \theta_r}Py = tr(V^{-1}\frac{\partial V}{\partial \theta_r}), \qquad r = 1, 2, ..., q,$$

gde je

$$P = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}. \tag{2.5}$$

Stoga je često procedura da se prvo reši (2.5), a da se potom β dobije iz (2.4).

Ocene dobijene metodom maksimalne verodostojnosti mogu biti i neefikasne i pristrasne. Ipak zna se da pod nekim uslovima ova ocena može biti asimtotski efikasna i da može imati asimtotski normalnu raspodelu.

2. **Metod limitirane maksimalne verodostojnosti** se razvio iz razloga što ocena dobijena metodom maksimalne verodostojnosti ne mora biti konzistentna ocena. Neyman i Scott(1948) su radili na jednom problemu u cilju da pokažu da kada se povećava broj parametara sa veličinom uzorka, ocena dobijena metodom maksimalne verodostojnosti ne mora biti konzistentna.[4]
Pretpostavimo sada da imamo dva posmatranja koja su sakupljena na m individua. Neka je μ_i sredina za i-tu individuu i neka su posmatranja nezavisna i normalno distribuirana sa varijansom σ^2. Cilj nam je oceniti varijansu σ^2.
Posmatramo model:

$$y_{ij} = \mu_i + \varepsilon_{ij},$$

gde su ε_{ij} nezavisni i normalno distribuirani, sa sredinom 0 i varijansom σ^2. Ovaj model se može posmatrati kao specijalni slučaj mešovitog linearnog modela (2.1), kada je $Z = 0$. Može se pokazati da MLE u ovom slučaju nije konzistentna ocena. Taj problem je vezan za to što MLE generalno nije nepristrasna ocena. Ta pristrasnost se ne može ukloniti sa povećanjem uzorka, ako je broj fiksnih parametara proporcionalan sa veličinom uzorka. Problem je što nam fiksni efekti nisu od interesa, a sa druge strane nam smetaju pri ocenjivanju varijanse. Potreban nam je model koji se neće baviti parametrima koji nisu od interesa, nasuprot metodi ML kojom se ocenjuju svi parametri u modelu. Dakle, neka imamo $m+1$ parametar, pri čemu parametri $\mu_1, ..., \mu_m$ nisu od interesa, tj. jedini cilj nam je oceniti varijansu. Broj parametara je proporcionalan veličini uzorka koja iznosi $2m$.

Kako bi eliminisali fiksne efekte izvršićemo jednostavnu transformaciju na podacima.[4] Neka je

$$z_i = y_{i1} - y_{i2}.$$

Osnovna razlika je što su sada fiksni efekti uključeni u raspodelu slučajnih promenljivih z_i, $i = 1, 2, ..., m$. Nakon transformacije imamo m opažanja i samo jedan parametar koji ocenjujemo. Lako je uočiti da $z_1, ..., z_m$ imaju normalnu raspodelu sa sredinom 0 i varijansom $2\sigma^2$. Može se pokazati da će ocena metodom maksimalne verodostojnosti zasnovana na ovim podacima biti konzistentna.[4]

Sada ćemo dati uopšten postupak dolaženja do REML ocene. Pretpostavimo b.u.o da $rank(X) = p$ i neka je A matrica dimenzije $n \times (n-p)$ takva da

$$rank(A) = n - p, \qquad A'X = 0.$$

Dalje, neka je $z = A'y$. Očigledno važi:

$$z : N(0, A'VA).$$

Tada je zajednička funkcija gustine verovatnoće za z data sa:

$$f_R(z) = \frac{1}{(2\pi)^{\frac{n-p}{2}} |A'VA|^{\frac{1}{2}}} e^{-\frac{1}{2} z'(A'VA)^{-1} z},$$

gde R u indeksu označava restrikciju. Dalje je log-verovatnoća u oznaci l_R dat sa:

$$l_R(\theta) = c - \frac{1}{2} log(|A'VA|) - \frac{1}{2} z'(A'VA)^{-1} z, \tag{2.6}$$

gde je c neka konstanta. Diferenciranjem ove fukcije dobijamo:

$$\frac{\partial l_R}{\partial \theta_i} = \frac{1}{2}\{y'P\frac{\partial V}{\partial \theta_i}Py - tr(P\frac{\partial V}{\partial \theta_i})\},$$

gde je $i = 1, 2, ...q$, pri čemu je q dimenzija vektora varijansi θ, a

$$P = A(A'VA)^{-1}A' = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}XV^{-1}.$$

REML ocena za θ dobija se maksimiziranjem funkcije (2.6). Analogno, kao i u slučaju ML ocene, maksimum se dobija rešavanjem jednačina:

$$\frac{\partial l_R}{\partial \theta_i} = 0, \quad i = 1, 2, ...q.$$

Bitno je napomenuti da REML ocena ne zavisi od A. Može se pokazati da ako A zamenimo matricom $B = AT$, gde je T bilo koja regularna matrica reda $(n-p)(n-p)$, REML ocena se neće promeniti. Dakle, matrica A nije jedinstvena. To je važno jer ocena ne zavisi od transformacije koju smo izabrali.
U posmatranom primeru, za transformaciju $z_i = y_{i1} - y_{i2}$, matrica A bi izgledala:

$$A = \begin{bmatrix} 1 & -1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & -1 \end{bmatrix}'.$$

Međutim možemo izvršiti i bilo koju drugu transformaciju sa $B = AT$, gde je T regularna matrica dimenzije $m \times m$, a pri tome se ocena dobijena ovim metodom neće promeniti.[4]

Treba imati na umu da se ovom metodom procenjuje isključivo varijansa. Parametri β se ne ocenjuju jer su oni eliminisani pre ocene. Uprkos tome, β nekada dobija na isti način kao i u metodu maksimalne verodostojnosti. Ocenu za β dobijamo tako što u izraz za ML ocenu za $\hat{\beta}$ uvrstimo $\hat{V} = V(\hat{\theta})$, gde je $\hat{\theta}$ ocena dobijena REML metodom. Takava ocena se nekad naziva "REML ocena"[4] za β.

2.3 Metodi ocenjivanja za ne-Gausove modele

Prethodne metode su se zasnivale na pretpostavci o normalnoj raspodeli. Međutim u praksi to često nije slučaj. Postoje razni realni primeri koji ne zadovoljavaju normalnu raspodelu. Iz tog razloga u ovom poglavlju će biti razmatrani metodi ocenjivanja za ne-Gausove mešovite linearne modele. Kako ne postoji generalni metod, razmotrićemo neke pojedinačne slučajeve.

2.3.1 Longitudinalni linearni mešoviti modeli

1. **Metod ponderisanih najmanjih kvadrata** (*Weighted Least Squares*-WLS) se koristi u analizi longitudinalnih podataka za ne-Gausove longitudinalne linearne mešovite modele koje smo ranije razmatrali.
   Pretpostavimo da se podaci prikupljaju na individuama tokom vremena. Sa y označimo vektor zapažanja koja mogu biti u korelaciji.[4] Neka je X poznata matrica i b.u.o pretpostavimo da je X punog ranga kao i da važi

   $$E(y) = X\beta,$$

   gde je β vektor nepoznatih regresijskih koeficijenata. Tada se WLS ocena za β dobija minimizacijom izraza

   $$(y - X\beta)'W(y - X\beta), \tag{2.7}$$

   gde je W poznata simetrična ponder matrica. Tada je za regularnu matricu W minimum dat sa

   $$\hat{\beta}_W = (X'WX)^{-1}X'Wy. \tag{2.8}$$

   Kada za matricu W izaberemo jediničnu matricu dobijamo običan metod najmanjih kvadrata i u tom slučaju je ocena parametra β

   $$\hat{\beta}_I = (X'X)^{-1}X'y. \tag{2.9}$$

   S druge strane poznato je da je optimalan izbor matrice W ustvari matrica V^{-1}, gde je $V = Var(y)$. U tom slučaju dobijamo najbolji linearni nepristrasan ocenjivač

   $$\hat{\beta}_{BLUE} = (X'V^{-1}X)^{-1}X'y. \tag{2.10}$$

   Međutim kada je matrica V nepoznata, nije moguće izračunati $\hat{\beta}_{BLUE}$ ocenu. U slučaju kad je β poznato možemo oceniti V. Tada je nepristrasan, ali ne i konzistentan ocenjivač za V dat sa

   $$\tilde{V} = (y - X\beta)(y - X\beta)'$$

   Ako je V u potpunosti nepoznat, onda imamo $\frac{n(n+1)}{2}$ nepoznatih parametara koje treba oceniti. Međutim, u tom slučaju i ne očekujemo da ćemo imati konzistentan ocenjivač jer bi nam neke informacije iz V morale biti dostupne.

   Radi jednostavnosti sada ćemo razmotriti specijalan balansiran slučaj. Dakle, pretpostavimo da su podaci sakupljeni na istom skupu vremenskih trenutaka.

Neka je y_{ij} predstavlja posmatranje na i-toj jedinki u vremenskom trenutku t_j, gde $j = 1, 2, ..., k$ i neka je

$$y_i = (y_{ij})_{1 \leq j \leq k},$$

gde su $y_1, y_2, ..., y_m$ međusobno nezavisni, $j = 1, 2, ..., m$, i važi

$$E(y_i) = X_i, \qquad Var(y_i) = V_0,$$

gde je $V_0 = (v_{qr})_{1 \leq q,r \leq k}$ poznata kovarijansna matrica. Ako je k fiksno možemo naći konzistentan ocenjivač. Specijalno ako je β poznato konzistentan ocenjivač će biti

$$\hat{V} = diag(\hat{V_0}, ..., \hat{V_0}), \tag{2.11}$$

gde je

$$\hat{V_0} = \frac{1}{m}(y_i - X_i\beta)(y_i - X_i\beta)'. \tag{2.12}$$

Suština je u tome da ako nam je poznato V možemo koristiti $\hat{\beta}_{BLUE}$ iz 2.10, kao ocenu za β. S druge strane, ukoliko nam je poznato β, ocenu za V dobijamo jednostavno iz (2.11) i (2.12).
Međutim, jasno je da postoji direktna veza između ovih parametara koji se trebaju oceniti, ali se postavlja pitanje šta možemo uraditi ako su nam i V i β nepoznati. U tom slučaju posmatramo iterativni postupak. Početni zadatak je oceniti β običnom metodom najmanjih kvadrata. Potom se tako dobijena ocena $\hat{\beta}_I$ uvrštava u (2.11), kao bi dobili ocenu za V. Dalje nastavljamo postupak tako što dobijenu ocenu $\hat{V}$ vraćamo u izraz (2.10) kako bi dobili sledeću ocenu za β. Ovaj postupak nazivamo iterativni postupak ponderisanih najmanjih kvadrata i označavamo sa I-WLS.[4]

Može se pokazati da ako važi pretpostavka o normalnosti i ovaj postupak konvergira, onda se granična ocena poklapa sa ocenom metodom maksimalne verodostojnosti.[4]

2. **Jackknife metod** je prvi uveo Quenouille(1949), a kasnije ga je razvio Tukey(1958).[4] Za sada znamo da iterativni postupak I-WLS konvergira ocenama za β i V. Glavni cilj je dobiti efikasniju ocenu. U tom slučaju ocenu za $\Sigma = Var(\hat{\beta})$ možemo dobiti zamenom V sa $\hat{V}$ u kovarijansnoj matrici $(X'V^{-1}X)^{-1}$ iz (2.10). Međutim kako nije uzeta u obzir dodatna varijacija uzrokovana ocenjivanjem V, ocena Σ će verovatno potceniti realnu varijansu.[4]
S druge strane, za ocenjivanje pristrasnosti i varijanse ocene nepoznatog parametra možemo koristiti jackknife metod. Prednost ove metode je što se može koristiti za različite metode ocenjivanja parametara.
Glavna ideja ovog metoda je da se uzorak podeli na određen broj grupa. Broj grupa može varirati, ali je u praksi najčešći slučaj da se uzorak deli na onoliko grupa koliki je obim uzorka. Potom se prave novi uzorci na taj način što sve grupe osim jedne ulaze u uzorak. Ako je veličina grupe jednaka jedinici,

onda se postupak svodi na izbacivanje jednog elementa iz uzorka. U ovom postupku dobija se uzorak obima za jedan manji od polaznog, pri čemu broj elemenata polaznog uzorka ostaje nepromenjen.
Posmatrajmo sada ne-Gausove modele. Pretpostavimo da je vektor $\psi = (\beta', \theta')$ ocenjen sa $\hat{\psi}$, gde je $\hat{\psi}$ M-ocena koja je rešenje sledeće jednačine:

$$\sum_{i=1}^{m} f_i(\psi, y_i) + a(\psi) = 0. \tag{2.13}$$

f_i i a su vektor vrednosti iste dimenzije kao i ψ, takve da je $E(f_i(\psi, y_i)) = 0$, za $i = 1, 2, ...m$. M-ocena podrazumeva na primer ML i REML ocene kao granice iterativnog postupka ponderisanih najmanjih kvadrata. Sada iz uzorka izbacujemo i-ti element i dobijamo ocenu $\hat{\psi}_{-i}$ kao rešenje jednačine

$$\sum_{i=1}^{m} f_i(\psi, y_i) + a_{-i}(\psi) = 0. \tag{2.14}$$

Neka je

$$\Sigma = Var(\hat{\psi}),$$

asimptotska kovarijansna matrica za ocenu $\hat{\psi}$. Tada je jackknife ocena za Σ data sa

$$\hat{\Sigma}_{Jack} = \frac{m-1}{m} \sum_{i=1}^{m} (\hat{\psi}_{-i} - \hat{\psi})(\hat{\psi}_{-i} - \hat{\psi})'. \tag{2.15}$$

Jiang i Lahiri(2004) su pokazali da je ocena iz (2.15) dobijena jackknife metodom konzistentna.[4]

2.4 Procena slučajnih efekata

2.4.1 BLUP

Dok se kod ocenjivanja fiksnih efekata koristi najbolji linearni nepristrasni ocenjivač BLUE (*Best linear unbiased estimator*), s druge strane se kod ocenjivanja slučajnih efekata koristi najbolji linearni nepristrasan prediktor BLUP (*Best linear unbiased predictor*). Najbolji prediktor podrazumeva minimizaciju srednje kvadratne greške (*Mean squared error - MSE*). Nepristrasnost podrazumeva da je $E(\hat{\alpha}) = \alpha$. Ovaj način predviđanja ima dugu istoriju koju je započeo Charles Roy Henderson(1948) u toku svog istraživanja vezanog za uzgoj životinja (animal breeding[4]). Kasnije je Robinson(1991) detaljnije izveo BLUP sa nekoliko primera i njegovom primenom.

Osnovni problem jeste predvideti funkciju:

$$K'\beta + M'\alpha$$

Prvo ćemo razmotriti na koji način se vrši predviđanje kada su nam fiksni efekti i varijanse komponeti poznati. Tada je najbolji prediktor, u smislu minimalne MSE, dat pomoću uslovne sredine:

$$E(M'\alpha|y) = M'E(\alpha|y).$$

Najbolji prediktor može biti linearan ili nelinearan. Nelinearni prediktori su često teški za izvođenje, pa je stoga njihova upotreba veoma ograničena. Mi ćemo se bazirati na linarne prediktore. U tom slučaju potrebno nam je da znamo očekivanje i varijansu slučajne promenljive y.
Neka je

$$Var(y) = ZGZ' + R = V.$$

Ako pretpostavimo normalnost imamo:

$$\begin{bmatrix} \alpha \\ y \end{bmatrix} : N(\begin{bmatrix} 0 \\ X\beta \end{bmatrix}, \begin{bmatrix} G & GZ' \\ ZG & V \end{bmatrix}).$$

Tada je najbolji prediktor za $M'\alpha$:

$$M'GZ'V^{-1}(y - X\beta).$$

Nekada se najbolji prediktor za $M'\alpha$ dobija kao najbolji prediktor za $\eta = K'\beta + M'\alpha$:

$$\hat{\eta_B} = K'\beta + M'GZ'V^{-1}(y - X\beta), \tag{2.16}$$

Pri tome se B u formulaciji $BLUE$ odnosi na najbolji (*Best*) prediktor.

Može se pokazati da je $\hat{\eta_B}$ najbolji linearni prediktor za η i to bez pretpostavke o normalnosti. Međutim, u slučaju kada su nam fiksni efekti nepoznati, pri čemu su nam poznate varijanse, ne možemo koristi (2.16) kao prediktor. Tada je uobičajeno zameniti β sa ocenom $\hat{\beta}$, što dobijamo pomoću metoda maksimalne verodostojnosti, pri čemu koristimo pretpostavku o normalnosti.
Dakle,
$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y \tag{2.17}$$

Poznato je da je $\hat{\beta}$, definisan na ovaj način, najbolji linearni nepristrasan ocenjivač čije izvođenje ne zahteva normalnost.
Henderson(1973)[4] je pokazao da ako u (2.16) β zamenimao sa $\hat{\beta}$ dobijamo najbolji linearni nepristrasni prediktor za η, gde najbolji podrazumeva da ima minimalnu srednje kvadratnu grešku od svih linearnih nepristrasnih prediktora gde je:

$$MSE(\hat{\eta}) = E[(\hat{\eta} - \eta)(\hat{\eta} - \eta)'].$$

Dakle, ponovo smo dobili prediktor bez pretpostavke o normalnosti:

$$\hat{\eta}_{BLUP} = K'\hat{\beta} + M'GZ'V^{-1}(y - X\hat{\beta}), \tag{2.18}$$

gde je $\hat{\beta}$ najbolji linearni nepristrasni ocenjivač - $BLUE$ iz (2.17), a:

$$\hat{\alpha} = GZ'V^{-1}(y - X\hat{\beta}), \tag{2.19}$$

koji se takođe zove $BLUP$ za slučajne efekte α.

Originalno Hendersonovo izvođenje BLUP-a zasniva se onome što on naziva zajednička procena maksimalne verodostojnosti fiksnih i slučajnih efekata. Ako posmatramo Gausov mešoviti model (2.1), gde

$$\alpha : N(0, G)$$

$$\varepsilon : N(0, R)$$

i slučajni efekti i greške su nezavisni, a matrice G i R regularne, onda se može pokazati da se logaritam zajedničke funkcije gustine za α i y može prikazati kao:

$$c - \frac{1}{2}[(y - X\beta - Z\alpha)'R^{-1}(y - X\beta - Z\alpha) + \alpha' G^{-1}\alpha] \tag{2.20}$$

gde je c neka konstanta. Henderson(1950) je predložio da se do maksimalnog verodostojnog ocenjivača dođe diferenciranjem po α i β i izjednačavanjem tih izvoda sa nulom. Na taj način se došlo do mešovitog modela jednačina, oblika

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\alpha} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}.$$

Kao rešenja ovog modela jednačina dobijamo upravo (2.17) i (2.19). Kasnije je Henderson(1963) pokazao da takvo rešenja zaista i jeste $BLUP$. Harville(1990), kao i Robinson(1991) su takođe dali svoj pristup u pokazivanju najboljeg linearnog

nepristrasnog prediktora. Ono što je specifično jeste da ovakav prediktor ne zahteva normalnost pa se može primeniti i na ne-Gausove mešovite modele. Konačno je dokaz da je *BLUP* najbolji linearni nepristrasan prediktor dao Jiming Jiang(1998) koji se bazira na upoređivanju grešaka.
Sada ćemo prikazati jedan primer koji je dao Robinson(1991)[4] kako bi ilustrovao *BLUE* i *BLUP* [4].

Primer 2.4.1 *Posmatramo prinos mleka kod mlečnih krava pri čemu su genetski uticaji oca uzeti za slučajne efekte, a uticaj stada za fiksne efekte. Primenjujemo mešoviti linearni model. Efekte stada označavamo sa* β_j, $j = 1,2,3$ *, a sa* α_i, $i = 1,2,3,4$ *označavamo uticaj očeva A, B, C i D. Za matricu R uzimamo jediničnu matricu, a za G jediničnu matricu pomnoženu sa 0.1. Pretpostavljamo da su očevi nepovezani i da su varijanse:*

$$\tau^2 = Var(\alpha_i),$$

$$\sigma^2 = Var(\varepsilon_{ij}).$$

Podaci su dati u sledećoj tabeli:

Stado	*1*	*1*	*2*	*2*	*2*	*3*	*3*	*3*	*3*
Otac	*A*	*D*	*B*	*D*	*D*	*C*	*C*	*D*	*D*
Prinos mleka	*110*	*100*	*110*	*100*	*100*	*110*	*110*	*100*	*100*

Mešoviti model jednačina u matričnom obliku[4] izgleda:

$$\begin{bmatrix} 2 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 3 & 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 4 & 0 & 0 & 2 & 2 \\ 1 & 0 & 0 & 11 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 11 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 12 & 0 \\ 1 & 1 & 2 & 0 & 0 & 0 & 15 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \hat{\alpha}_3 \\ \hat{\alpha}_4 \end{bmatrix} = \begin{bmatrix} 210 \\ 310 \\ 420 \\ 110 \\ 110 \\ 220 \\ 500 \end{bmatrix},$$

Kao rešenje ovog modela jednačina, pomoću programskog paketa MATLAB dobijamo:

$$\hat{\beta} = \begin{bmatrix} 105.64 & 104.28 & 105.46 \end{bmatrix}',$$

$$\hat{\alpha} = \begin{bmatrix} 0.4 & 0.52 & 0.76 & -1.67 \end{bmatrix}'.$$

Dakle, najbolji prediktor za $K'\beta + M'\alpha$ možemo zapisati:

$$\hat{\eta} = K'\hat{\beta} + M'GZ'V^{-1}(y - X\hat{\beta})$$

gde je $K'\hat{\beta}$ najbolji, linearni, nepristrasni (BLUE) ocenjivač za $K'\beta$.
Pojam nepristrasnosti je malo drugačiji kada je u pitanju prediktor, jer je potrebno oceniti kako α tako i linearnu kombinaciju β.

Očekivanje prediktora $L'y$ je:

$$
\begin{aligned}
E(\hat{\eta}) &= E(K'\hat{\beta} + M'GZ'V^{-1}(y - X\hat{\beta})) && (2.21)\\
&= K'\beta + E(M'GZ'V^{-1}(y - X\hat{\beta})) && (2.22)\\
&= K'\beta + M'E(\alpha) && (2.23)
\end{aligned}
$$

Nepristrasnost se obično definiše da je očekivana vrednost ocenjivača jednaka stvarnoj vrednosti te promenljive. Međutim, ovde to ne možemo koristiti s obzirom da imamo uključene slučajne efekte. To je razlog što u (2.23) imamo $E(\alpha)$. Kada za α imamo da je $E(\alpha) = 0$ onda je taj deo od malog značaja u (2.23).
Možemo zaključiti da je prediktor nepristrasan ako važi

$$E(K'\hat{\beta} + M'\hat{\alpha}) = K'\beta.$$

Do sada smo pretpostavljali da je $Cov(\varepsilon, \alpha) = 0$. Sada ćemo pokazati da u slučaju kada je $Cov(\varepsilon, \alpha) = \sigma^2 k$, gde je k poznati vektor, osobine ocenjivača ostaju nepromenjene.[11] Slučaj kada je kovarijansa između slučajnih grešaka i slučajnih efekata jednaka nuli je specijalan slučaj sledeće teoreme.

Teorema 2.4.1 *Pretpostavimo da je*

$$Var(\varepsilon) = \sigma^2 V,$$

$$Cov(\varepsilon, \alpha) = \sigma^2 k.$$

gde je $\sigma^2 > 0$, V je pozitivno definitna simetrična matrica i k poznati vektor. Pretpostavimo da se $K'\beta + \alpha$ može oceniti. Tada je $K'\hat{\beta} + \hat{\alpha}$ najbolji linearni nepristrasni prediktor za sve prediktore oblika $K'\beta + \alpha$ gde je

$$\hat{\alpha} = Cov(\varepsilon, \alpha)[Var(\varepsilon)]^{-1}(y - X\hat{\beta}) = kV^{-1}(y - X\hat{\beta}).$$

Pri tome je $\hat{\beta}$ BLUE ocenjivač dat sa:

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y$$

Dokaz. *Neka je prediktor oblika:*

$$K'\hat{\beta} + \hat{\alpha} = K'\hat{\beta} + k'V^{-1}(y - X\hat{\beta}).$$

Tada je njegova očekivana vrednost

$$E(K'\hat{\beta} + k'V^{-1}(y - X\hat{\beta})) = K'\beta + k'V^{-1}(X\beta - X\beta) = K'\beta$$

Odavde vidimo da je $K'\hat{\beta} + \hat{\alpha}$ **nepristrasan** *prediktor za $K'\beta + \alpha$.*
$K'\hat{\beta} + \hat{\alpha}$ **linearan** *prediktor što se vidi iz:*

$$
\begin{aligned}
K'\hat{\beta} + \hat{\alpha} &= K'\hat{\beta} + k'V^{-1}(y - X\hat{\beta})\\
&= (K' - k'V^{-1}X)\hat{\beta} + k'V^{-1}y\\
&= (K' - k'V^{-1}X)(XV^{-1}X')^{-1}X'V^{-1}y + k'V^{-1}y\\
&= [(K' - k'V^{-1}X)(XV^{-1}X')^{-1}X'V^{-1} + k'V^{-1}]y\\
&\equiv L'y
\end{aligned}
$$

Kako je $K'\hat{\beta} + \hat{\alpha} = L'y$ *nepristrasan prediktor očigledno je da važi:*

$$L'X = K'.$$

Da bi pokazali da je i **najbolji** *prediktor potrebno je da pokažemo da ima minimalnu varijansu. Neka je* $P'y$ *neki drugi linearan i nepristrasan prediktor oblika* $K'\beta + \alpha$. *Tada važi:*

$$P'X = K'.$$

Srednja kvadratna greška za prediktor $P'y$ *je:*

$$\begin{aligned} MSE(P'y) = E[P'y - (K'\beta + \alpha)]^2 &= Var[P'y - (K'\beta + \alpha)] \\ &= Var(P'y - \alpha) \\ &= Var(P'y - \alpha + L'y - L'y) \\ &= Var[(P-L)'y + L'y - \alpha] \\ &= Var[(P-L)'y] + Var(L'y - \alpha) \\ &\quad +2Cov[(P-L)'y, L'y - \alpha] \end{aligned}$$

Sada je

$$\begin{aligned} Cov[(P-L)'y, L'y - \alpha] &= (P-L)'Var(y)L - Cov[(P-L)'y, \alpha] \\ &= (P-L)'\sigma^2 VL - (P-L)'Cov(y, \alpha) \\ &= \sigma^2 (P-L)'VL - (P-L)'Cov(\varepsilon, \alpha) \\ &= \sigma^2 (P-L)'VL - \sigma^2 (P-L)'k \\ &= \sigma^2 (P-L)'(VL - k). \end{aligned}$$

Dalje,

$$\begin{aligned} VL - k &= V[V^{-1}X[(X'V^{-1}X)^{-1}]'(K - X'V^{-1}k) + V^{-1}k] - k \\ &= X[(X'V^{-1}X)^{-1}]'(K - X'V^{-1}k) \end{aligned}$$

Kako su $P'y$ *i* $L'y$ *nepristrasni prediktori, onda važi:*

$$P'X = L'X = K'.$$

Odavde sledi da je kovarijansa:

$$Cov[(P-L)'y, L'y - \alpha] = \sigma^2 (P-L)'X[(X'V^{-1}X)^{-1}]'(K - X'V^{-1}k) = 0.$$

Dakle, srednja kvadratna greška je:

$$\begin{aligned} MSE(P'y) = E[P'y - (K'\beta + \alpha)]^2 &= Var[(P-L)'y] + Var(L'y - \alpha) \\ &= Var[(P-L)'y] + Var[L'y - (K'\beta + \alpha)] \\ &= Var[(P-L)'y] + E[L'y - (K'\beta + \alpha)]^2 \\ &= Var[(P-L)'y] + MSE(L'y) \end{aligned}$$

Odavde sledi

$$MSE(P'y) \geq MSE(L'y),$$

pa važi:

$$Var[(P-L)'y] = 0 \leftrightarrow P = L.$$

Dakle, $K'\hat{\beta} + \hat{\alpha}$ *je jedinstveni najbolji, linearni, nepristrasni prediktor (BLUP) za sve prediktore oblika* $K'\beta + \alpha$. ∎

2.4.2 Mešoviti model jednačina

Najbolji linearni nepristrasan prediktor u svom izrazu sadrži V^{-1}. Traženje inverzne matrice kada je V velike dimenzije nije praktično. U ovom delu ćemo prikazati drugi način za pronalaženje ovakvog prediktora.
Kao što smo već rekli, Henderson je formulisao mešoviti model jednačina (Mixed Model Equations - MME) za izračunavanje BLUP prediktora za α i BLUE ocenjivača za β.
Mešoviti model jednačina se obično predstavlja u matričnom obliku:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\alpha} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix},$$

odakle dobijamo dve jednačine:

$$X'R^{-1}X\hat{\beta} + X'R^{-1}Z\hat{\alpha} = X'R^{-1}y \tag{2.24}$$

$$Z'R^{-1}X\hat{\beta} + (Z'R^{-1}Z + G^{-1})\hat{\alpha} = Z'R^{-1}y. \tag{2.25}$$

Kada iz jednačine (2.25) izrazimo $\hat{\alpha}$ i ubacimo u jednačinu (2.24) dobijamo:

$$X'R^{-1}X\hat{\beta} + X'R^{-1}ZWZ'R^{-1}(y - X\hat{\beta}) = X'R^{-1}y,$$

gde je

$$W = (Z'R^{-1}Z + G^{-1})^{-1}.$$

Dalje,

$$\begin{aligned} X'R^{-1}X\hat{\beta} - X'R^{-1}ZWZ'R^{-1}X\hat{\beta} &= X'R^{-1}y - X'R^{-1}ZWZ'R^{-1}y \\ X'(R^{-1} - R^{-1}ZWZ'R^{-1})X\hat{\beta} &= X'(R^{-1} - R^{-1}ZWZ'R^{-1})y \\ X'V^{-1}X\hat{\beta} &= X'V^{-1}y \end{aligned}$$

gde je

$$V^{-1} = R^{-1} - R^{-1}ZWZ'R^{-1}.$$

Ovako definisano V^{-1} je zaista inverzna matrica od V što se vidi iz:

$$\begin{aligned} VV^{-1} &= (R + ZGZ')(R^{-1} - R^{-1}ZWZ'R^{-1}) \\ &= I + ZGZ'R^{-1} - ZWZ'R^{-1} - ZGZ'R^{-1}ZWZ'R^{-1} \\ &= I + ZGZ'R^{-1} - Z(I + GZ'R^{-1}Z)WZ'R^{-1} \\ &= I + ZGZ'R^{-1} - ZG(G^{-1} + Z'R^{-1}Z)WZ'R^{-1} \\ &= I + ZGZ'R^{-1} - ZGW^{-1}WZ'R^{-1} \\ &= I + ZGZ'R^{-1} - ZGZ'R^{-1} \\ &= I \end{aligned}$$

Ocenjivač za β dobijen pomoću MME jednak je onom kojeg smo dobili i ranije. Takođe, prediktor za slučajne efekte se poklapa.
Ako u prediktor $\hat{\alpha}$:

$$\hat{\alpha} = GZ'V^{-1}(y - X\hat{\beta}),$$

dobijen BLUP metodom, V^{-1} zamenimo sa $R^{-1} - R^{-1}ZWZ'R^{-1}$, dobijamo:

$$
\begin{aligned}
GZ'V^{-1}(y - X\hat{\beta}) &= GZ'(R^{-1} - R^{-1}ZWZ'R^{-1})(y - X\hat{\beta}) \\
&= G(Z'R^{-1} - Z'R^{-1}ZWZ'R^{-1})(y - X\hat{\beta}) \\
&= G(I - Z'R^{-1}ZW)Z'R^{-1}(y - X\hat{\beta}) \\
&= G(W^{-1} - Z'R^{-1}Z)WZ'R^{-1}(y - X\hat{\beta}) \\
&= G((Z'RZ + G^{-1}) - Z'R^{-1}Z)WZ'R^{-1}(y - X\hat{\beta}) \\
&= (GZ'RZ + I - GZ'R^{-1}Z)WZ'R^{-1}(y - X\hat{b}) \\
&= IWZ'R^{-1}(y - X\hat{\beta}) \\
&= WZ'R^{-1}(y - X\hat{\beta}) \\
&= \hat{\alpha}
\end{aligned}
$$

Dakle, najbolji linearni nepristrasan prediktor za $K'\beta + \alpha$ je $K'\hat{\beta} + \hat{\alpha}$, gde $\hat{\beta}$ i $\hat{\alpha}$ možemo dobiti pomoću MME.

2.4.3 Empirijski BLUP

U praksi se često dešava da su fiksni efekti kao i varijanse komponenti nepoznati. U tom slučaju nije moguće izračunati BLUP ocenu, odnosno potrebno je oceniti te varijanse. Dakle, vektor varijansi θ je potrebno zameniti odgovarajućim ocenjivačem $\hat{\theta}$. Na taj način dobijamo prediktor koji se najčešće naziva EBLUP, gde oznaka E potiče od *Empirical.*
Kackar i *Harville*(1981) su pokazali da takav prediktor ostaje nepristrasan,[4] pri čemu su pored normalnosti podataka koristili i sledeće osobine ocenjivača $\hat{\theta} = \hat{\theta}(y)$:

$$\hat{\theta}(-y) = \hat{\theta}(y),$$

$$\hat{\theta}(y - X\beta) = \hat{\theta}(y).$$

Druga osobina se naziva translatorna invarijantnost ocenjivača $\hat{\theta}$. Postoji više translatorno invarijantnih ocenjivača za θ, od kojih su neki od najpoznatijih ANOVA, ML i REML. Pored toga pretpostavili su da postoji očekivana vrednost ocene EBLUP koja i nije baš očigledna kako EBLUP nije linearan ocenjivač. Kasnije je postojanje te očekivane vrednosti dokazao Jiming Jiang(1999b,2000a).[4]

Pored EBLUP ocene postoji i empirijska Bayes-ova ocena - PEB. Međutim, Harville je pokazao da se u slučaju mešovitih efekata oni poklapaju.[4] Jedna od osnovnih razlika između ta dva pristupa jeste što je PEB više teorijske prirode, odnosno rad je fokusiran na jednostavnije modele jer su kao takvi pogodniji sa teorijskog stanovišta. S druge strane, rad vezan za EBLUP se uglavnom vrši od strane određenih specijalizovanih stručnjaka i stoga se primenjuje na složenije modele.

Ocenu srednje kvadratne greške MSE za EBLUP takođe dobijamo zamenom θ sa njegovom ocenom $\hat{\theta}$. Takvu ocenu za MSE nazivamo naivna ocena jer kao takva potcenjuje ukupnu grešku. Kako bi to ilustrovali sa $\tilde{\eta}$ označićemo EBLUP mešovitih efekata. Dakle, neka je

$$\tilde{\eta} = K'\tilde{\alpha} + M'\tilde{\beta}$$

EBLUP ocena mešovitih efekata za $\eta = K'\alpha + M'\beta$, gde su $\tilde{\alpha}$ i $\tilde{\beta}$ BLUP i BLUE ocene za α i β i pri tome je vektor varijansi θ zamenjen ocenom $\hat{\theta}$. Kackar i Harville su pokazali da pod pretpostavkom normalnosti važi:

$$MSE(\tilde{\eta}) = MSE(\hat{\eta}) + E(\tilde{\eta} - \hat{\eta})^2,$$

gde je $\hat{\eta}$ BLUP ocena za η data ranije u izrazu (2.18).
Može se pokazati[4] da je:

$$MSE(\tilde{\eta}) = g_1(\theta) + g_2(\theta),$$

gde su:

$$g_1(\theta) = K'(G - GZ'V^{-1}ZG)K,$$
$$g_2(\theta) = (M - X'V^{-1}ZGK)'(X'V^{-1}X)'(M - X'V^{-1}ZGK).$$

Jasno je da ukoliko umesto θ koristimo ocenjivač $\hat{\theta}$ možemo potceniti srednje kvadratnu grešku jer se ne uzima u obzir varijansa koja nastaje prilikom procene θ. Ovaj problem je od velikog značaja kada su u pitanju velika ulaganja. Nije retkost da se na osnovu EBLUP-a ulažu velika novčana sredstva. Kackar i Harville(1984) su došli do približne vrednosti srednje kvadratne greške EBLUP ocene za linearne mešovite modele pri čemu su izveli varijabilnost ocene $\hat{\theta}$. Tačnost aproksimacije greške do danas nije pokazana. Prasad i Rao(1990) su izučavali tačnost drugog reda za dva specijalna slučaja longitudinalnih mešovitih modela.[4] Kasnije su rezultati prošireni na generalne mešovite modele (2.1).

Za Gausov mešoviti model ANOVA sa REML ocenom za θ je pokazano da:

$$MSE(\tilde{\eta}) = g_1(\theta) + g_2(\theta) + g_3(\theta) + o(d_*^{-2}),$$

gde je:

$$g_3(\theta) = tr[\{(\partial/\partial\theta')V^{-1}ZGM\}'V\{(\partial/\partial\theta')V^{-1}ZGM\}H^{-1}],$$
$$H = E(\partial l_R/\partial\theta\partial\theta')$$

Pri tome je l_R funkcija data ranije u (2.6) i:

$$d_* = min_{1\leq i\leq s} d_i \qquad d_i = ||Z_i'PZ_i||_2,$$
$$P = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}XV^{-1}.$$

Takođe isti rezultat se dobija i za ML ocenu.

Jiming Jiang(2002) je predložio metod jackknife koji je doveo do drugog reda aproksimacije srednje kvadratne greške za EBLUP u slučaju longitudinalnih linearnih mešovitih modela. Označimo sada, radi jednostavnosti zapisa $MSE(\hat{\eta})$ sa $b(\theta)$, gde je $\hat{\eta}$ BLUP ocena. Jackknife ocenjivač za $MSE(\tilde{\eta})$ je dat sa:

$$\widehat{MSE}(\tilde{\eta}) = \widehat{MSAE}(\tilde{\eta}) + \widehat{MSE}(\hat{\eta}),$$

gde je

$$\widehat{MSAE}(\tilde{\eta}) = \frac{m-1}{m}\sum_{i=1}^{m}(\tilde{\eta}_{-i} - \tilde{\eta})^2,$$
$$\widehat{MSE}(\tilde{\eta}) = b(\hat{\theta}) - \frac{m-1}{m}\sum_{i=1}^{m}(b(\hat{\theta}_{-i}) - b(\hat{\theta})).$$

Pri tome m predstavlja broj klastera, dok je sa $\hat{\theta}_{-i}$ označen ocenjivač za θ dobijen na podacima bez i-tog klastera. Analogno, $\tilde{\eta}$ predstavlja EBLUP za η gde su fiksni parametri ocenjeni bez podataka iz i-tog klastera.
Pored toga, važi:

$$E(\widehat{MSE}(\tilde{\eta})) = MSE(\tilde{\eta}) + o(m^{-1}).$$

Ovaj rezultat se koristi kada je ocenjivač za θ REML ili ML, za Gausove mešovite linearne modele longitudinalnog tipa. Jackknife metod se najčešće primenjuje na generalne mešovite linearne modele gde se ocena EBLUP tada zamenjuje ocenom EBP (*Empirical Bayes prediction*).

Glava 3

Primena linearnih mešovitih modela u selekciji životinja

3.1 Selekcija u ovčarstvu pomoću linearnih mešovitih modela

Kako bi ilustrovali linearne mešovite modele Harville i Fenech(1985)[4] su analizirali telesnu masu jagnjadi na rođenju. Prikupili su podatke o masi 62 pojedinačnih jagnjadi. Jagnjad su potomci od 23 oca (ovna) i svako jagnje ima različitu majku. Starost majke je evidentirana kao kovarijetet, dok drugi kovarijetet predstavlja populaciona linija. U ovom slučaju postoje dve kontrolne linije i tri selekcione. Jiming Jiang[4] je nešto drugačije grupisao prikupljene podatke kako bi bolje predstavio mešovite linearne modele, što ćemo i predstaviti u nastavku.

Podaci su prikazani u tabeli 3.1. U ovom modelu, otac odnosno ovan predstavlja slučajni faktor, tj. slučajni efekat. Slučajni efekti očeva su označeni sa s_{ij} (s - od *sire* što znači otac), $i = 1, 2, ..., 5$ i $j = 1, 2, ..., n_i$, gde su $n_1 = n_2 = n_3 = 4$, $n_4 = 3$ i $n_5 = 8$. s_{ij} su nezavisni i normalno distribuirani sa sredinom 0 i varijansom σ_s^2. Starost majke predstavlja kategorijalnu promenljivu. Majke su podeljene u tri starosne grupe: **1** (1-2 godine), **2** (2-3 godine) i **3** (više od 3 godine), te je starost majke posmatrana kao fiksna promenljiva.

Dalje, neka je $x_{ijk,1} = 1$ ako godine k-te majke koja odgovara liniji i i ocu j pripadaju kategoriji 1, a u drugom slučaju neka je $x_{ijk,1} = 0$. Analogno, $x_{ijk,2} = 1$ ako godine k-te majke koja odgovara liniji i i ocu j pripadaju kategoriji 2, a u drugom slučaju važi $x_{ijk,1} = 0$. Ostali fiksni efekti su linijski i njih označavamo sa $l_i, i = 1, 2, ..., 5$. Konačno, slučajne greške e_{ijk}, $i = 1, 2, ...5$, $j = 1, 2, ...n_i$ i $k = 1, 2, ..., n_{ij}$ su dodate u model kako bi objasnile varijaciju koja nastaje usled uticaja životne sredine i drugih neobjašnjenih faktora. e_{ijk} su po pretpostavci nezavisne i normalno distribuirane sa sredinom 0 i varijansom σ_e^2. Takođe su i e_{ijk} i s_{ij} međusobno nezavisni. n_{ij} predstavlja broj merenja u ćeliji (i, j). Na primer $n_{11} = 1$, $n_{13} = 6$, $n_{42} = 2$.

Sada mešoviti linearni model se može predstaviti kao:

$$y_{ijk} = l_i + a_1 x_{ijk,1} + a_2 x_{ijk,2} + s_{ij} + e_{ijk},$$

gde $i = 1, 2, ..., 5$, $j = 1, 2, ...n_i$ i $k = 1, 2, ..., n_{ij}$. Ovaj model se može prikazati u standardnoj formi modela (2.1)

$$y = X\beta + Zs + e,$$

gde su $y = [y_{111}, y_{121}, ..., y_{585}]'$ vektor posmatranja, $\beta = [l_1, l_2, ..., l_5, a_1, a_2]'$ vektor fiksnih efekata, X matrica kovarijeteta koji odgovaraju β, $s = [s_{11}, s_{12}, ..., s_{58}]'$ vektor koji odgovara uticaju oca, Z dizajn matrica za s i $e = [e_{111}, e_{121}, ..., e_{585}]$ vektor slučajnih grešaka. Matrica Z je standardna dizajn matrica koja se sastoji od nula i jedinica, gde u svakoj vrsti postoji tačno jedna jedinici i u svakoj koloni najmanje jedna.
Model je fitovan pomoću *SAS PROC MIXED* sa opcijom REML.[4] Opcija REML daje ocenu varijanse komponenti. Na taj način[4] se dobija:

$$\hat{\sigma_s^2} = 0.511,$$

$$\hat{\sigma_e^2} = 2.996.$$

Pre svega, cilj je odrediti fiksne efekte. Ocene fiksnih efekata su date u tabeli 3.2.[4] Vidimo da se svi linijski efekti značajno razlikuju od 0, pri čemu je kao nivo značajnosti uzeta vrednost od 0.01. Međutim, efekti starosti nisu značajni. Dakle, možemo zaključiti da starost majke ne utiče značajno na telesnu masu jagnjadi na rođenju.
SAS PROC MIXED[4] daje i alternativne metode za fitovanje linearnih mešovitih modela. Npr. ako se REML opcija zameni sa opcijom MIVQUE0 (specijalan slučaj MINQUE) kao ocene varijansi[4] dobijaju se:

$$\hat{\sigma_s^2} = 0.323,$$

$$\hat{\sigma_e^2} = 3.116.$$

Koristeći ovaj metod dobijaju se slične ocene za fiksne efekte, dok se ocene varijansi sasvim razlikuju.[4]
Međutim, *PROC MIXED* ne daje mogućnost ANOVA ocene parametara.[4] Jedina mogućnost da se dođe do takve ocene je pomoću GLM procedure u SAS-u, PROC GLM.[4] One koriste Hendersonov metod III i na taj način se dobijaju ocene varijansi:

$$\hat{\sigma_s^2} = 0.764,$$

$$\hat{\sigma_e^2} = 2.796.$$

Ponovo su dobijene različite ocene u odnosu na REML i MIVQUE0.

Posmatranje	6.2	13.0	9.5	10.1	11.4	11.8	12.9	13.1	10.4
Otac	11	12	13	13	13	13	13	13	14
Linija	1	1	1	1	1	1	1	1	1
God.	1	1	1	1	1	2	3	3	1
Posmatranje	8.5	13.5	10.1	11.0	14.0	15.5	12.0	11.5	10.8
Otac	14	21	22	22	22	22	23	24	24
Linija	1	2	2	2	2	2	2	2	2
God.	2	3	2	3	3	3	1	1	3
Posmatranje	9.0	9.0	12.6	11.0	10.1	11.7	8.5	8.8	9.9
Otac	31	31	31	32	32	32	32	32	32
Linija	3	3	3	3	3	3	3	3	3
God.	2	3	3	1	2	2	3	3	3
Posmatranje	10.9	11.0	13.9	11.6	13.0	12.0	9.2	10.6	10.6
Otac	32	32	32	33	33	34	41	41	41
Linija	3	3	3	3	3	3	4	4	4
God.	3	3	3	1	3	2	1	1	1
Posmatranje	7.7	10.0	11.2	10.2	10.9	11.7	9.9	11.7	12.6
Otac	41	41	41	42	42	43	43	51	51
Linija	4	4	4	4	4	4	4	5	5
God.	3	3	3	1	1	1	3	1	1
Posmatranje	9.0	11.0	9.0	12.0	9.9	13.5	10.9	5.9	10.0
Otac	52	52	53	53	54	55	56	56	57
Linija	5	5	5	5	5	5	5	5	5
God.	1	3	3	3	3	2	2	3	2
Posmatranje	12.7	13.2	13.3	10.7	11.0	12.5	9.0	10.2	
Otac	57	57	57	58	58	58	58	58	
Linija	5	5	5	5	5	5	5	5	
God.	2	3	3	1	1	1	3	3	

Tabela 3.1: Telesna masa jagnjadi po rođenju[4]

Effect	Line	Age	Est.	S.E	t-value	$Pr > \|t\|$
Line	1		10.5008	0.8070	13.01	< .0001
Line	2		12.2998	0.7569	16.25	<.0001
Line	3		11.0425	0.6562	16.83	<.0001
Line	4		10.2864	0.7882	13.05	<.0001
Line	5		10.9625	0.5438	20.16	<.0001
Age		1	−0.0097	0.5481	−0.02	0.9861
Age		2	−0.1651	0.6435	−0.26	0.7989

Tabela 3.2: Ocene fiksnih efekata(REML)[4]

Effect	Line	Age	Est.	S.E	t-value	$Pr > \|t\|$
Line	1		10.5637	0.7730	13.67	$<$ .0001
Line	2		12.3028	0.7277	16.91	$<$.0001
Line	3		10.9962	0.6134	17.93	$<$.0001
Line	4		10.2640	0.7478	13.72	$<$.0001
Line	5		10.9650	0.5255	20.86	$<$.0001
Age		1	-0.0109	0.5509	-0.02	0.9844
Age		2	-0.1393	0.6495	-0.21	0.8313

Tabela 3.3: Ocene fiksnih efekata(MIQUE0)[4]

Dakle, do sada su ocenjeni fiksni efekti i dobijene su ocene varijansi komponenti. Dalje je cilj oceniti slučajne efekte, tj. u ovom slučaju su to efekti koji potiču od oca. Potrebni su nam najbolji linearani nepristrasani prediktori, odnosno BLUP ocene. Ti prediktori su dobijeni pomoću *PROC MIXED*.[4] Rezultat su prikazani u tabeli 3.4. Standard Pred Error predstavlja kvadratni koren ocene srednje kvadratne predviđenje greške, MSPE (*Mean Square Prediction Error*) od EBLUP ocene za s_{ij}, j-tog oca iz linije i. Ocena MSPE se u *PROC MIXED* dobija ubacivanjem REML ocena varijansi u formulu za MSPE, pretpostavljajući da su kao takve varijanse komponenti poznate.[4]
Ovo je poznato kao naivna metoda ocenjivanja greške MSPE. Kao što smo i pre diskutovali ova metoda može potceniti grešku MSPE. Metode koje poboljšavaju tačnost ove ocene su predloženi ranije.

Sire	Line	Estimate	Standard Pred Error
11	1	−0.6256	0.6693
12	1	0.3658	0.6693
13	1	0.5050	0.6156
14	1	−0.2452	0.6441
21	2	0.1750	0.6701
22	2	0.1588	0.6296
23	2	−0.0423	0.6717
24	2	−0.2914	0.6457
31	3	−0.2667	0.6184
32	3	−0.2182	0.5850
33	3	0.3212	0.6397
34	3	0.1637	0.6701
41	4	−0.2015	0.6187
42	4	0.0695	0.6454
43	4	0.1319	0.6436
51	5	0.3047	0.6356
52	5	−0.2437	0.6308
53	5	−0.1177	0.6327
54	5	−0.1549	0.6656
55	5	0.3940	0.6684
56	5	−0.6311	0.6318
57	5	0.5762	0.5913
58	5	−0.1274	0.5769

Tabela 3.4: BLUP ocene slučajnih efekata[4]

3.2 Selekcija u govedarstvu pomoću linearnih mešovitih modela

U oplemenjivanju životinja je, kao što smo već rekli, jedan od osnovnih ciljeva uspešna selekcija, odnosno potrebno je odabrati najbolju životinju koja će na potomstvo preneti svoj genetski potencijal. Odabir životinja za dalju reprodukciju se u svetu najčešce vrši na osnovu oplemenjivačkih vrednosti koje se izračunavaju pomoću najboljeg linearnog nepristrasnog prediktora, tj. BLUP ocene.

2011. godine izvršeno je jedno istraživanje od strane Poljoprivrednog fakulteta Univerziteta Novi Sad i Instituta za stočarstvo u Zemunu.[7] Osnova tog istraživanja je bila odabir bikova očeva koji će se koristiti za dalji priplod. Od interesa su bile osobine mlečnosti krava koje potiču od bikova simentalske rase u R. Srbiji. Istraživanje obuhvata 2121 prvotelku, simentalske rase, koje su locirane na imanjima individualnih poljoprivrednih proizvođača na teritoriji Srbije. Laktacije su sakupljene u toku jedne godine. Analiza podataka je zasnovana na principu najmanjih kvadrata za mešovite modele. Cilj je proceniti oplemenjivačke vrednosti bikova na osnovu osobina mlečnosti njihovih kćerki, a pri tome eliminisati faktore životne sredine. Ocene su vršene za sledeće osobine: trajanje laktacije (DL), prinos mleka (MY), sadržaj mlečne masti (CMF), prinos mlečne masti (MFY) i količinu mleka (kg/dan) korigovanu na 4% mlečne masti (4% FCM). Jednačina modela sadrži slučajne efekte koji potiču od oca i fiksne efekte: uticaj regiona, godinu i sezonu teljenja. Svaki bik (otac) ima 20 i više kćeri rasprostranjenih u 2 do 3 regiona. Dakle, kao što je ranije navedeno za ocenu oplemenjivačkih vrednosti bikova i samo njihovo rangiranje na osnovu proizvodnje njihovih kćeri, korištena je BLUP metoda.[7] Na taj način dobijeni su rezultati koji nam pokazuju znatnu superiornost nekih bikova u odnosu na ostale.

Jednačina mešovitog linerarnog modela[7] izgleda:

$$y_{ijklm} = \mu + B_i + R_j + G_k + S_l + \varepsilon_{ijklm}$$

gde su:
y_{ijklm} osobina koja je merena na m-toj kravi, kćerki oca i, proizvedenoj u j-toj regiji i oteljenoj u k-toj godini i l-toj sezoni;
μ generalni prosek, odnosno sredina;
B_i slučajni efekat i-tog bika(oca);
R_j fiksni efekat koji predstavlja uticaj j-tog regiona;
G_k fiksni efekat k-te godine teljenja;
S_l fiksni efekat l-te sezone teljenja;
ε_{ijklm} slučajna greška.

Rezultati koji su dobijeni za oplemenjivačke vrednosti i rangiranje bikova na osnovu osobina mlečnosti kćerki pokazuju da su, kada je u pitanju prinos mleka, najbolje rangirani strani bikovi koji su pozajmljeni iz drugih zemalja ili oni čije je seme uvezeno za plansko osemenjavanje domaćih plotkinja simentalske rase. Neki od ovih bikova imaju u svojoj genetskoj osnovi udela drugih rasa poput Crvenog Holštajna. S druge strane, domaći bikovi su najlošije rangirani. Bolji rezultati za domaće bikove nisu postignuti ni za ostale osobine.

Proizvodnja mleka i mlečne masti vezana je isključivo za jedan pol. Međutim znanje o kvalitetu bikova može biti od velikog značaja za poboljšanje populacije. Osnovni zahtev, kada je u pitanju oplemenjivanje životinja jeste poboljšanje njihovog potencijala za proizvodnju kako mleka tako i mesa. Linearni modeli doprinose genetskom poboljšanju proizvodnih i reproduktivnih osobina, pre svega kroz dobijanje tačnijih procena genetskih vrednosti grla. Primena BLUP metoda je jedna od vodećih kada je u pitanju genetski napredak populacije i kao takav primenjuje se širom sveta.[7] Prednost mu se daje u odnosu na ostale metode, jer je kao takav najmanje pristrasan.

3.3 Selekcija u svinjarstvu pomoću linearnih mešovitih modela

U ovom delu biće izvršena samostalna obrada prikupljenih podataka. Pre svega napomenimo da se dobar kvalitet svinje odnosi na visok nivo mišićnog tkiva u određenim delovima poput buta, leđa, vrata i slično. Ocenjivanje mesnatosti, koja je od velikog interesa u svinjarstvu, vrši se na osnovu sastava i nekih karakteristika koji su od komercijalnog značaja. Balansirana ishrana predstavlja dobar put do sigurnog unapređenja određenih osobina kod svinja, ali je pored toga, neizostavni faktor koji utiče na posmatranu osobinu i genetska predispozicija. U ovom slučaju izabrali smo da posmatramo koeficijent koji predstavlja odnos bočne slanine i telesne mase pomnožen sa 100 radi lakše analize. Kao i do sada, glavni cilj je izvršiti selekciju svinja očeva odnosno majki koje će se koristiti za dalji priplod, tj. potrebno je odabrati one životinje koje imaju najveću mesnatost, a kako je debljina slanine u negativnoj korelaciji sa količinom mesa onda samim tim i najmanji pomenuti koeficijent. Selekciju odnosno odabir najboljih životinja radićemo pomoću najboljeg linearnog nepristrasnog prediktora, tj. BLUP ocene.

Podaci su prikupljeni od strane Poljoprivrednog fakulteta Novi Sad i prilagođeni radi lakše ilustracije linearnih mešovitih modela. Podaci sadrže informacije o ukupno 562 svinje. Na slici 3.3 ispod izdvojeni su podaci za prvih deset svinja. Prva kolona predstavlja šifru životinje, odnosno grla (ID), druga pol (sex), treća farmu na kojoj se odgovarajuće grlo uzgaja (F) i četvrta rasu grla (Ge). Potom slede šifre oca (Sire), odnosno majke (Dam), a na samom kraju izmereni koeficijent bočne slanine za svako grlo. Analizom je obuhvaćeno 5 različitih rasa svinja na kojima je merena debljina bočne slanine i koje se uzgajaju na 22 farme u Srbiji. Grla potiču od 102 oca i 336 majki. Treba napomenuti da su sva posmatrana grla muškog pola, te stoga osobina pol nije uzeta u dalju analizu.

ID	sex	F	Ge	Sire	Dam	k
20206	Male	5	2	2276	2920	20.86
23216	Male	31	4	6320	6213	15.00
25616	Male	31	4	6320	2652	15.74
27849	Male	31	4	6320	9990	12.63
28184	Male	31	4	6320	6213	8.69
29177	Male	31	4	6320	12259	9.90
37749	Male	31	1	3908	10006	4.90
38418	Male	31	1	3908	15207	20.00
40097	Male	31	1	3908	8737	10.41
40198	Male	31	1	3908	14593	20.00

Tabela 3.5: Koeficijent bočne slanine

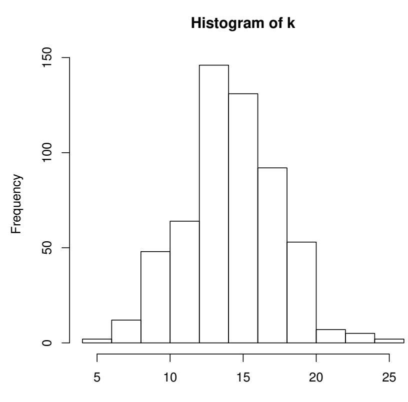
Jasno je da, kao i do sada konstruišemo linearni mešoviti model, gde otac, majka ili oboje predstavljaju slučajne efekte, ostale promenljive su fiksne, a pri tome zavisna promenljiva je koeficijent k. Cilj analize jeste oceniti slučajne parametre.

Za analizu smo koristili poznati statistički paket R jer u njemu postoji mogućnost ocene parametara za linearne mešovite modele. Ocene su dobijene pomoću funkcije **lmer** koja je ugrađena u paket **lme4**. Razlog za takav odabir je što prethodno pomenuta funkcija koristi metod limitirane maksimalne verodostojnosti (REML) (detaljnije u poglavlju 2.2). Kako REML ocena zahteva normalnu distribuciju pre toga smo ovu osobinu i ispitali na našim podacima.

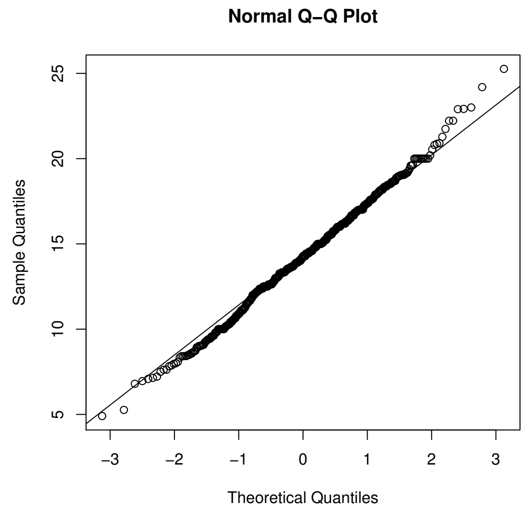
Poznato je da normalnu distribuciju rezultata možemo ispitati na dva načina i to pomoću grafičke i numeričke metode. Numeričke metode provere imaju prednost jer ne uzimaju u obzir subjektivno mišljenje istraživača. S druge strane mogu biti nedovoljno osetljive na male uzorke, te previše osetljive na velike. Ovo znači da u slučaju malih uzoraka testovi mogu pokazivati da se radi o normalnoj distribuciji iako to zaista nije slučaj, dok u slučaju velikih uzoraka mogu pokazivati da se ne radi o normalnoj distribuciji i onda kada je ona normalna. Iz tog razloga prvo smo obavili proveru pomoću grafičke metode, a potom i numeričke.

Na prva tri grafika se može uvideti da podaci prate normalnu distribuciju. Grafička metoda provere normalnosti je odrađena u statističkom paketu R pomoću sledećih funkcija:

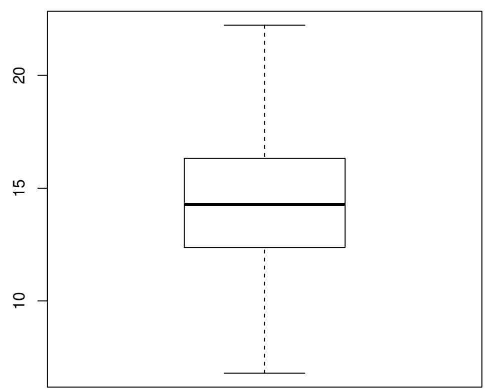
```
> hist(k)
> qqnorm(k)
> boxplot(k)
```



Slika 3.1: Histogram koeficijenta k.



Slika 3.2: Q-Q plot za koeficijent k.



Slika 3.3: Boxplot za koeficijent k.

Pored grafičke metode, kao što je već prethodno spomenuto, odrađeni su i statistički testovi normalnosti. Na narednoj slici prikazan je rezultat iz paketa SPSS (Trial Version) radi lepše preglednosti, pri tome su identični reziltati dobijeni i u paketu R pomoću funkcija:

```
>  shapiro.test(k)
> ks.test(k,pnorm,mean(k),sd(k))
```

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
k	.031	562	.200*	.997	562	.543

a. Lilliefors Significance Correction
*. This is a lower bound of the true significance.

Slika 3.4: Testovi normalnosti za koeficijent k.

Nulta hipoteza u ovim testovima govori da podaci prate normalnu distribuciju, dok alternativna podrazumeva da postoje značajne razlike između distribucija. Testiranje je vršeno na nivou poverenja od 95%. p vrednosti koje smo dobili su veće od 0.05 i pri tome prihvatamo nultu hipotezu, tj. smatra se da su podaci normalno distribuirani.

Pored toga, korištena je još jedna numerička metoda za ispitivanje mere asimetričnosti i homogenosti (zarubljenosti). Ovu metodu smo uradili pomoću koeficijenata *skewness* i *kurtosis* u statističkom paketu R koristeći istoimene funkcije. Dobijeni su sledeći rezultati:

```
> skewness(k)
[1] 0.07689206

> kurtosis(k)
[1] 3.07593
```

Kao što se može videti skewness je blizu nule, a kurtosis je blizu vrednosti 3, te stoga pretpostavka o normalnosti podataka može biti prihvaćena.

Sada kada imamo normalnost možemo preći na analizu našeg linearnog mešovitog modela. Pre svega, treba napomenuti da je za analizu u paketu R bilo potrebno učitati nekoliko paketa poput *lme4*, *devtools*, *moments*.

Prvi model koji je rađen uzima farmu i rasu svinje kao fiksne faktore, dok je otac (Sire) uzet kao slučajan faktor. Model pravimo pomoću funkcije *lmer* na sledeći način:

```
> ML<-lmer(k~F+Ge+(1|Sire),data=data1)
> summary(ML)
Linear mixed model fit by REML ['lmerMod']
Formula: k ~ F + Ge + (1 | Sire)
   Data: data1

REML criterion at convergence: 2741.7

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.0373 -0.5472 -0.0125  0.5785  4.2926

Random effects:
 Groups   Name        Variance Std.Dev.
 Sire     (Intercept) 5.405    2.325
 Residual             5.765    2.401
Number of obs: 562, groups:  Sire, 102

Fixed effects:
             Estimate Std. Error t value
(Intercept) 14.232080   0.478105  29.768
F           -0.012322   0.004782  -2.577
Ge           0.138836   0.084117   1.651

Correlation of Fixed Effects:
   (Intr) F
F  -0.547
Ge -0.635  0.039
>
```

Oznaka $(1|Sire)$ označava da se radi o jednostavnijem linearnom mešovitom modelu kakav smo mi do sada i radili. Umesto jedinice mogu da stoje i druge oznake koje znatno komplikuju model. Vidimo da smo na ovaj način uspeli da konstruišemo model u kojem slučajni efekat proizilazi od oca. Međutim, sada nas intresuju i koeficijenti za svakog oca pojedinačno. Dakle interesuju nas BLUP ocene koeficijenata. To možemo dobiti pomoću:

```
> coef(ML)
```

Na ovaj način dobijamo 102 koeficijenta koja nam govore o kvalitetu očeva. Nas, kao što smo i rekli, interesuju oni koji imaju manji koeficijent k, jer nam to ukazuje na veću mesnatost kod njihovih potomaka, a samim tim i na kvalitet genetskog materijala kod očeva. Na slici 3.6 izdvojeno je prvih deset očeva sa njihovim koeficijentima, dok je na slici 3.7 izdvojeno deset očeva sa najlošijim genetskim materijalom, po ovom modelu.

Sire	BLUP
5358	9.47
40377	10.14
1277	10.32
115039	10.54
78911	10.86
16670	10.95
6822	10.97
71132	11.28
16412	11.31
52323	11.34

Tabela 3.6: BLUP ocene za očeve(1-10)

Sire	BLUP
70879	16.46
11913	16.48
114770	16.55
40096	16.80
107474	16.89
6714	17.00
2276	17.33
51037	17.35
3850	18.93
13626	19.21

Tabela 3.7: BLUP ocene za očeve(93-102)

Možemo videti da otac 5358 ima, po ovom modelu, najmanju BLUP ocenu, tj. najbolju (najveću) oplemenjivačku vrednost, što ukazuje da njegovi potomci imaju u proseku najveću mesnatost, kada je u pitanju bočna slanina. Kako se koeficijent k kreće okvirno od 5 do 25, kao što možemo videti i na histogramu na slici 3.1, koefcijenti k u vrednosti od 8, 6.79, 7.82, 10.09, 10,57, njegove dece su sasvim zadovoljavajući. S druge strane, otac 13626 ima najgori rezultat, tj. ima najveću BLUP ocenu. Ako proverimo vrednosti k njegovih potomaka vidimo da su vrednosti od 16.19, 20.8, 22.9 izrazito velike. Jasno je da se najbolji otac ne određuje sa najboljim prosekom rezultata njegove dece, već na to utiču i ostali faktori iz modela. Ipak, ove vrednosti su date da bi se stekla jasnija slika o razlici najboljeg i najlošijeg oca po ovom modelu.

Pored ovog modela, konstruisan je još jedan u koji je, pored oca kao slučajnog faktora, uključena i majka. Postupak je rađen analogno, a dobijeni su rezultati koji se mogu videti u nastavku.

```
> ML1<-lmer(k~F+Ge+(1|Sire)+(1|Dam),data=data1)
> summary(ML1)
Linear mixed model fit by REML ['lmerMod']
Formula: k ~ F + Ge + (1 | Sire) + (1 | Dam)
   Data: data1

REML criterion at convergence: 2735

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.3896 -0.4953 -0.0164  0.5118  4.7220

Random effects:
 Groups   Name        Variance Std.Dev.
 Dam      (Intercept) 1.210    1.100
 Sire     (Intercept) 4.959    2.227
 Residual             4.818    2.195
Number of obs: 562, groups:  Dam, 336; Sire, 102

Fixed effects:
             Estimate Std. Error t value
(Intercept) 14.228969   0.475873  29.901
F           -0.012049   0.004823  -2.498
Ge           0.131121   0.083928   1.562

Correlation of Fixed Effects:
   (Intr) F
F  -0.554
Ge -0.634  0.042
```

Sire	BLUP
5358	9.95
1277	10.52
40377	10.61
115039	10.70
16670	10.91
78911	11.13
6822	11.19
71132	11.42
16237	11.45
16412	11.50

Tabela 3.8: BLUP ocene za očeve(1-10)

Sire	BLUP
11913	16.38
70879	16.32
114770	16.41
40096	16.50
107474	16.75
6714	16.82
2276	17.13
51037	17.15
3850	18.53
13626	18.65

Tabela 3.9: BLUP ocene za očeve(93-102)

Ponovo smo prikazali rezultate samo za očeve, jer se u ovim modelima veći značaj pridaje muškim grlima, tj. očevima na nivou populacije, zbog same prirode majki, odnosno kolčine potrebnog vremena dobijanja potomaka. Drugim rečima očekuje se da će muška grla imati više potomaka. Uključili smo uticaj majke, kako bi stekli jasniju sliku o očevima i uporedili sa prethodnim modelom. Vidimo da ne postoje značajne razlike u ocenama, te da je ponovo otac 5358 rangiran najviše. Međutim, na nižim pozicijama postoje neke izmene, kao što je na primer slučaj sa ocem 1277, koji je u prvom modelu bio na trećoj poziciji, a sada je na drugoj.

Dakle, po oba ova modela bi izabrali oca 5358 za dalju reprodukciju. Pored toga treba imati u vidu i fiksne faktore koji su uticali da se ispolje određene karakteristike kod njegovih potomaka, u ovom slučaju naš posmatrani koeficijent k.

Treba naglasiti da su ovi modeli rađeni u cilju ilustracije linearnih mešovitih modela i BLUP ocene, te da bi u slučaju stvarne selekcije trebalo uključiti više faktora u model. Pored toga, za dalju analizu i uspešniju selekciju, u smislu poboljšanja tačnosti samog modela, bilo bi neophodno veće znanje iz oblasti genetike, oplemenjivanja i ishrane svinja kako bi se model unapredio.

Prilog

U ovom delu rada prikazani su podaci koji su korišćeni za analizu u cilju selekcija svinja nerastova, pri čemu je koeficijent *k* dobijen na način spomenut ranije, pomoću debljine bočne slanine (BS) i telesne mase (TM).

ID	F	Ge	Sire	Dam	BS	TM	k
20206	5	2	2276	2920	24	115	20.87
23216	31	4	6320	6213	15	100	15
25616	31	4	6320	2652	17	108	15.74
27849	31	4	6320	9990	12	95	12.63
28184	31	4	6320	6213	10	115	8.7
29177	31	4	6320	12259	11	111	9.91
37749	31	1	3908	10006	5	102	4.9
38418	31	1	3908	15207	21	105	20
40097	31	1	3908	8737	10	96	10.42
40198	31	1	3908	14593	19	95	20
40507	109	1	15987	9479	12	110	10.91
40508	109	1	15987	9479	11	112	9.82
46184	64	1	16237	16238	8	105	7.62
47228	87	2	536	5901	12	117	10.26
47229	87	2	536	2680	15	110	13.64
47230	87	2	536	2680	10	100	10
47232	87	2	536	10486	10	81	12.35
47235	87	2	547	2358	12	102	11.76
47243	87	2	535	12	10	94	10.64
47244	87	2	535	13927	16	110	14.55
48520	87	2	536	2354	16	103	15.53
51087	3	1	13562	8500	17	115	14.78
52044	3	1	13562	723	25	180	13.89
52046	3	1	13562	723	20	178	11.24
52766	4	2	10911	4730	18	106	16.98
52767	4	2	10911	4730	12	102	11.76
52768	4	2	11913	12861	19	105	18.1
52769	4	2	10911	4946	14	103	13.59
53227	4	2	10911	12879	15	109	13.76

ID	F	Ge	Sire	Dam	BS	TM	k
53228	4	2	10911	12879	15	108	13.89
53553	31	1	8809	12454	19	113	16.81
54861	6	1	16670	9834	10	90	11.11
55762	114	2	16412	16411	8	113	7.08
55785	87	2	535	9788	12	100	12
56469	24	2	13626	4085	26	125	20.8
56470	24	2	13626	4085	30	131	22.9
56477	24	1	3850	4059	30	124	24.19
56478	24	1	3850	4059	24	132	18.18
56961	23	2	6822	13524	6	80	7.5
58026	27	1	15228	3132	11	109	10.09
58699	87	2	535	7617	14	96	14.58
58700	87	2	535	7617	12	86	13.95
58701	87	2	535	7617	16	99	16.16
58702	87	2	535	7617	10	106	9.43
58703	87	2	535	7617	12	108	11.11
58732	160	1	15677	317	15	104	14.42
58733	160	1	15677	317	14	103	13.59
59084	63	2	9374	17089	19	116	16.38
59085	63	2	9374	17089	20	125	16
60324	92	4	6714	10846	19	98	19.39
61333	109	1	16237	349	12	130	9.23
61367	6	1	16670	10778	9	106	8.49
61368	6	1	16670	10778	11	108	10.19
61564	109	1	16237	349	14	126	11.11
62231	64	10	13011	474	12	118	10.17
63745	6	1	16670	16984	14	120	11.67
63746	6	1	16670	16984	19	108	17.59
63747	6	1	16670	16984	13	120	10.83
69652	87	1	15677	19047	9	94	9.57
69954	31	1	8809	19393	16	95	16.84
70211	160	1	6795	19754	17	105	16.19
70345	31	4	6320	19953	12	128	9.38
70749	160	1	8809	20307	15	107	14.02
70750	160	1	8809	20307	12	100	12
71500	31	1	8809	21611	16	103	15.53
73647	31	4	8813	24058	17	90	18.89
73895	31	4	8813	24529	15	110	13.64
74417	33	10	25427	10019	10	110	9.09
74418	33	10	25427	20193	16	100	16
74424	33	10	25427	12446	5	95	5.26
74430	33	10	25427	10614	15	115	13.04
74438	33	10	25427	25418	15	100	15
74450	33	10	25427	12446	24	95	25.26
74456	33	10	25427	14350	8	115	6.96

ID	F	Ge	Sire	Dam	BS	TM	k
74457	33	10	25427	24111	16	105	15.24
74459	33	10	25427	10947	11	100	11
74461	33	10	25427	10018	15	115	13.04
74465	33	10	25427	3953	21	104	20.19
74466	33	10	25427	3953	10	108	9.26
74469	33	10	25427	8821	15	100	15
74471	33	10	25427	12773	20	107	18.69
74472	33	10	25427	12773	17	108	15.74
74946	31	1	3908	25735	13	105	12.38
74947	31	1	3908	25735	11	102	10.78
75787	160	1	8809	26274	13	100	13
75791	160	1	6795	26274	14	120	11.67
75797	160	1	15677	26274	13	105	12.38
75811	160	1	8809	26286	19	110	17.27
75813	160	1	8809	26286	17	100	17
75840	160	1	6795	26289	16	105	15.24
75841	160	1	6795	26289	14	95	14.74
76253	4	2	11913	26985	17	103	16.5
76439	14	1	27287	17982	10	80	12.5
76440	14	1	27287	17982	13	84	15.48
76441	14	1	27287	17982	14	80	17.5
76442	14	1	27287	22824	22	110	20
76444	14	1	27287	22824	23	112	20.54
76521	14	1	27287	14322	17	115	14.78
76523	14	1	27287	26895	22	110	20
76524	14	1	27287	22819	10	106	9.43
76526	14	1	27287	22819	15	100	15
76527	14	1	27287	14038	12	120	10
76528	14	1	27287	14038	10	127	7.87
76529	14	1	27287	26238	22	119	18.49
76561	14	1	27287	9424	16	115	13.91
76562	14	1	27287	20838	14	115	12.17
76563	14	1	27287	20838	16	115	13.91
76564	14	1	27287	20838	20	100	20
76565	14	1	27287	779	16	120	13.33

ID	F	Ge	Sire	Dam	BS	TM	k
76574	14	1	27287	9424	16	110	14.55
76579	14	1	27287	17864	22	120	18.33
76580	14	1	27287	10627	20	125	16
76581	14	1	27287	10627	20	115	17.39
76582	14	1	27287	1791	23	120	19.17
76583	14	1	27287	10627	23	115	20
76638	14	1	27287	26239	17	127	13.39
79836	3	1	13562	29689	17	130	13.08
80083	4	2	11913	29763	19	103	18.45
80084	4	2	11913	29763	17	105	16.19
80282	87	2	535	29963	10	87	11.49
80283	87	2	535	29963	10	104	9.62
80307	87	2	535	29992	15	100	15
81918	31	1	8809	32211	16	95	16.84
82172	160	1	3850	32463	18	104	17.31
82829	87	2	536	33422	11	81	13.58
82898	87	2	535	33557	12	89	13.48
83008	87	2	535	33657	10	94	10.64
84080	6	1	16670	35039	12	120	10
84083	6	1	16670	35039	10	100	10
84084	6	1	16670	35039	18	115	15.65
84436	87	2	535	35335	11	117	9.4
84593	31	1	15987	35475	17	105	16.19
84622	160	5	9889	35505	11	96	11.46
85153	87	2	536	36464	15	92	16.3
86190	33	10	25427	37755	16	100	16
86191	33	10	25427	37755	15	110	13.64
86192	33	10	25427	37755	20	120	16.67
86195	33	10	25427	37757	13	108	12.04
86994	31	1	30794	38429	17	108	15.74
87985	31	1	40096	24017	25	115	21.74
88008	31	1	40096	2511	18	100	18
88049	31	1	40096	10895	16	92	17.39
88050	31	1	40096	10895	17	107	15.89
88076	31	1	40096	10895	20	94	21.28

ID	F	Ge	Sire	Dam	BS	TM	k
88120	31	1	40096	40052	16	105	15.24
88195	31	1	30794	40177	20	100	20
88196	31	1	30794	40177	16	109	14.68
88197	31	1	30794	40177	22	130	16.92
88202	31	1	40096	40177	14	112	12.5
88601	16	2	40377	8303	9	118	7.63
88602	16	2	40377	8303	10	118	8.47
89477	16	1	41500	5754	13	95	13.68
89478	16	1	41500	5754	19	110	17.27
89479	16	1	41500	5754	13	89	14.61
90033	27	5	1277	42347	9	90	10
91382	31	1	44211	29438	17	100	17
91383	31	1	44211	29438	15	99	15.15
91401	31	1	44211	28744	19	108	17.59
91402	31	1	44211	8734	19	106	17.92
91403	31	1	44211	8734	14	105	13.33
91439	31	1	44211	10895	20	109	18.35
91441	31	1	44211	10895	17	106	16.04
91468	31	1	44211	38441	12	95	12.63
91485	31	1	44211	31243	16	96	16.67
91486	31	1	44211	26702	19	112	16.96
91491	31	1	44211	29438	16	99	16.16
91549	31	1	44211	25657	14	110	12.73
91550	31	1	44211	25657	14	103	13.59
91919	27	5	1277	44693	7	97	7.22
92910	6	4	46284	819	11	112	9.82
92911	6	4	46284	819	7	98	7.14
92912	6	4	46284	819	13	96	13.54
92913	6	4	46284	819	12	96	12.5
92918	6	4	46284	35050	10	104	9.62
92943	6	4	46284	16974	10	102	9.8
92945	6	4	46284	184	15	100	15
92946	6	4	46284	819	14	102	13.73
92988	6	4	46284	44103	14	95	14.74
92989	6	4	46284	9845	14	110	12.73

ID	F	Ge	Sire	Dam	BS	TM	k
92993	6	4	46284	6889	14	116	12.07
93119	31	1	44211	46560	17	116	14.66
93129	31	1	30794	46570	21	118	17.8
93226	6	4	46284	46730	9	113	7.96
93227	6	4	46284	46730	16	116	13.79
94625	5	2	49341	10885	20	115	17.39
95066	15	1	50060	8936	19	120	15.83
95283	31	4	50487	30934	19	106	17.92
95319	31	4	50487	29503	20	107	18.69
95320	31	4	50487	26714	18	102	17.65
95350	31	4	50487	28502	17	98	17.35
95351	31	4	50487	24925	16	110	14.55
95352	31	4	50487	28502	19	102	18.63
95370	31	4	50487	16982	16	97	16.49
95371	31	4	50487	16982	15	97	15.46
95387	31	4	50487	47425	14	105	13.33
95695	15	2	50815	12305	24	141	17.02
95938	6	4	46284	50990	16	125	12.8
95939	6	4	46284	50990	14	125	11.2
95941	6	2	46284	50990	9	94	9.57
96009	5	2	51037	10882	23	110	20.91
96129	23	2	51144	41937	14	105	13.33
96130	23	2	51144	41937	10	95	10.53
96447	31	4	51392	7158	18	112	16.07
96671	6	1	16670	51570	9	104	8.65
96672	6	1	16670	51570	9	105	8.57
97208	6	1	52323	44150	8	96	8.33
97641	160	1	12641	52914	13	98	13.27
97644	160	1	12641	52914	12	90	13.33
97992	6	1	16670	53061	11	107	10.28
97993	6	1	16670	53061	11	110	10
98508	83	1	53231	12117	11	100	11
98511	83	1	53231	12117	8	95	8.42
98512	83	1	53231	12117	10	112	8.93
99142	31	1	44211	53531	14	101	13.86

ID	F	Ge	Sire	Dam	BS	TM	k
99246	31	1	53551	25735	13	105	12.38
99247	31	1	53551	48634	17	110	15.45
99248	31	1	53551	48634	17	100	17
99265	31	1	53551	30895	15	105	14.29
100474	84	1	55088	12425	14	93	15.05
100475	84	1	55088	12439	12	96	12.5
100476	84	1	55088	3990	13	100	13
101361	31	4	56069	46639	15	119	12.61
101362	31	4	56069	29357	16	103	15.53
101367	31	4	56069	28675	16	110	14.55
101368	31	4	56069	3798	16	110	14.55
101369	31	4	56069	31153	17	93	18.28
101370	31	4	56069	28675	14	93	15.05
101371	31	4	56069	31153	16	100	16
102419	31	10	57564	11301	14	130	10.77
105821	14	2	68193	17973	15	105	14.29
105822	14	2	68193	17973	12	114	10.53
105845	14	2	68193	20806	20	140	14.29
105846	14	2	68193	20806	20	124	16.13
107827	31	4	50487	69251	15	100	15
107828	31	4	50487	69251	16	115	13.91
107829	31	4	50487	69251	17	104	16.35
107830	31	4	50487	69251	16	114	14.04
108317	31	4	50487	69772	16	96	16.67
108459	31	4	50487	70065	20	107	18.69
108460	31	4	50487	70065	16	94	17.02
108461	31	4	50487	70065	18	108	16.67
108616	160	1	6795	70196	12	97	12.37
108624	160	1	6795	70207	13	90	14.44
109031	31	1	53551	70433	15	113	13.27
109041	31	1	40096	70509	16	106	15.09
109052	31	1	40096	70599	17	95	17.89
109053	31	1	40096	70599	16	97	16.49
109248	31	4	51392	70811	16	105	15.24
109263	31	1	3908	70838	19	100	19

ID	F	Ge	Sire	Dam	BS	TM	k
109264	31	1	3908	70838	16	102	15.69
109396	31	1	30794	70850	15	116	12.93
109412	33	10	70879	8682	13	110	11.82
109413	33	10	70879	21658	15	105	14.29
109417	33	10	70879	8840	23	100	23
109419	33	10	70879	10953	20	110	18.18
109424	33	10	70879	26795	20	105	19.05
109425	33	10	70879	26795	21	105	20
109436	33	10	70879	20102	20	110	18.18
109492	14	1	27287	70953	30	135	22.22
109643	15	1	71102	15351	17	123	13.82
109681	3	1	71132	4360	16	177	9.04
109682	3	1	71132	12786	18	170	10.59
109846	31	1	44211	71236	16	108	14.81
110265	31	1	40096	71867	16	96	16.67
110404	33	10	48683	72215	19	105	18.1
110905	31	4	72597	40160	16	113	14.16
110909	31	4	72597	14385	14	108	12.96
111331	31	4	73222	6745	17	105	16.19
111337	31	4	73222	27627	20	105	19.05
111338	31	4	73222	15206	17	105	16.19
111376	31	4	73222	29124	19	105	18.1
111377	31	4	73222	24730	15	100	15
111378	31	4	73222	29124	16	103	15.53
111490	31	4	50487	73272	16	91	17.58
111542	31	4	51392	73282	16	107	14.95
111645	33	10	25619	73284	18	102	17.65
111657	31	1	44211	73456	17	100	17
111659	31	1	44211	73456	15	105	14.29
111754	31	1	40096	73575	22	112	19.64
111757	31	1	40096	73575	17	104	16.35
111759	31	1	40096	73575	16	108	14.81
111760	31	1	40096	73575	17	120	14.17
111795	31	4	73588	283	19	95	20
111811	31	4	73588	69113	16	119	13.45

ID	F	Ge	Sire	Dam	BS	TM	k
111833	31	4	73588	68924	17	100	17
111834	31	4	73588	68924	15	100	15
111899	31	4	73588	69198	19	109	17.43
111900	31	4	73588	69198	18	113	15.93
111928	31	4	73588	69113	13	90	14.44
111929	31	4	73588	5600	17	108	15.74
111930	31	4	73588	29335	16	105	15.24
111996	31	4	73588	37365	14	99	14.14
112035	31	4	73588	11907	16	94	17.02
112049	31	4	73588	11907	19	100	19
112050	31	4	73588	11907	17	97	17.53
112070	31	4	73588	286	16	97	16.49
112071	31	4	73588	286	16	108	14.81
112090	31	4	73606	70065	19	97	19.59
112131	31	4	73606	35550	15	94	15.96
112173	31	4	73606	7882	17	98	17.35
112174	31	4	73606	32002	14	95	14.74
112187	31	4	73606	73546	16	99	16.16
112188	31	4	73606	73546	19	102	18.63
112191	31	4	73606	29124	13	106	12.26
112228	31	4	73606	287	16	98	16.33
112229	31	4	73606	29124	17	103	16.5
112410	31	4	73606	73735	14	112	12.5
112435	33	10	73750	19349	15	95	15.79
112440	33	10	73750	19362	17	92	18.48
113194	31	4	73222	74095	14	93	15.05
113853	31	1	74828	46556	18	124	14.52
113916	31	1	74828	46560	13	103	12.62
114626	16	1	75633	5754	14	98	14.29
114658	16	1	75633	30301	15	91	16.48
114659	16	1	75633	30301	17	100	17
114782	33	10	73750	75756	19	103	18.45
114862	160	1	6795	75819	15	110	13.64
114863	160	1	6795	75819	15	110	13.64
114912	160	1	12641	75830	13	97	13.4

ID	F	Ge	Sire	Dam	BS	TM	k
115434	33	10	25427	76138	20	135	14.81
115781	31	1	44211	76727	15	102	14.71
115782	31	1	44211	76727	15	97	15.46
116567	160	1	77785	75782	15	109	13.76
116568	160	1	77785	27829	10	100	10
116569	160	1	77785	75806	14	105	13.33
116570	160	1	77785	75806	16	107	14.95
116576	160	1	77785	70207	16	112	14.29
116577	160	1	77785	70207	12	115	10.43
116578	160	1	77785	70207	12	105	11.43
116579	160	1	77785	70207	14	115	12.17
116583	160	1	77785	26289	14	97	14.43
116584	160	1	77785	26289	12	94	12.77
116585	160	1	77785	28122	14	98	14.29
116615	31	1	74828	77791	17	106	16.04
117091	84	2	78692	9644	14	100	14
117135	84	2	78692	7352	16	98	16.33
117136	84	2	78692	7352	8	99	8.08
117183	84	2	78694	12316	13	110	11.82
117184	84	2	78694	15751	9	105	8.57
117320	6	1	16670	78782	9	95	9.47
117575	6	4	78911	44136	13	104	12.5
117588	6	4	78911	40702	8	95	8.42
117589	6	4	78911	40702	9	107	8.41
117591	6	4	78911	35010	17	112	15.18
117653	6	4	46284	78914	13	115	11.3
118373	6	1	16670	79070	11	118	9.32
118560	6	4	78911	79278	11	101	10.89
118972	6	4	46284	79506	15	100	15
118981	6	4	46284	79526	14	104	13.46
118982	6	4	46284	79526	10	103	9.71
119001	6	4	78911	79541	10	112	8.93
119002	6	4	78911	79541	14	118	11.86
119114	31	4	50487	79720	16	93	17.2
119456	4	2	10911	80117	14	100	14

ID	F	Ge	Sire	Dam	BS	TM	k
119457	4	2	10911	80117	12	95	12.63
119458	4	2	10911	80117	11	90	12.22
120301	31	4	80778	72287	15	108	13.89
120302	31	4	80778	69067	15	100	15
120317	31	4	80778	72287	15	111	13.51
120318	31	4	80778	72287	15	98	15.31
120319	31	4	80778	72287	14	122	11.48
120343	33	10	25427	80803	19	116	16.38
120344	33	10	25427	80803	16	103	15.53
120347	33	10	25427	80803	22	99	22.22
120348	33	10	25427	80803	18	98	18.37
120493	14	1	27287	81049	22	120	18.33
120494	14	1	73937	81049	22	130	16.92
120495	14	1	73937	81049	22	125	17.6
120511	14	1	27287	81095	19	110	17.27
121057	31	4	50487	81887	16	108	14.81
122557	15	2	84055	29583	18	140	12.86
122675	6	1	84075	79596	11	107	10.28
123405	160	5	9889	84615	13	90	14.44
123406	160	5	9889	84615	11	95	11.58
125338	31	1	40096	87019	16	110	14.55
125854	160	1	88001	26289	12	97	12.37
125855	160	1	88001	26289	12	90	13.33
125892	31	1	44211	88164	17	110	15.45
125903	31	1	57562	88174	13	96	13.54
125904	31	1	57562	88174	14	97	14.43
126354	31	4	89382	77226	17	105	16.19
126355	31	4	89382	77226	15	89	16.85
126356	31	4	89382	77226	19	106	17.92
126357	31	4	89382	77226	17	96	17.71
126361	31	4	89382	31153	15	100	15
126362	31	4	89382	81806	14	96	14.58
126403	31	4	89382	25922	17	100	17
126423	31	4	89382	72377	15	104	14.42
126424	31	4	89382	14079	12	103	11.65

ID	F	Ge	Sire	Dam	BS	TM	k
126453	31	4	89382	4168	16	105	15.24
127046	67	2	90395	58050	12	110	10.91
127049	67	2	90395	8408	10	110	9.09
127053	67	2	90395	8487	17	120	14.17
127135	109	10	60279	90435	12	106	11.32
127136	109	10	60279	90435	12	109	11.01
127668	6	2	5358	91207	8	100	8
127669	6	2	5358	91207	7	103	6.8
127670	6	2	5358	91207	9	115	7.83
128280	6	4	78911	92923	9	107	8.41
128281	6	4	78911	92923	10	100	10
128529	15	2	93669	4314	16	125	12.8
130094	6	2	95635	13442	9	88	10.23
130145	6	2	5358	95917	11	109	10.09
130146	6	2	5358	95924	11	104	10.58
132690	160	1	77785	107023	10	100	10
132691	160	1	77785	107023	16	112	14.29
133255	31	4	107474	70077	20	105	19.05
133341	31	4	107474	72275	15	90	16.67
133369	31	4	107474	26714	18	98	18.37
133370	31	4	107474	26714	22	96	22.92
133426	31	4	107474	70823	22	115	19.13
133427	31	4	107474	14353	15	100	15
133469	31	4	107474	41316	20	108	18.52
133490	31	4	107474	69054	14	96	14.58
133502	31	4	107474	72374	13	93	13.98
133523	31	4	107474	48614	19	102	18.63
133524	31	4	107474	48614	19	112	16.96
133612	31	4	107474	30210	15	99	15.15
133614	31	4	107474	29503	14	93	15.05
133649	31	4	107474	30873	19	114	16.67
133698	31	4	72597	107484	20	108	18.52
133724	31	4	50487	107495	15	108	13.89
133993	31	4	107620	75434	13	98	13.27
134185	31	4	50487	107742	20	105	19.05

ID	F	Ge	Sire	Dam	BS	TM	k
134626	31	4	50487	108310	17	108	15.74
134647	33	10	25427	108345	22	115	19.13
134952	160	1	6795	108631	11	98	11.22
134959	160	1	8738	108631	12	90	13.33
134979	160	1	77785	108635	14	111	12.61
134980	160	1	77785	108635	13	100	13
134981	160	1	77785	108635	14	110	12.73
134992	160	1	8738	108635	12	96	12.5
134993	160	1	8738	108635	12	98	12.24
134994	160	1	8738	108635	12	97	12.37
135032	31	4	56069	108702	17	113	15.04
135353	33	10	25427	109075	19	120	15.83
135359	33	10	73750	109088	17	101	16.83
135485	160	1	77785	109183	10	96	10.42
135495	160	1	12641	109183	14	101	13.86
135496	160	1	12641	109183	11	95	11.58
135501	160	1	77785	109184	10	100	10
135833	33	10	109435	80735	16	106	15.09
135834	33	10	109435	80735	16	90	17.78
136220	31	4	50487	110296	17	110	15.45
136248	33	10	110307	109428	18	95	18.95
136249	33	10	110307	109428	15	93	16.13
136261	33	10	25427	110309	15	106	14.15
136834	31	4	111330	72486	16	98	16.33
136835	31	4	111330	72486	20	106	18.87
137113	31	4	111778	3798	16	110	14.55
137960	33	10	113489	40188	17	95	17.89
137961	33	10	113489	40188	18	95	18.95
137962	33	10	113489	40188	17	104	16.35
137963	33	10	113489	40188	16	94	17.02
137964	33	10	113489	40188	15	94	15.96
138057	15	2	84054	113570	19	122	15.57
138106	31	4	107474	113632	18	99	18.18
138398	6	1	84075	114467	14	100	14
138399	6	1	84075	114467	11	122	9.02

ID	F	Ge	Sire	Dam	BS	TM	k
138444	160	1	77785	114728	12	96	12.5
138445	160	1	77785	114728	14	94	14.89
138456	160	1	8738	114735	12	96	12.5
138457	160	1	8738	114735	12	97	12.37
138519	33	10	114770	109089	21	105	20
138520	33	10	114770	109089	19	97	19.59
138521	33	10	114770	109089	21	109	19.27
139319	6	10	115039	10158	15	125	12
139320	6	10	115039	10158	15	110	13.64
139325	6	10	115039	10703	11	122	9.02
139326	6	10	115039	10158	17	130	13.08
139327	6	10	115039	10158	10	111	9.01
139351	6	10	115039	115037	15	130	11.54
139352	6	10	115039	115037	14	130	10.77
139402	31	1	115244	25657	16	93	17.2
139403	31	1	115244	111479	16	107	14.95
139496	31	1	115244	28744	17	112	15.18
139500	31	1	115244	46570	14	100	14
139664	33	10	114770	115445	16	103	15.53
139665	33	10	114770	115445	15	95	15.79
139678	33	10	112463	115450	15	115	13.04
139755	31	1	115582	73575	17	105	16.19
139756	31	1	115582	73575	19	108	17.59
140320	160	1	12641	116555	12	99	12.12
140321	160	1	12641	116555	12	96	12.5
140322	160	1	12641	116555	13	99	13.13
140323	160	1	12641	116555	13	96	13.54
140945	31	1	118625	71499	16	104	15.38
140977	31	1	118625	73448	16	96	16.67
141014	31	1	118625	46472	18	100	18
141853	33	10	120342	23808	16	100	16
141854	33	10	120342	23808	14	105	13.33
143113	6	1	122696	29304	11	98	11.22
143580	160	1	124366	114728	15	95	15.79
143581	160	1	124366	114728	14	100	14

ID	F	Ge	Sire	Dam	BS	TM	k
143582	160	1	124366	114728	13	96	13.54
143583	160	1	124366	110948	16	95	16.84
143584	160	1	124366	110948	13	98	13.27
143585	160	1	124366	110948	16	100	16
143597	160	1	124366	93696	13	90	14.44
143599	160	1	124366	110947	12	96	12.5
143600	160	1	124366	114728	12	95	12.63
143608	160	1	124366	108635	13	94	13.83
143609	160	1	124366	108635	13	100	13
143692	160	1	124366	124375	12	90	13.33
143694	160	1	124366	124375	13	98	13.27
143695	160	1	124366	124375	12	95	12.63
143696	160	1	124366	124375	10	94	10.64
143697	160	1	12641	124375	13	97	13.4
143698	160	1	12641	124375	11	90	12.22
144197	31	1	44211	125889	16	100	16
144215	160	1	125891	116565	10	96	10.42
144216	160	1	125891	116565	14	98	14.29
144226	160	1	125891	109184	13	95	13.68
144227	160	1	125891	124372	12	96	12.5
144228	160	1	125891	124372	13	95	13.68
144229	160	1	125891	124372	13	96	13.54
144230	160	1	125891	124372	14	90	15.56
144231	160	1	125891	93695	12	99	12.12
144232	160	1	125891	93695	13	96	13.54
144414	67	2	90484	127145	9	103	8.74
145105	160	10	44695	132498	13	120	10.83
145106	160	10	44695	132498	14	120	11.67
145107	160	10	44695	132498	14	100	14
145117	160	10	44695	132498	12	96	12.5
145118	160	10	44695	132498	14	97	14.43
145119	160	10	44695	132498	14	90	15.56
145126	160	10	44695	132498	10	90	11.11
145127	160	10	44695	132498	10	98	10.2
145128	160	10	44695	132498	10	97	10.31

ID	F	Ge	Sire	Dam	BS	TM	k
145129	160	10	44695	132498	14	99	14.14
145130	160	10	44695	132498	10	90	11.11
145131	160	10	44695	132498	13	98	13.27
145132	160	10	44695	132498	13	98	13.27
145133	160	10	44695	132498	14	96	14.58
145134	160	10	44695	132498	13	95	13.68
145176	160	10	44695	132518	14	111	12.61
145177	160	10	44695	132518	15	112	13.39
145178	160	10	44695	132518	14	110	12.73
145179	160	10	44695	132518	12	100	12
145295	6	4	78911	133052	13	85	15.29
145743	160	4	133902	111239	14	97	14.43
145746	160	4	133902	111239	13	96	13.54
145747	160	4	133902	111238	12	95	12.63
145748	160	4	133902	92084	13	90	14.44
145749	160	4	133902	92084	12	98	12.24
145750	160	4	133902	92084	13	98	13.27
145751	160	4	133902	92084	12	97	12.37
145752	160	4	133902	92076	12	97	12.37
146268	160	1	125891	134978	12	97	12.37
146270	160	1	125891	134978	12	99	12.12
146271	160	1	125891	134978	12	95	12.63
146272	160	1	125891	134978	12	96	12.5
146803	33	10	136260	80802	17	105	16.19
146822	160	1	125891	136453	12	95	12.63
146838	160	1	125891	136471	12	100	12
146898	160	4	133902	136722	13	100	13
146899	160	4	133902	136722	11	90	12.22
146991	31	10	137959	109089	16	115	13.91
147998	160	10	14254	145120	13	95	13.68
148003	160	10	14254	145171	12	95	12.63
148004	160	10	14254	145171	13	99	13.13
148005	160	10	14254	145171	13	100	13

Zaključak

Prvenstveno u ovom radu glavni cilj je bio dati osnovni uvid u linearne mešovite modele, kao i u oblasti u kojima se mogu primeniti sa posebnom pažnjom usmerenom na oplemenjivanje životinja. Glavni akcenat stavljen je na ocene parametara u mešovitim modelima. BLUP, kao način ocene slučajnih efekata u mešovitim modelima, pri čemu se istovremeno ocenjuju i fiksni efekti, je od velike važnosti u proceni uzgojnih odnosno oplemenjivačkih vrednosti. Smatra se najboljom i najobjektivnijom ocenom. Pre svega linearni mešoviti modeli imaju primenu u oblastima u kojima veliki uticaj imaju genetski faktori. Stoga i ne čudi da ih najčešće susrećemo u oplemenjivanju životinja, biljaka, psihologiji, medicini i slično. U poslednje vreme, linearni mešoviti modeli nalaze značajnu primenu u detektovanju postojanja biološkog napada ili bioterorisma. Pomoću takvih modela, utvrđuje se, na određenim grupama ili regijama, koliko pojavljivanje posmatrane bolesti odstupa od očekivane vrednosti, dobijene na osnovu istorijskih pdataka.

Raznolikost individua i njihovih osobina rezultat su uticaja naslednih faktora i spoljašnje sredine. Među svim faktorima koji utiču na osobine neke jedinke, najmanje se genetski menjaju tokom njenog života, pa su od posebnog značaja kada je u pitanju selekcija životinja. Linearnom regresijom možemo opisati uticaj fiksnih faktora na neku posmatranu osobinu. Međutim, ukoliko želimo uvrstiti uticaj genetike, kako same individue tako i njenih roditelja, to možemo uraditi pomoću linearnih mešovitih modela. Drugim rečima takav uticaj smatramo slučajnim, te ga uvrštavamo u model kao slučajni efekat. Na taj način možemo izabrati najbolju individuu koju ćemo koristiti za dalju reprodukciju. Pre svega u takvim modelima nas zanima ocena slučajnih efekata, jer tako biramo najbolju životinju, pri čemu je osnovni kriterijum oplemenjivačka vrednost, odnosno genetski potencijal. S druge strane, kako jedinka određenog genotipa može ispoljiti svoje bitne karakteristike u određenim uslovima životne sredine, a u drugim ne nužno, neophodno je posmatranje međusobnog delovanja fiksnih i slučajnih efekata, jer se samo na taj način mogu ispoljiti željene osobine kod date individue.

Glavni cilj primene linearnih mešovitih modela u selekciji životinja je definitivno ekonomski aspekt. U slučaju kad imamo genetski kvalitetno grlo i pri tome utičemo na uslove spoljašnje sredine, daljom analizom i reprodukcijom možemo unapređivati naše stado. U Srbiji se godišnje potroši na stotine hiljada evra na uvoz semena domaćih životinja. Pored toga, Srbija ima veliki potencijal kada je kvalitet bikova u pitanju, kao i kvalitet nerastova (muških svinja). Za saradnju su zainteresovane zemlje poput Austrije i Nemačke, dok se procene oplemenjivačkih vrednosti rade upravo pomoću BLUP ocena. Na ovaj način se vrši selekcija životinja, odnosno procena njihovog genetskog kvaliteta. Od interesa mogu biti kvalitet mesa, mlečnost ili količina mleka, kvalitet samog mleka, brzina konja i slično u zavisnosti od vrste životinje, rase i cilja samog odgajivača.

O autoru

Ivana Jančić je rođena u Kotoru, 24.05.1990. godine. Osnovnu školu *„Dašo Pavičić"* završila je u Herceg Novom, gde 2005. i upisuje gimnaziju *„Ivan Goran Kovačić"*. Gimnaziju *„Laza Kostić"* završava u Novom Sadu 2009. godine gde i upisuje *Prirodno-matematički Fakultet*, smer Primenjena Matematika, modul Matematika finansija. Osnovne bachelor studije završava 2013. godine kada i upisuje master studije istog smera. Položila je sve ispite predviđene nastavnim planom i programom master studija u junskom ispitnom roku 2015.

Poseduje znanje engleskog i osnovno znanje nemačkog i italijanskog jezika.

Literatura

[1] Mrode, R.A. *Linear models for the prediction of animal breeding values,* CABI Publishing (2005), 332-416.

[2] Agresti Alan, *Categorical Data Analysis,* John Wiley & Sons, Inc., Hoboken, New Jersey (2002).

[3] Vitomir Vidović, *Linearni modeli u oplemenjivanju životinja,* Stylos d.o.o.,Stylos art, Novi Sad (2013).

[4] Jiming Jiang, *Linear and Generalized Linear Mixed Models and Their Applications,* Springer Science + Business Media, LLC (2007).

[5] Walter W.Stroup, *Generalized Linear Mixed Models Modern Concepts, Methods and Applications,* Taylor & Francis Group, LLC (2013).

[6] Annette J. Dobson, *An introduction to generalized linear models,* Chapman & Hall/CRC (2002).

[7] V. Pantelić, M. Plavšić, S. Trivunović , S. Aleksić , Lj. Sretenović, D. Ostojić-Andrić, D. Nikšić, *The evaluation of breeding value of simmental bulls for milk performance in Serbia*, Institute for Animal Husbandry, Belgrade-Zemun (2011)

[8] Bodo Winter, *A very basic tutorial for performing linear mixed effects analyses*, University of California, Cognitive and Information Sciences (2014).

[9] Danijela Rajter-Ćirić, *Verovatnoća*, Univerzitet u Novom Sadu, Prirodno matematički fakultet u Novom Sadu (2009).

[10] C. Radhakrishna Rao, Sujit Kumar Mitra, *Generalized inverse of a Matrix and its Applications*, Indian Statistrical Institute.

[11] http://www.public.iastate.edu/ dnett/S611/40BLUP.pdf

[12] https://cran.r-project.org/web/packages/lme4/lme4.pdf

[13] https://cran.r-project.org/web/packages/lme4/lme4.pdf

UNIVERZITET U NOVOM SADU
PRIRODNO-MATEMATIČKI FAKULTET
KLJUČNA DOKUMENTACIJSKA INFORMACIJA

Redni broj:
RBR

Identifikacioni broj:
IBR

Tip dokumentacije: Monografska dokumentacija
TD

Tip zapisa: Tekstualni štampani materijal
TZ

Vrsta rada: Master rad
VR

Autor: Ivana Jančić
AU

Mentor: dr Zagorka Lozanov-Crvenković
MN

Naslov rada: Linearni mešoviti modeli i njihova upotreba u selekciji domaćih životinja
NR

Jezik publikacije: srpski (latinica)
JP

Jezik izvoda: srpski/engleski
JI

Zemlja publikovanja: Republika Srbija
ZP

Uže geografsko područje: Vojvodina
UGP

Godina: 2017.
GO

Izdavač: Autorski reprint
IZ

Mesto i adresa: Univerzitet u Novom Sadu, Prirodno-matematički fakultet, Trg Dositeja Obradovića 4
MA

Fizički opis rada: 3/72/0/11/8/0/1
(broj poglavlja/strana/lit citata/tabela/slika/grafika/priloga)
FO

Naučna oblast: Matematika
NO

Naučna disciplina: Statistika
ND

Predmetna odrednica/Ključne reči: mešoviti linearni modeli, oplemenjivanje životinja, najbolji linearni nepristrasan prediktor, normalna raspodela, statistika, Kronekerov produkt, selekcija životinja.
PO

UDK:

Čuva se: u biblioteci Departmana za matematiku i informatiku, Novi Sad
ČU

Važna napomena:
VN

Izvod: U ovom radu akcenat je stavljen na linearne mešovite modele. U prvom delu objašnjeni su osnovni matematički pojmovi koji su bili potrebni za analizu u nastavku rada. Definisana je podela slučajnih promenljivih, normalna raspodela, kao i odgovarajuća funkcija gustine raspodele za jednodimenzionalnu i višedimenzionalnu slučajnu promenljivu. U drugom delu definisani su linearni mešoviti modeli, a potom je data i njihova klasifikacija. Obrađeni su metodi ocenjivanja parametara za Gausove i ne-Gausove linearne mešovite modele. Centralni deo se odnosi na ocenu, pre svega slučajnih efekata. Za ocenu je korišćen najbolji linearni nepristrasan prediktor - BLUP, koji podrazumeva minimizaciju srednje kvadratne greške. Predstavljen je mešoviti model jednačina, kao metod za izračunavanje ocena fiksnih i slučajnih efekata. Treći, poslednji deo, se odnosi na primenu ovih modela u oplemenjivanju životinja. Na samom kraju obrađeni su stvarni podaci u cilju selekcije svinja nerastova.
IZ

Datum prihvatanja teme od strane NN Veća: 8.07.2016.
DP

Datum odbrane:
DO

Članovi komisije:

Predsednik: dr Danijela Rajter-Ćirić, redovni profesor, Prirodno-matematički fakultet, Univerzitet u Novom Sadu

Mentor: dr Zagorka Lozanov-Crvenković, redovni profesor, Prirodno-matematički fakultet, Univerzitet u Novom Sadu.

Član: dr Ivana Štajner-Papuga, redovni profesor, Prirodno-matematički fakultet, Univerzitet u Novom Sadu

DO

UNIVERSITY OF NOVI SAD
FACULTY OF NATURAL SCIENCES AND MATHEMATICS
KEY WORDS DOCUMENTATION

Accession number:
ANO

Identification number:
INO

Document type: Monographic type
DT

Type of record: Text printed material
TR

Contents code: Master thesis
CC

Author: Ivana Jančić
AU

Mentor: Zagorka Lozanov-Crvvenković, Ph.D.
MN

Title: Linear mixed models an their application in selection of domestic animals
TI

Language of text: Serbian(Latin)
LT

Language of abstract: Serbian/English
LA

Country of publication: Republic of Serbia
CP

Locality of publication: Vojvodina
LP

Publication year: 2017.
PY

Publisher: Author's reprint

PU

Publ.place: University of Novi Sad, Faculty of Science, Trg Dositeja Obradovića 4
PP

Physical description: 3/72/0/11/8/0/1
(chapters/pages/literature/tables/pictures/graphs/additional lists)
PD

Scientific field: Mathematics
SF

Scientific discipline: Statistics
SD

Subject / Key words: linear mixed model, animal breeding, best linear unbiased predictor, normal distribution, statistics, Kronecker product, animal selection.
SKW

UC:

Holding data: The Library of the Department of Mathematics and Informatics, Novi Sad
HD

Note:
N

Abstract: The main subject of this master thesis is linear mixed models. The first chapter provides an overview of some mathematical terms which was necessary for analysis in the following part. Also, classification of random variables, normal distribution and probability density function for random variable are defined in this part. Linear mixed models are described in the second part. We classified linear mixed model as Gaussian mixed models and non-Gaussian mixed models. Analysis in estimation have been made in both of this models. The central part of this thesis is reserved for the best linear unbiased predictor-BLUP, based on the minimised value of the mean square error. There, a method for solving both fixed effects and random effects - linear mixed model equation is represented . The third part is based on application of these models in animal breeding. At the end of this thesis an example of real data describes application of linear mixed model in swine selection.
AB

Accepted by Scientific Board on: 8.07.2016.
ASB

Defended:
DE

Thesis defend board:

President: Danijela Rajter-Ćirić, full Professor, Faculty of Science, University of Novi Sad

Mentor: Zagorka Lozanov-Crvenković, Ph.D., full professor, Faculty of Sciences, University of Novi Sad

Member: Ivana Štajner-Papuga, Ph.D., full professor, Faculty of Sciences, University of Novi Sad

DB