

**T.C.
AKDENİZ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**KANTİTATİF ÖZELLİKLERDE GENOMİK DEĞERLENDİRME: TAHMİN
İSABETİNİ ETKİLEYEN FAKTÖRLER VE GBLUP YÖNTEMİNİN İSABETİ
İÇİN BİR ÜST SINIR**

Emre KARAMAN

**DOKTORA TEZİ
ZOOTEKNİ ANABİLİM DALI**

2016

**T.C.
AKDENİZ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**KANTİTATİF ÖZELLİKLERDE GENOMİK DEĞERLENDİRME: TAHMİN
İSABETİNİ ETKİLEYEN FAKTÖRLER VE GBLUP YÖNTEMİNİN İSABETİ
İÇİN BİR ÜST SINIR**

Emre KARAMAN

**DOKTORA TEZİ
ZOOTEKNİ ANABİLİM DALI**

2016

T.C.
AKDENİZ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

KANTİTATİF ÖZELLİKLERDE GENOMİK DEĞERLENDİRME: TAHMİN
İSABETİNİ ETKİLEYEN FAKTÖRLER VE GBLUP YÖNTEMİNİN İSABETİ
İÇİN BİR ÜST SINIR

Emre KARAMAN

DOKTORA TEZİ
ZOOTEKNİ ANABİLİM DALI

Bu tez 28/10/2016 tarihinde aşağıdaki jüri tarafından oy birliği ile kabul edilmiştir.

Prof. Dr. Mehmet Ziya FIRAT

Prof. Dr. İbrahim CEMAL

Doç. Dr. Burak KARACAÖREN

Doç. Dr. Doğan NARİNÇ

Yrd. Doç. Dr. Aşkın GALİÇ



ÖZET

KANTİTATİF ÖZELLİKLERDE GENOMİK DEĞERLENDİRME: TAHMİN İSABETİNİ ETKİLEYEN FAKTÖRLER VE GBLUP YÖNTEMİNİN İSABETİ İÇİN BİR ÜST SINIR

Emre KARAMAN

Doktora Tezi, Zootekni Anabilim Dalı
Danışman : Prof. Dr. Mehmet Ziya FIRAT
Ekim 2016, 66 sayfa

Bu çalışmada, genomik tahmin R^2 değerinin asimptotik davranışının belirlenmesi amaçlanmıştır. İnsanlardaki bağlantı dengesizliğini yansıtmaması ve çok sayıda bireye ait genomik verinin elde edilmesi amacıyla, 1,000 Genom Projesi (1,000 Genomes Project) kapsamında 85 Caucasian (Beyaz ırk) bireyinden elde edilen haplotip verileri kullanılarak 10,000 birey 111 kuşak boyunca şansa bağlı olarak eşleştirilecek şekilde bir simülasyon gerçekleştirilmiştir. Genom uzunluğu 0.5 Morgan olarak kısıtlanmış ve simülasyonlar yalnızca ilk 5 kromozomdan seçilen 0.1 Morgan uzunluğundaki bir alanda yer alan lokuslar kullanılarak gerçekleştirilmiştir. Toplam uzunluğu 30 Morgan olan bir genom için elde edilecek genomik tahmin R^2 değeri bu çalışmadakinin 60 katı büyüklüğünde bir referans popülasyon gerektirecektir. Kalıtım derecesi 0.8 olan bir özellik için QTL ve marker allel frekanslarının dağılımları bakımından farklılık gösteren bazı senaryolar oluşturulmuştur. Her bir senaryodaki marker sayısı 4,200, QTL sayısı ise 70 olarak belirlenmiştir. Genomik değerlendirilmede isabet bir tahmin edilebilirlik problemi şeklinde ele alınarak güvenilirlik ve dolayısıyla da genomik tahmin R^2 değeri için bir üst sınır önerilmiştir. Ayrıca bazı genomik tahmin yöntemleri de, GBLUP, BayesB ve BayesC, isabetleri bakımından karşılaştırılmıştır. Sonuçlar, genom boyu 30 Morgan olarak alındığında Bayesçi değişken seçim yöntemleri BayesB ve BayesC'nin görece düşük referans popülasyon büyüklüklerinde (< 6,000 birey) GBLUP yöntemine göre bir üstünlüklerinin olmadığını ortaya koymuştur. Diğer yandan, referans popülasyon büyüklüğü arttıkça söz konusu yöntemlerin GBLUP yöntemine göre daha isabetli tahminler sağladığı ortaya konmuştur. Referans popülasyon büyüklüğü yaklaşık yarım milyona ulaştığında her üç tahmin yöntemi de benzer genomik tahmin R^2 değerleri vermiş olup, bu değerler genomik kalıtım derecesine yaklaşmaktadır.

ANAHTAR KELİMELELER: Genomik kalıtım derecesi, Tüm genom regresyon yöntemleri, Güvenilirlik, Genetik değer

JÜRİ: Prof. Dr. Mehmet Ziya FIRAT (Danışman)

Prof. Dr. İbrahim CEMAL

Doç. Dr. Burak KARACAÖREN

Doç. Dr. Doğan NARİNÇ

Yrd. Doç. Dr. Aşkın GALİÇ

ABSTRACT

GENOMIC EVALUATION OF QUANTITATIVE TRAITS: FACTORS AFFECTING THE PREDICTION ACCURACY, AND AN UPPER BOUND FOR ACCURACY OF GBLUP

Emre KARAMAN

**PhD Thesis, in Animal Science
Supervisor : Prof. Dr. Mehmet Ziya FIRAT
October 2016, 66 pages**

This study aims at characterizing the asymptotic behavior of genomic prediction R^2 . Haplotypes derived from whole-genome sequence of 85 Caucasian individuals from the 1,000 Genomes Project were used to simulate random mating in a population of 10,000 individuals for 111 generations to create the LD structure in humans for a large number of individuals. To reduce computational demands, only SNPs within a 0.1 Morgan region of each of the first 5 chromosomes were used in simulations, and therefore, the total genome length simulated was 0.5 Morgan. When the genome length is 30 Morgan, to get the same genomic prediction R^2 as with a 0.5 Morgan genome would require a reference population 60 fold larger. Three scenarios were considered varying in minor allele frequency distributions of markers and QTL, for $h^2=0.8$ resembling height in humans. Total number of markers was 4,200 and QTL were 70 for each scenario. In this study, we considered the prediction accuracy in terms of an estimability problem, and thereby provided an upper bound for reliability of prediction, and thus, for prediction R^2 . Genomic prediction methods GBLUP, BayesB and BayesC were compared. Our results imply that variable selection methods BayesB and BayesC applied to a 30 Morgan genome have no advantage over GBLUP when the size of reference population was small (<6,000 individuals), but are superior as more individuals are included in the reference population. All methods become asymptotically equivalent in terms of prediction R^2 , which approaches genomic heritability when the size of the reference population reaches about half a million individuals.

KEYWORDS: Genomic heritability, Whole genome regression methods, Reliability, Genetic value

COMMITTEE: Prof. Dr. Mehmet Ziya FIRAT (Supervisor)

Prof. Dr. İbrahim CEMAL

Assoc. Prof. Dr. Burak KARACAÖREN

Assoc. Prof. Dr. Doğan NARİNÇ

Assist. Prof. Dr. Aşkın GALİÇ

ÖNSÖZ

Biyoteknoloji alanında yaşanan gelişmelere paralel olarak pek çok tür için tek nükleotid polimorfizmine (single nucleotide polymorphism-SNP) dayalı yüksek yoğunluklu marker panelleri geliştirilmiştir. Günümüzde; buğday ve mısır gibi çeşitli bitkiler ya da sığır, koyun, tavuk ve domuz gibi çiftlik hayvanlarının yanı sıra, insan ve fareler için de tüm genomu yayılan on binlerce SNP barındıran paneller kullanılabilir durumdadır.

Genomik bilginin elde edilebilir hale gelmesi, bitki ve hayvan ıslahında kullanılan istatistiksel yöntemlerin de değişmesini ve gelişmesini sağlamıştır. Örneğin, Amerika Birleşik Devletleri ve Kanada'da Holstein ırkı boğaların seçimi 2009 yılından bu yana genomik damızlık değer tahminlerine göre yapılmaktadır. Bireylerin tüm genomu yayılan ve sayıları yüz binleri bulan SNP markerler ile genotiplendirilmesinin önemli bir sonucu da insanlarda kişiselleştirilmiş tıp uygulamalarının olanaklı hale gelmesidir. Genom boyu ilişki analizleri (genome-wide association studies-GWAS), bazı hastalık ya da özelliklerle ilişkili kimi genleri tespit etmede başarılı olmuş olsa da, söz konusu çalışmalarla belirlenen bu genler bireyler arasındaki varyasyonun ancak çok küçük bir kısmını açıklamaktadır. İnsan boyu üzerine gerçekleştirilen bir çalışmada, araştırmacılar boy uzunluğunu etkilediği belirlenen sınırlı sayıda lokus yerine bireylere ait yaklaşık 300,000 SNP genotipini içeren bir panel kullanarak ilgili özellik için kalıtım derecesini %50 oranında tahmin etmeyi başarmışlardır.

Gerek bitki ve hayvan ıslahı uygulamalarında gerekse insanlar üzerindeki çalışmalarda kullanımı oldukça yeni olan ve «genomik değerlendirme» olarak adlandırdığımız çalışma konusunu seçmem için beni cesaretlendiren danışman hocam Prof. Dr. Mehmet Ziya Fırat'a teşekkür ederim. Kendisinin danışmanlığındaki lisans üstü eğitimim sürecinde edindiğim bilgi birikim yanında, O'nun bana ve çalışmalarına dair her zaman hissettirdiği inanç ve güven duygusu olmasa bu tez konusunu seçmem de mümkün olmazdı.

Konuyla ilgili araştırmalar yapmak amacıyla çalışma gruplarını ziyaret ettiğim süreçte bana her konuda destek olan ve sahip olduğu alçakgönüllülükle tüm sorularımı usanmadan yanıtlayan Prof. Dr. Rohan L. Fernando'ya teşekkür ederim. Kendisi her zaman benim için özel insanlardan biri olarak kalacaktır. Özellikle, bu çalışmanın bir kısmının yayına dönüştürülmesi (Karaman vd 2016) sürecindeki destek ve yönlendirmeleri için Dorian J. Garrick'e teşekkür ederim. Yazmış olduğu simülasyon programını koşulsuz şartsız benimle paylaşan ve orada bulunduğum süreçte bana her konuda içtenlikle yardımcı olan Hao Cheng'e teşekkür ederim. Ayrıca, ilgili çalışma grubunda yer alan ve gerek okul gerekse okul dışındaki hayatımı oldukça kolaylaştıran Wan-Ling Hsu, Jian Zeng ve Xiaochen Sun'a teşekkürlerimi sunarım.

Lisansüstü eğitimine başladığım günden bu yana her zaman desteğini gördüğüm ve özellikle de yürütücülüğünü yaptığı projelerde yer almamı sağlayarak

kendimi bilimsel açıdan geliřtirmem için bana imkanlar sunan Prof. Dr. Tlin Aksoy'a teřekkr ederim. Bildiklerini, ğrendiklerini ve sahip olduklarını her zaman benimle paylařan ve gerek eleřtirileri gerekse nerileri ile daha donanımlı bir bilim insanı olmam için beni her daim ynlendiren Doç. Dr. Doęan Nariñ'e teřekkrlerimi sunarım. Kendisi ile geçirdięim zamanların gelecekteki akademik yařantım zerinde de nemli etkileri olacaęından eminim. Ayrıca, farklı bir disiplinden gelerek bařladıęım lisansst eęitim hayatım boyunca her konuda bana destek olan blm hocalarım ve mesai arkadaşlarıma teřekkr ederim.

Bugnlere gelmem için okça fedakarlık yapan; para, zaman ve en nemlisi de emek harcayan, anne ve babama teřekkr ederim. En ok da ahlaklı ve erdemli bir insan olmanın her trl bařarıdan daha nemli olduęunu bana ęrettikleri için...

Son olarak, uzun bir yolu birlikte yrdęm, belki de birlikte bydęm on altı yıllık hayat arkadaşım, eřim Gl Tanker Karaman'a bana verdięi kořulsuz destek için ok teřekkr ederim.

İÇİNDEKİLER

ÖZET	i
ABSTRACT	iii
ÖNSÖZ	vi
İÇİNDEKİLER	viii
SİMGELER ve KISALTMALAR DİZİNİ	x
ŞEKİLLER DİZİNİ	x
ÇİZELGELER DİZİNİ	xi
1. GİRİŞ	1
2. KURAMSAL BİLGİLER VE KAYNAK TARAMALARI	5
2.1. Bazı Temel Kavramlar	5
2.1.1. DNA, gen, allel	5
2.1.2. Rekombinasyon, mutasyon, bağlantı dengesizliği	6
2.1.3. Genetik değer, kalıtım derecesi, genomik kalıtım derecesi	8
2.1.4. Regresyon analizi	10
2.1.4.1. Basit doğrusal regresyon	10
2.1.4.2. Çoklu doğrusal regresyon	11
2.1.5. Genomik değerlendirme	11
2.1.5.1. Tek aşamalı yöntemler	12
2.1.5.2. İki aşamalı yöntemler	13
2.1.5.2.1 Bayesci regresyon (BR)	14
2.1.5.2.2 BayesA	16
2.1.5.2.3 BayesB	17
2.1.5.2.4 BayesC	19
2.1.6. Genomik değerlendirmede isabet	20
2.1.7. Genomik değerlendirme çalışmaları ve isabeti etkileyen faktörler	23
3. MATERYAL VE METOD	26
3.1. Veri Setleri ve Simülasyonlar	26
3.2. Referans ve Test Popülasyonları	27
3.3. Fenotipik Değerlerin Simülasyonu	27
3.4. Marker Etkilerinin Tahmini	28
3.5. Eklemeli Genetik Değerlerin Tahmini ve İsabet	28
3.6. Tahmin İsabeti ile Sabit Etkiler Modelinde Tahmin Edilebilirlik Arasındaki Bağlantı	29
4. BULGULAR	32
4.1. Panelde Yalnızca Markerlerin Bulunduğu Durum	32
4.1.1. GBLUP yöntemine ilişkin tahminler	32
4.1.2. Tahmin yöntemlerinin karşılaştırılması	37
4.2. Panelde Markerler ve QTL'lerin Bulunduğu Durum	37
4.2.1. GBLUP yöntemine ilişkin tahminler	37
4.2.2. Tahmin yöntemlerinin karşılaştırılması	40
4.3. Gerçekleşen İlişkiler ve Tahmin Edilebilirlik	40
5. TARTIŞMA	45

6. SONUÇ	51
7. KAYNAKLAR	53
8. EKLER	61
Ek-1	61
Ek-2	62
Ek-3	64
Ek-4	65
ÖZGEÇMİŞ	

SİMGELER ve KISALTMALAR DİZİNİ

Simgeler

A	Akrabalık matrisi
c	Rekombinasyon oranı
cM	Centi Morgan
G	Genomik ilişki matrisi
h^2	Kalıtım derecesi
h_M^2	Genomik kalıtım derecesi
L	Ortalama kromozom uzunluğu
M	Morgan
M_e	Etkin kromozom bölgelerinin sayısı
M_u	Mutasyon oranı
N	Normal dağılım
N_e	Etkin popülasyon büyüklüğü
n_R	Referans popülasyondaki birey sayısı
n_T	Test popülasyondaki birey sayısı
p	Marker sayısı
q^2	Genetik varyansın markerler tarafından açıklanan oranı
R^2	Fenotipik değer ve tahmin edilen genetik değer arasındaki korelasyonun karesi
r	İsabet derecesi
r^2	Güvenilirlik

Kısaltmalar

BLUP	Best Linear Unbiased Prediction - En İyi Doğrusal Sapmasız Tahmin
BR	Bayesian Regression - Bayesci Regresyon
BRR	Bayesian Ridge Regression - Bayesci Ridge Regresyon
COR	Correlation - Korelasyon
COV	Covariance - Kovaryans
DI	Düşük ilişki grubu
EKK	En küçük kareler
GBLUP	Genomic Best Linear Unbiased Prediction - Genomik En İyi Doğrusal Sapmasız Tahmin
LD	Linkage Disequilibrium - Bağlantı Dengesizliği
MAF	Minör allel frekansı
MCMC	Marcov Chain Monte Carlo
PBLUP	Pedigree based BLUP - Pedigriye dayalı BLUP
QTL	Quantitative Trait Loci - Kantitatif Özellik Lokusu
RP	Referans popülasyon
TP	Test popülasyonu
SNP	Single Nucleotide Polymorphism - Tek Nükleotid Polimorfizmi
VAR	Variance - Varyans
YI	Yüksek ilişki grubu

ŞEKİLLER DİZİNİ

Şekil 2.1.	Genom	6
Şekil 2.2.	Aile içi ve popülasyon düzeyi LD için hipotetik bir durum	7
Şekil 2.3.	Kayıp kalıtım derecesi	10
Şekil 2.4.	Genomik değerlendirmede kullanılan yaklaşımlar	12
Şekil 3.1.	Simülasyon şeması	26
Şekil 4.1.	Panelde yalnızca markerler bulunduğunda GBLUP yönteminin isabeti	33
Şekil 4.2.	Panelde yalnız markerler bulunduğunda tahmin yöntemlerinin isabeti	38
Şekil 4.3.	Panelde markerlar ve QTL'ler bulunduğunda GBLUP yönteminin isabeti	39
Şekil 4.4.	Panelde marker ve QTL'ler bulunduğunda tahmin yöntemlerinin isabeti	41
Şekil 5.1.	Senaryo 1'deki bir tekerrürde ($n_R = 8,000$) test popülasyonundaki bir bireyin referans popülasyondaki tüm bireylerle marker ve QTL düzeyindeki ilişkilerine ait serpilme grafiği	46
Şekil 5.2.	Senaryo 1'deki bir tekerrürde BayesB yöntemi için markerlerin model frekansları	49

ÇİZELGELER DİZİNİ

Çizelge 1.1. Bazı türler için mevcut SNP panelleri	1
Çizelge 3.1. Her bir senaryodaki MAF dağılımları	27
Çizelge 4.1. Senaryo 1'e ait sonuçlar	34
Çizelge 4.2. Senaryo 2'ye ait sonuçlar	35
Çizelge 4.3. Senaryo 3'e ait sonuçlar	36
Çizelge 4.4. Farklı genomik ilişki grupları için tahmin isabeti ve tahmin edilebilirlik değerleri	43

1. GİRİŞ

Moleküler genetik alanında yaşanan önemli gelişmeler ve teknolojik ilerlemeler sayesinde genetik değerlendirme çalışmaları “genomik çağ” adı verilen yeni bir döneme doğru evrilmiştir. Yirmi birinci yüzyılın başlarında insan genomu dizileme projesinin (Human Genome Project) tamamlanmasıyla başlayan süreçte edinilen bilgi, deneyim ve ortaya konan yeni yöntem ve teknolojiler diğer türlere de aktarılacak genomik değerlendirme çağının başlamasını sağlamıştır (Fan vd 2010). Aday gen yaklaşımı ya da gen haritalamacılığı ile tespit edilmeye çalışılan, canlılarda karmaşık yapıdaki özellikleri ortaya çıkartan genler, genom boyu ilişki analizi çalışmalarıyla tek nükleotid polimorfizmine (single nucleotide polymorphism-SNP) bağlı olarak daha hızlı bir şekilde belirlenebilir hale gelmiştir (Goddard ve Hayes 2009).

Pek çok tür için ticari olarak piyasaya sürülen SNP panelleri (Çizelge 1.1) genomik değerlendirme çalışmalarının hız kazanmasını sağlamıştır. Adından da anlaşılacağı gibi genomik değerlendirme hayvanlarda (Meuwissen vd 2001; Goddard ve Hayes 2007; Hayes vd 2009b; Wolc vd 2013) ve bitkilerde (Bernardo ve Yu 2007; Heffner vd 2009; Jannink vd 2010) genetik değer tahmininde veya insanlarda kişiselleştirilmiş tıp uygulamalarında hastalığa yatkınlığı tespit etmede genomik bilginin kullanılması anlamına gelmektedir. Genomik değerlendirme uygulamalarında fenotipik değerleri (y) ve genotipleri bilinen bir referans popülasyon (RP) yardımıyla marker etkileri tahmin edilmekte ve yalnızca genotiplerine dair bilginin bulunduğu test popülasyonu (TP) bireyleri için genetik değer tahmininde bulunmaktadır (Meuwissen vd 2001).

Genomik değerlendirmede sağlanacak isabetin çok sayıda faktörden etkilendiği bilinmektedir. Bunlar; marker etkilerinin tahmininde kullanılan yöntem (Habier vd 2007; Luan vd 2009), özelliğin kalıtım derecesi (h^2) ve genetik yapısı (Luan vd 2009; Daetwyler vd 2010b), referans popülasyonun büyüklüğü (n_R) ve

Çizelge 1.1. Bazı türler için mevcut SNP panelleri

Tür	Firma	Panel adı	SNP Sayısı
İnsan	Affymetrix	Genome-Wide Human SNP Array 6.0	~900,000
Sığır	Illumina	BovineHD	~770,000
Koyun	Illumina	OvineHD array	~600,000
Keçi	Illumina	GoatSNP50	~50,000
Tavuk	Affymetrix	Axiom® Genome-Wide Chicken Genotyping Array	~600,000
Domuz	Affymetrix	Axiom® Porcine Genotyping Array	~650,000
Köpek	Illumina	CanineHD	~170,000
At	Affymetrix	Axiom® Equine Genotyping Array	~650,000
Mısır	Illumina	MaizeSNP50	~50,000
Somon Balığı	Affymetrix	Axiom® Salmon Genotyping Array	12,000
Buğday	Affymetrix	Axiom® Wheat HD Genotyping Array	~800,000

SNP: Single nucleotide polimorphism - Tek nükleotid polimorfizmi
(Affymetrix Inc 2016; Illumina Inc 2016)

yapısı (Goddard 2009; Daetwyler vd 2010b; Habier vd 2010b; Pszczola vd 2011; Makowsky vd 2011; Clark vd 2012) olarak sıralanabilir. Kullanılacak tahmin yöntemi ile referans popülasyonun yapısı ve büyüklüğü, üzerinde çalışılan özelliğin kalıtım derecesi ve genetik yapısı dikkate alınarak belirlenebilir.

Genomik değerlendirmede uygulamada yaşanan zorluklardan biri RP’u oluşturan birey sayısının etkisi tahmin edilecek markerlerin sayısından (p) çok daha az olmasıdır, $n_R \ll p$. Bu sorunun üstesinden gelmenin bir yolu, marker etkilerini, β_j ’leri, tahmin etmek amacıyla, marker etkilerine dair ön bilgiyi gözlenen verilerle bir araya getiren Bayesçi regresyonu kullanmaktır (Meuwissen vd 2001). Marker etkileri için varsayımları önsel dağılımlar bakımından birbirinden ayrılan çok sayıda Bayesçi regresyon yaklaşımı bulunmaktadır. Yaygın olarak marker etkilerinin tüm lokuslar için sabit bir varyansla normal dağılımdan geldiği varsayılmaktadır, $\beta_j \sim N(0, \sigma_\beta^2)$. $\sigma_\beta^2/\sigma_e^2$ oranı bilindiğinde bunun ridge regresyon ve en iyi doğrusal yansız tahmin (best linear unbiased prediction-BLUP) ile eşdeğer olduğu gösterilmiştir (de los Campos vd 2013a). Bayesçi regresyon, ridge regresyon ve BLUP ile olan bu ilişkisi dolayısıyla Bayesçi ridge regresyon (BRR) veya «random regression BLUP» (RR-BLUP) olarak adlandırılmaktadır. RR-BLUP yönteminden elde edilen genomik değerlendirme sonuçlarının, akrabalık matrisinin tüm genoma yayılan markerler yolu ile hesaplandığı ($\mathbf{G}$) birey modeli sonuçları (genomik best linear unbiased prediction-GBLUP) ile aynı olduğu gösterilmiştir (Nejati-Javaremi 1997; Fernando 1998; Habier vd 2007; van Raden 2008; Strandén ve Garrick 2009).

GBLUP yönteminin ıslah çalışmalarında yaygın olarak kullanılmasının çeşitli nedenleri vardır. Bunlar; (i) çiftlik hayvanlarında seleksiyon çalışmaları uzun yıllardır birey modeli altında BLUP (Henderson 1984) sonuçlarına dayalı olarak yapıldığından GBLUP yönteminin mevcut bilgisayar yazılımları ile kolayca uygulanabilir olması, (ii) BLUP yönteminde tahmin hatalarının varyanslarının hesaplanabilmesi için teoreminin mevcut olması ve (iii) insanlar, bitkiler ve hayvanlarda ilgilenilen pek çok özelliğin karmaşık yapıda olması ve tüm genoma yayılan ve her biri küçük etkiye sahip çok sayıda gen tarafından determine edilmesi (Hayes ve Goddard 2001; Meuwissen ve Goddard 2010; Yang vd 2010) dolayısıyla tüm lokuslar için eşit varyans varsayımının uygun olması (Meuwissen vd 2001) olarak sıralanabilir. Bu varsayım sağlanmadığında, β_j ’nin π olasılıkla sıfır olduğunu, $(1 - \pi)$ olasılıkla t dağılımı (BayesB) (Meuwissen vd 2001) veya normal dağılımdan (BayesC) (Kızılkaya vd 2010; Habier vd 2010a) geldiğini varsayan Bayesçi karışım modelleri kullanılabilir.

Bireylerin gerçek eklemeli genetik değerleri (u) ile bu değerlerin tahmini ($\hat{u}$) arasındaki korelasyonun karesi, $r_{(u,\hat{u})}^2$, güvenilirlik olarak adlandırılmakta ve genetik değer tahmininde isabeti değerlendirmenin bir ölçüsü olarak kullanılmaktadır. Daetwyler vd (2008) ve Goddard (2009); etkin popülasyon büyüklüğü (N_e), n_R , h^2 ve etkili kromozom segmentlerinin sayısını (M_e) kullanarak tahmin isabeti için bazı yakınsamalar önermişlerdir. Ancak, her iki çalışmada önerilen yakınsamalar marker ve QTL çiftleri arasında kusursuz bir bağlantı dengesizliği (linkage disequilibrium-LD) olduğu varsayımı altında türetilmiştir. Goddard vd (2011)

markerler ve QTL'ler arasında tam bir LD bulunmayabileceğini de hesaba katarak tahmin isabeti için yeni bir yakınsama ortaya koymuştur. Söz konusu çalışmadan hareketle, genomik değerlendirmede r^2 değeri, q^2 genetik varyansın markerler tarafından açıklanan oranını göstermek üzere, $q^2[n_R h_M^2 / (n_R h_M^2 + M_e / q^2)]$ eşitliği ile tahmin edilebilir. Uygulamada, bireylere ait gerçek eklemeli genetik değerler bilinmediğinden, isabetin bir ölçüsü olarak bireylerin fenotipik değerleri, y , ile tahmin edilen eklemeli genetik değerleri, $\hat{u}$, arasındaki korelasyonun karesi $R_{(y,\hat{u})}^2$ kullanılmaktadır (Buradan sonra R^2 olarak anılacaktır). Bu durumda, Goddard vd (2011) tarafından önerilen yakınsama:

$$R^2 \approx h_M^2 [n_R h_M^2 / (n_R h_M^2 + M_e)] \quad (1.1)$$

şeklinde olacaktır (Ek-1). Burada h_M^2 genomik kalıtım derecesi, bir başka ifade ile markerler tarafından açıklanan varyansın fenotipik varyansa oranıdır (de los Campos vd 2013b).

de los Campos vd (2013b), gerçek veriler ve simülasyonlardan yararlanarak marker ve QTL'ler için farklı minör allel frekansları (minor allele frequency-MAF) bakımından değişen çeşitli senaryolar altında ve GBLUP yöntemini kullanarak, insanlarda karmaşık yapıdaki özellikler için h_M^2 ve R^2 arasındaki ilişkiyi incelemiştir. Söz konusu çalışmada, veri setleri birbiri ile akraba olmayan bireyler içerdiğinde ve markerler ile QTL'ler benzer MAF dağılımlarına sahip olduğunda elde edilen R^2 ve h_M^2 değerleri sırası ile 0.071 ve 0.737, QTL'lerin MAF dağılımları markerlerden farkı olduğunda elde edilen R^2 ve h_M^2 değerleri ise sırası ile 0.049 ve 0.573 olarak verilmiştir. de los Campos vd (2013b) TP bireyleri ile RP bireyleri ilişkili olmadığında genomik kalıtım derecesi h_M^2 'nin, R^2 için iyi bir gösterge olmadığını bildirmişlerdir. Araştırmacılar, b markerlerden elde edilen genetik ilişki vektörü ile QTL'lerden elde edilen genetik ilişki vektörü arasında kurulan regresyon eğrisinin eğim parametresinin ortalama değeri olmak üzere, R^2 için üst sınırın $[1 - (1 - b)^2]h^2$ olduğunu iddia etmişlerdir. Aynı zamanda, söz konusu çalışmanın sonuçları ışığında, R^2 için üst sınırın aralarında akrabalık bulunmayan bireyler için h^2 'nin %20'si olduğu belirtilmiştir. Bu sonuçların aksine, yakınsama (1.1) R^2 için asimptotik değerin h_M^2 olduğunu göstermektedir (Ek-1).

İnsanlara ilişkin olarak etkin örnek büyüklüğünü, $N_e = 10,000$, ortalama kromozom uzunluğunu (L) 1.57 Morgan, kromozom sayısını (k) 23 alacak olursak, $M_e = \frac{2N_e L k}{\log(N_e L)} \approx 7 \times 10^4$ olacaktır (Takahata 1993; Goddard vd 2011). Bu durumda, yakınsama (1.1)'den $n_R = 5,300$ olmak üzere $h_M^2 = 0.737$ ve $h_M^2 = 0.573$ için tahmin edilen R^2 değerleri de sırası ile 0.037 ve 0.022 olacaktır. R^2 için tahmin edilen bu değerler, aralarında ilişki bulunmayan bireyler için de los Campos vd (2013b) tarafından bildirilen değerlere benzerdir. Yakınsama (1.1)'e göre R^2 'nin genomik kalıtım derecesinin, h_M^2 'nin, %90'ına ulaşması için yaklaşık olarak yarım milyondan fazla birey barındıran bir RP'a ihtiyaç vardır.

Literatürde, yüksek isabetli genetik değer tahmini için en uygun referans popülasyonun nasıl olması gerektiğine (Pszczola vd 2011; Perez-Cabal vd 2012) ve TP ve RP bireyleri arasındaki ilişkinin tahminlerin isabeti üzerindeki etkisine dair çok sayıda çalışma bulunmaktadır (Habier vd 2007; Legarra vd 2008; Habier vd 2010b; Pszczola vd 2011; Makowsky vd 2011; Clark vd 2011). Ancak, pek çok çalışmada bireyler arasındaki genetik ilişkiler akrabalık durumları üzerinden yapılmakta olup bireylerin bu şekilde elde edilen akrabalık bilgileri de ne kadar geçmişe dönük kayıt tutulduğuna ya da bu ilişki için hangi katsayının (ilişki katsayılarının kareli ortalaması, ortalama ilişki vb.) kullanıldığına bağlıdır. Buna karşın, tüm genoma yayılan markerler yolu ile hesaplanan $\mathbf{G}$ matrisi bireyler arasındaki gerçek akrabalığı daha iyi ortaya koyacaktır. Diğer yandan, TP bireyleri tahmin edilen genetik değerlerine, $\hat{u}$, göre sıralanmak istendiğinde TP ve RP bireyleri arasındaki ikili ilişkiler de yanıltıcı olacaktır.

Bu çalışmada, R^2 için de los Campos vd (2013b) tarafından verilen üst limitin mi geçerli olduğu, yoksa R^2 'nin yakınsama (1.1)'den görüleceği gibi n_R arttıkça h_M^2 'ye mi yaklaştığını test etmek için simülasyonlardan yararlanılacaktır. Bununla birlikte, Bayesçi genomik tahmin yöntemlerinin karmaşık özellikler için GBLUP'a göre daha yüksek isabet sağlayabileceği iddiası da (de los Campos vd 2013b) sınanacaktır. Ayrıca, genomik değerlendirme bir tahmin edilebilirlik sorunu olarak ele alınarak tahmin isabeti, R^2 , için bir üst sınır ortaya konacaktır.

2. KURAMSAL BİLGİLER VE KAYNAK TARAMALARI

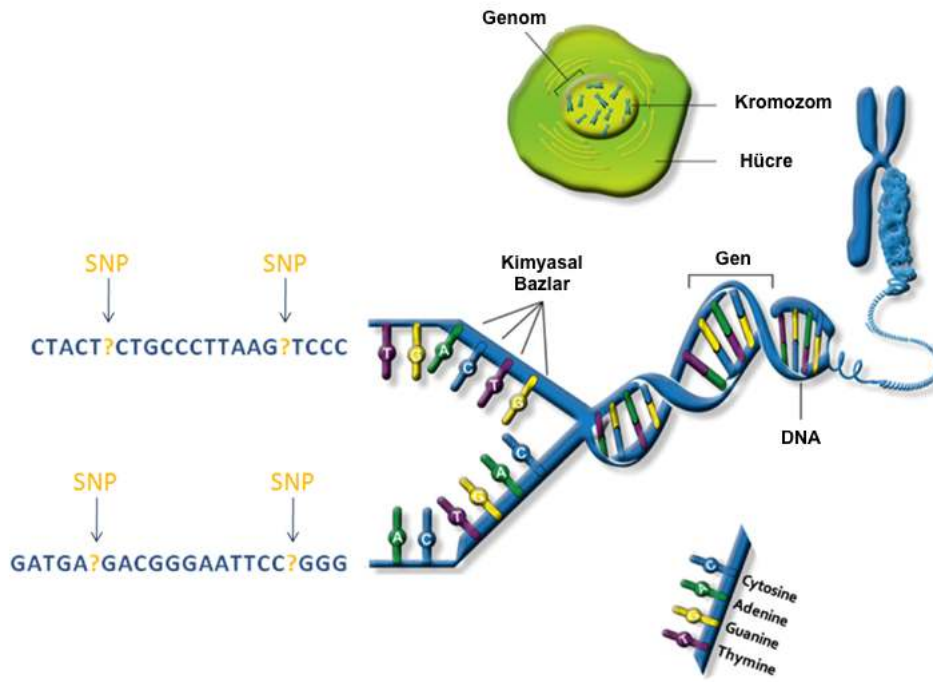
2.1. Bazı Temel Kavramlar

2.1.1. DNA, gen, allel

Tüm canlı organizmalar hücrelerden meydana gelmekte olup, her bir hücre, içerisinde dört temel bileşenden (A, C, G ve T) oluşan ve deoksiribonükleik asit (DNA) adı verilen yapıyı barındırır. DNA, üreme yolu ile genetik yapının bir sonraki nesile aktarılmasını sağlamasının yanında, canlılara ait özelliklerin veya bir başka ifade ile proteinlerin kodlandığı bir yapıdır (Klug ve Cummings 2003). Gen, bir proteinin üretilmesinde görev alan DNA dizisinin adıdır. Bir geni ya da lokusu meydana getiren bu DNA dizisi bireyler arasında farklılık gösterebilir ve bu dizinin, ya da diğer adıyla genin, alternatif formları allel olarak adlandırılmaktadır. Bireylerin sahip oldukları genlerdeki farklılık, üretilecek olan protein miktarını ya da türünü belirleyerek bireyler arasındaki farklılığa neden olmaktadır. Bu genler kalitatif özellikleri belirleyebileceği gibi kantitatif özellikleri de belirleyebilir. Allel etkileri; resesif, eklemeli ve dominant olabilir. Eğer allel etkisi eklemeli ise bireyin her iki ebeveyninden alınacak allellerin toplam etkisi söz konusudur (Falconer ve Mackay 1996). Tüm canlıların sahip oldukları genler her bir hücre içerisinde yer alan kromozomlara dağılmış olup, o bireyin genomunu meydana getirmektedir (Şekil 2.1).

Hemen hemen bütün türlerde, moleküler biyoloji ve kantitatif genetik araçları kullanarak kromozomlarda çeşitli özellikler ile ilişkili bölgeleri belirlemek ve bu bölgelerdeki DNA dizisi farklılıklarını ortaya çıkarmak üzerine çalışılmaktadır (Goddard ve Hayes 2009). Genomda belirli bir kromozomda yer alan ve bir verim özelliği ile ilişkili bu DNA dizisi **marker** olarak adlandırılmaktadır. Hayvan ıslahında, kantitatif bir özellik üzerinde etkisi olan lokusla (quantitative trait loci-**QTL**) bağlantı dengesizliği (linkage disequilibrium-**LD**) içerisinde olan markerler kullanılarak seleksiyonda ve dolayısıyla damızlık değer tahmininde isabetinin artırılması amaçlanmaktadır (Meuwissen vd 2001). Genetik markerlerin çok sayıda farklı türü olmakla birlikte, bir lokustaki allellerin DNA dizisinde tek bir baz çifti bakımından farklılık gösterdiği **SNP** (single nucleotide polymorphism) markerler yaygın olarak kullanılmaktadır (Meuwissen vd 2013).

Bir bireyin sahip olduğu marker allelleri genotip, çeşitli araçlar kullanarak bireyin taşıdığı bu marker allellerinin belirlenmesine genotiplendirme adı verilmektedir. Her birey bir gen bakımından dişi ve erkek ebeveynlerinden yalnızca o genin bir formunu alacağından, belirli bir lokusta yalnızca iki adet allel bulunduracaktır. Çiftlik hayvanlarında, bitkilerde ya da insanlarda pek çok özellik veya protein çok sayıda genin ve içinde bulunan çevrenin etkisi ile ortaya çıkmaktadır ve karmaşık yapıdadır (Goddard ve Hayes 2009).

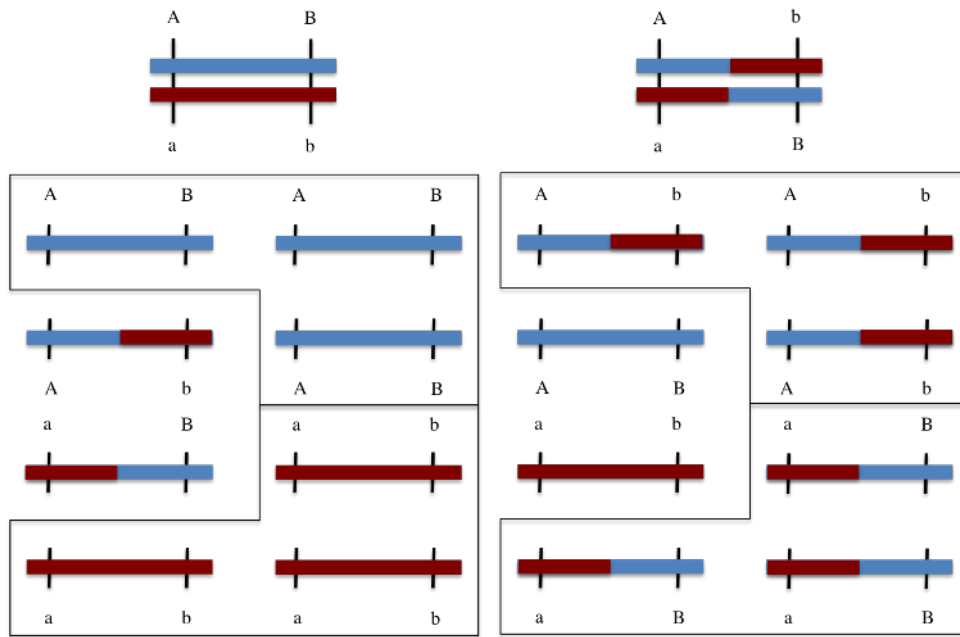


Şekil 2.1. Genom
(Anonim 2016)

2.1.2. Rekombinasyon, mutasyon, bağlantı dengesizliği

Bağlantı dengesizliği, LD, farklı lokuslarda yer alan allel genler arasındaki ilişkiyi ifade etmek için kullanılmaktadır (Legarra 2014). Örnek olarak A ve B allellerini bulunduran iki lokusu ele alalım ve ilgili popülasyonda A ve B allellerine ilişkin frekansları sırası ile p_A ve p_B ile gösterelim. Bu iki lokus ilişkisiz olduğunda AB haplotipinin frekansının p_{APB} olması beklenmektedir. AB haplotipinin frekansının p_{APB} 'dan sapma göstermesi durumunda söz konusu lokusların bağlantı dengesizliği içerisinde olduğu sonucuna ulaşılır (Falconer ve Mackay 1996). LD; doğal seleksiyon, mutasyon oranı (M_u), rekombinasyon oranı (c) ve göç gibi faktörlerin etkisi altındadır. Rekombinasyon, sözü edilen faktörler içerisinde LD'yi en fazla etkileyen faktördür. Bir kromozomda birbirine yakın iki lokus arasında rekombinasyon nadir görülürken, birbirinden uzakta bulunan lokuslar arasında görülme ihtimali daha fazladır. Çünkü, daha dar bir kromozom segmentinde crossing-over meydana gelme ihtimali, daha geniş bir kromozom segmentine göre daha düşüktür. İki marker arasında rekombinasyon olduğunda her iki lokusta bulunan allel genlerin bağımsız olma olasılığı artar ve bu iki lokus arasındaki LD azalır. Bu bilgiler ışığında, birbirine yakın iki lokus arasındaki LD'nin birbirine uzak iki lokus arasındaki LD'den daha yüksek olacağı açıktır (de Roos 2008).

Bir popülasyonda gözlenen LD, ilgili popülasyondaki etkin popülasyon büyüklüğüne (N_e) bağlıdır. Düşük bir N_e mevcut popülasyondaki allellerin yakın bir kuşaktaki ortak bir atadan geldiğine yani, bireylerin kromozomlarının büyük



Şekil 2.2. Aile içi ve popülasyon düzeyi LD için hipotetik bir durum

ölçüde paylaşıldığına işaret etmektedir (Zhang vd 2011). Bu tür bir popülasyonda uzak lokuslar arasında da LD gözlenmesi mümkündür (Goddard ve Hayes 2009). Sığır türünde N_e , evcilleştirme öncesi ve sonrası için yaklaşık olarak 50,000 ve 100 olarak tahmin edilmiştir. Bu durum insanlarda tam tersi olup, son 10,000 yılda 3,000'den yaklaşık 10,000 değerine yükseldiği tahmin edilmektedir (Goddard ve Hayes 2009). Bu durumda insanlar için yakın lokuslar arasında beklenen LD sığırlardakine benzer iken, birbirine uzak lokuslar için beklenen LD son derece düşük olacaktır. Bu nedenle, genomik değerlendirme amacıyla gereken marker yoğunluğu pek çok evcil hayvan türü için insanlara kıyasla daha düşük olacaktır. Birbirinden d birim uzaklıkta bulunan iki lokus için LD şu şekilde tahmin edilebilir (Sved 1971):

$$LD = \frac{1}{4N_e d + 1} \quad (2.1)$$

Yukarıdaki formülden görüleceği gibi, N_e arttıkça LD azalmakta ve buna bağlı olarak da genetik değer tahmininde isabetin de azalması beklenmektedir.

Allel frekanslarının dağılımı izole edilmiş bir popülasyonda ya da bir aile içerisinde incelendiğinde daha yüksek LD'ye işaret edecektir. Çünkü, izole edilmiş popülasyonlarda ya da aile içerisinde ebeveynlere ait kromozomlar yukarıda da belirtildiği şekilde büyük ölçüde korunacaktır. Bunun nedeni bir kuşakta gerçekleşebilecek rekombinasyon sayısının oldukça az olmasıdır (Legarra 2014). Şekil 2.2'de iki birey için 0.25 rekombinasyon oranı üzerinden düzenlenmiş hipotetik bir durum gösterilmiştir (Legarra 2014).

2.1.3. Genetik değer, kalıtım derecesi, genomik kalıtım derecesi

Klasik kantitatif teoriye göre bireylerin sahip oldukları eklemeli genetik etkiler ilgili özelliği determine eden lokuslarda yer alan gen etkilerinin doğrusal bir fonksiyonudur (Falconer ve Mackay 1996). Ancak, uygulamada QTL'ler bilinmemekte ve bu nedenle QTL'lerle bağlantı dengesizliği içinde olan markerler kullanılarak bir regresyon modeli uydurulmaktadır (Meuwissen vd 2001). QTL ve marker etkileri (α ve β) sabit etkiler olsun. Aynı zamanda QTL ve marker genotipleri için varyans-kovaryans matrisleri, $\mathbf{V}_{HH}$ ve $\mathbf{V}_{XX}$, popülasyon parametreleri olsun. Klasik kantitatif teoriye göre şansa bağlı olma durumu bireylerin QTL ve marker genotiplerinin şansa bağlı olmasından kaynaklanmaktadır.

Kantitatif bir özellik, y_i , eşitlik (2.2)'deki gibi modellenenebilir olsun:

$$y_i = \mathbf{h}_i' \alpha + e_i \quad (2.2)$$

Burada $\mathbf{h}_i$ referans popülasyonundaki i 'inci bireyin QTL genotiplerini içeren vektör ve α , QTL etkilerini içeren vektördür. $\mathbf{q}_i$ test popülasyonundaki i 'inci bireyin genotiplerini içeren vektör olmak üzere, söz konusu bireyin genetik değeri $u_i = \mathbf{q}_i' \alpha$ ve bunun en küçük kareler ve ridge regresyon tahminleri de sırası ile $\hat{u}_i = \mathbf{q}_i' \hat{\alpha}$ ve $\tilde{u}_i = \mathbf{q}_i' \tilde{\alpha}$ şeklinde gösterilebilir.

Uygulamada QTL'ler bilinmemekte ve genetik değer tahminleri marker genotipleri kullanılarak yapılmaktadır. Bu durumda kullanılan model,

$$y_i = \mathbf{x}_i' \beta + \epsilon_i \quad (2.3)$$

şeklindedir. Burada, $\mathbf{x}_i$ RP'daki i 'inci bireyin genotiplerini içeren vektör ve β marker etkilerini içeren vektördür. $\mathbf{k}_i$ test popülasyonundaki i 'inci bireyin marker genotiplerini içeren vektör olmak üzere, söz konusu bireyin genetik değeri $u_i = \mathbf{q}_i' \alpha$, ancak bunun en küçük kareler ve ridge regresyon tahminleri sırası ile $\hat{u}_i = \mathbf{k}_i' \hat{\beta}$ ve $\tilde{u}_i = \mathbf{k}_i' \tilde{\beta}$ olacaktır.

En küçük kareler tahmincisi kullanılarak $\mathbf{k}_i' \beta = \mathbf{k}_i' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{h}_i' \alpha$ yazılabilir. Bu durumda,

$$\beta = \mathbf{V}_{XX}^{-1} \mathbf{V}_{XH} \alpha. \quad (2.4)$$

olacaktır.

Eşitlik (2.2)'den, fenotipik varyans:

$$\begin{aligned} \text{Var}(y_i | \mathbf{H}) &= \text{Var}(\mathbf{h}_i' \alpha | \mathbf{H}) + \text{Var}(e_i) \\ \sigma_y^2 &= \sigma_a^2 + \sigma_e^2 \end{aligned} \quad (2.5)$$

şeklinde yazılabilir. Burada, $\sigma_a^2 = \boldsymbol{\alpha}'\mathbf{V}_{HH}\boldsymbol{\alpha}$ olup eklemeli genetik varyansı temsil etmektedir. Eklemeli genetik varyansın fenotipik varyansa oranı kalıtım derecesi, h^2 , olarak adlandırılmaktadır:

$$h^2 = \frac{\sigma_a^2}{\sigma_y^2} = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2} = \frac{\boldsymbol{\alpha}'\mathbf{V}_{HH}\boldsymbol{\alpha}}{\boldsymbol{\alpha}'\mathbf{V}_{HH}\boldsymbol{\alpha} + \sigma_e^2} \quad (2.6)$$

Benzer şekilde, fenotipik varyansı marker etkileri modelinden (2.3),

$$\begin{aligned} Var(y_i | \mathbf{X}) &= Var(\mathbf{x}_i'\boldsymbol{\beta} | \mathbf{X}) + Var(\epsilon_i) \\ \sigma_y^2 &= \sigma_q^2 + \sigma_e^2 \end{aligned} \quad (2.7)$$

şeklinde yazabilir. Burada, $\sigma_q^2 = \boldsymbol{\beta}'\mathbf{V}_{XX}\boldsymbol{\beta}$ ya da (2.4)'den $\sigma_q^2 = \boldsymbol{\alpha}'\mathbf{V}_{HX}\mathbf{V}_{XX}^{-1}\mathbf{V}_{XH}\boldsymbol{\alpha}$ olup genomik varyansı verir. Markerler tarafından açıklanan varyansın toplam varyans içindeki payı genomik kalıtım derecesi, h_M^2 , olarak adlandırılmaktadır:

$$h_M^2 = \frac{\sigma_q^2}{\sigma_y^2} = \frac{\sigma_q^2}{\sigma_q^2 + \sigma_e^2} = \frac{\boldsymbol{\beta}'\mathbf{V}_{XX}\boldsymbol{\beta}}{\boldsymbol{\beta}'\mathbf{V}_{XX}\boldsymbol{\beta} + \sigma_e^2} = \frac{\boldsymbol{\alpha}'\mathbf{V}_{HX}\mathbf{V}_{XX}^{-1}\mathbf{V}_{XH}\boldsymbol{\alpha}}{\boldsymbol{\alpha}'\mathbf{V}_{HX}\mathbf{V}_{XX}^{-1}\mathbf{V}_{XH}\boldsymbol{\alpha} + \sigma_e^2} \quad (2.8)$$

$\frac{\sigma_q^2}{\sigma_a^2}$ eklemeli genetik varyansın markerler tarafından açıklanan oranı olup q^2 ile gösterilir (Dekkers 2007).

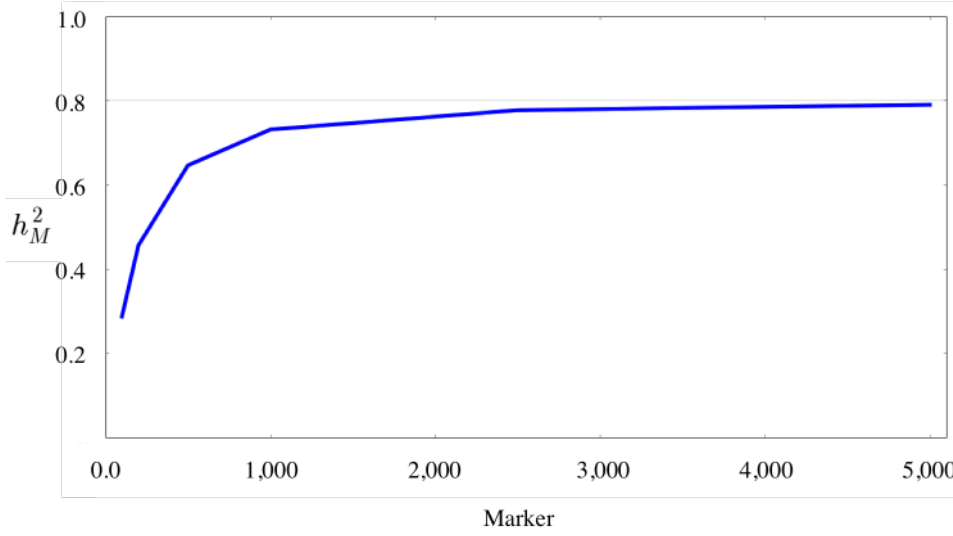
Eşitlik (2.5) ve eşitlik (2.7)'deki sonuçları kullanarak,

$$\begin{aligned} \sigma_e^2 &= \boldsymbol{\alpha}'\mathbf{V}_{HH}\boldsymbol{\alpha} - \boldsymbol{\beta}'\mathbf{V}_{XX}\boldsymbol{\beta} + Var(e_i) \\ &= \boldsymbol{\alpha}'\mathbf{V}_{HH}\boldsymbol{\alpha} - \boldsymbol{\alpha}'\mathbf{V}_{HX}\mathbf{V}_{XX}^{-1}\mathbf{V}_{XH}\boldsymbol{\alpha} + Var(e_i) \\ &= \sigma_a^2 - \sigma_q^2 + \sigma_e^2 \end{aligned}$$

elde edilir. Burada, $\sigma_a^2 - \sigma_q^2 = \boldsymbol{\alpha}'[\mathbf{V}_{HH} - \mathbf{V}_{HX}\mathbf{V}_{XX}^{-1}\mathbf{V}_{XH}]\boldsymbol{\alpha}$ markerler tarafından açıklanamayan genetik varyansı göstermektedir. $\frac{\sigma_a^2 - \sigma_q^2}{\sigma_a^2}$ kayıp kalıtım derecesi olup:

$$\frac{h^2 - h_M^2}{h^2} = \frac{\sigma_a^2 - \sigma_q^2}{\sigma_a^2} = \frac{\boldsymbol{\alpha}'[\mathbf{V}_{HH} - \mathbf{V}_{HX}\mathbf{V}_{XX}^{-1}\mathbf{V}_{XH}]\boldsymbol{\alpha}}{\boldsymbol{\alpha}'\mathbf{V}_{HH}\boldsymbol{\alpha}} \quad (2.9)$$

şeklinde gösterilir (de los Campos vd 2015). Kullanılan panelin marker yoğunluğu ne kadar yüksek olursa, QTL'lar ile yüksek ya da kusursuz LD içerisinde markerlerin panelde bulunma ihtimali ve hatta QTL'lerin kendilerinin de panelde bulunma ihtimali artacaktır. Bu durumda $\sigma_a^2 = \sigma_q^2$ olacaktır.



Şekil 2.3. Kayıp kalıtım derecesi

h_M^2 : genomik kalıtım derecesi; (—): simüle edilen kalıtım derecesi, h^2 .

Şekil 2.3 marker yoğunluğundaki artışın genomik kalıtım derecesi, h_M^2 , üzerindeki etkisini göstermektedir. Veri seti 10,000 bireye ait ve 1 Morgan uzunluğunda bir genomda eşit aralıklarla yerleşmiş 5,000 marker ve 50 QTL genotipini içermektedir. Bu hipotetik durumda marker sayısı 100 olduğunda marker yoğunluğu 1/1cM ve marker sayısı maksimum, yani 5,000 olduğunda marker yoğunluğu 50/1cM şeklindedir. Şekilden de açıkça görüleceği gibi marker sayısı arttıkça h_M^2 özelliğin kalıtım derecesine (h^2) yaklaşmaktadır.

2.1.4. Regresyon analizi

2.1.4.1. Basit doğrusal regresyon

Mevcut popülasyondaki bireylere ilişkin tek bir lokustaki genotiplere sahip olduğumuzu varsayalım. Bu durumda tek değişkenli basit regresyon modelini kullanmak uygun olacaktır:

$$y_i = \mu + x_i\beta + e_i \quad (2.10)$$

Burada, y_i i 'inci bireye ait gözlem değeri, μ genel ortalama, x_{ij} i 'inci bireyin ilgili lokustaki allellerden herhangi birinin sayısı (0, 1 veya 2), β bu lokusa ilişkin allel değişim etkisi ve e_i hata terimidir. Hata terimlerinin ilişkisiz ve normal dağılıma sahip olduğu varsayımı altında $var(e) = \mathbf{I}\sigma_e^2$ olmaktadır. Söz konusu modeli daha genel olarak aşağıdaki şekilde yazmak mümkündür:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{e} \quad (2.11)$$

Burada $\mathbf{X} = [\mathbf{1} \ \mathbf{x}]$, sabit etkiler μ ve β ile gözlem değerlerini ilişkilendiren desen matrisidir. En küçük kareler (EKK) yöntemi hata kareleri toplamını, $\mathbf{e}'\mathbf{e} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$, minimize eden $\boldsymbol{\beta}$ değerlerini bulmayı amaçlamaktadır. Parametre vektörü $\boldsymbol{\beta}$ 'nin EKK tahmincisi ve bu tahmincinin varyansı,

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

$$var(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\sigma_e^2$$

şeklinde (Garrick vd 2014). Genetik değer tahmini amacıyla, tek bir lokustaki genotipler değil tüm genomu kapsayan çok sayıda marker için elde edilen genotipler kullanılmaktadır.

2.1.4.2. Çoklu doğrusal regresyon

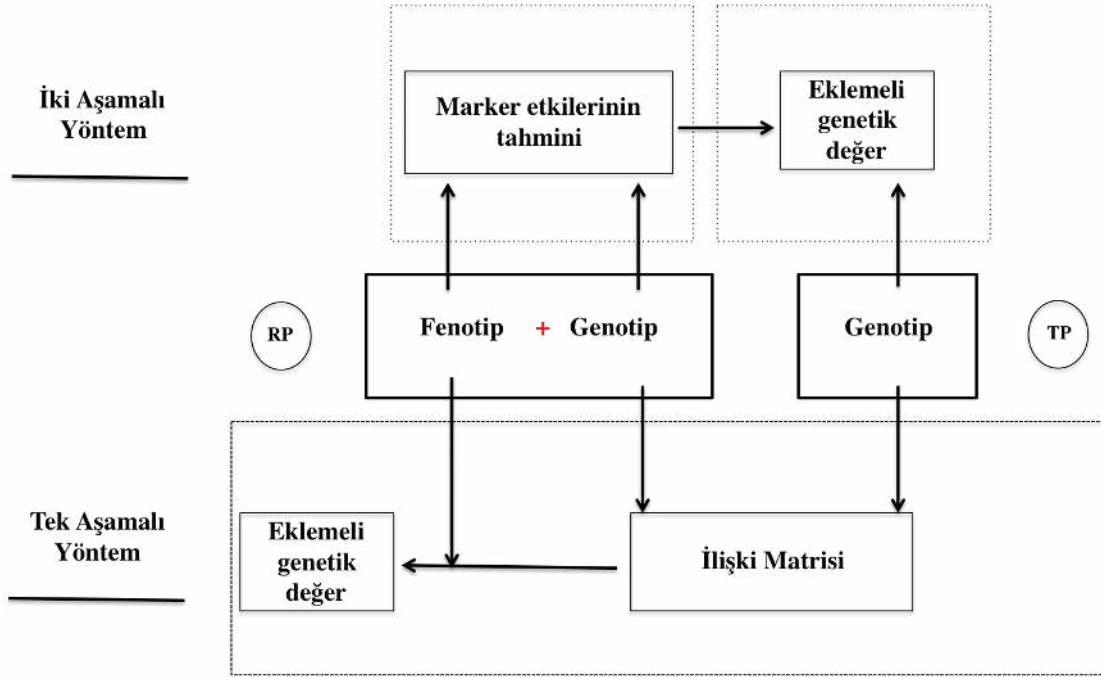
Marker etkilerinin tahmininde kullanılan model p marker sayısı olmak üzere,

$$y_i = \mu + \sum_{j=1}^p x_{ij}\beta_j + e_i \quad (2.12)$$

şeklinde. Bu modelde, x_{ij} i 'inci bireyin j 'inci marker lokusundaki genotipi, β_j j 'inci markere ait allel değişim etkisi ve e_i hata terimi olup, hata terimlerinin ilişkisiz ve normal dağılıma sahip olduğu varsayımı altında $var(\mathbf{e}) = \mathbf{I}\sigma_e^2$ şeklindedir. Model parametrelerinin EKK tahmini basit regresyon modelindeki gibidir. Ancak, genetik değer tahmin uygulamalarında 10,000'lerce hatta 100,000'lerce marker etkisi sınırlı sayıda ($<10,000$) bireye ait gözlem değerleri kullanılarak tahmin edilmektedir, $p \gg n$. Bu durumda $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ eşitliğinin birden fazla çözümü vardır (Garrick vd 2014). Bu sorunu gidermek amacı ile parametre tahmininde Bayesçi regresyon yaklaşımının kullanımı tercih edilmektedir.

2.1.5. Genomik değerlendirme

Genomik değerlendirme amacıyla kullanılan yöntemleri istatistiksel modele bağlı olarak dolaylı ya da doğrudan yöntemler olarak iki sınıfta ele almak mümkündür (Şekil 2.4) (Zhang vd 2011). Dolaylı yöntemlerde ilk olarak bir RP yardımıyla marker etkileri tahmin edilmekte ve yalnızca genotipleri bilinen bireyler için bu marker etkileri yardımı ile bireylerin eklemeli genetik değerleri elde edilmektedir. Doğrudan yöntemlerde ise bireylerin eklemeli genetik değerleri karışık model eşitlikleri yardımı ile aynı geleneksel BLUP yönteminde olduğu gibi tek adımda elde edilmektedir. Söz konusu karışık model eşitlikleri oluşturulurken RP'na ait fenotipik veriler yanında, RP ve TP'na ait genotipik bilgiler kullanılarak oluşturulan ilişki matrisinden yararlanılmaktadır (Zhang vd 2011; Meuwissen vd 2013).



Şekil 2.4. Genomik değerlendirmede kullanılan yaklaşımlar

RP: referans popülasyon; TP: test popülasyonu

2.1.5.1. Tek aşamalı yöntemler

Eklemeli genetik değer tahmininde uzun yıllardır BLUP (buradan itibaren pedigrîye dayalı BLUP-PBLUP olarak adlandırılacaktır) yöntemi kullanılmaktadır (Henderson 1984). PBLUP yöntemi, akrabalık matrisi ($\mathbf{A}$) yardımı ile eklemeli genetik değer tahmini yapılacak bireylerin tüm akrabalarına ait bilgilerden yararlanmaktadır. Söz konusu matris bireyler arasındaki eklemeli genetik ilişkileri yansıtmaktadır ve bu matrisin elemanları bireyler arasında aynı atadan gelen genlerin beklenen oranını ifade etmektedir. Ancak, söz konusu bu oran Mendel yasası gereği gerçek orandan farklılık gösterebilir (Zhang vd 2011). Tüm genomu kapsayan marker panelleri sayesinde bireyler arasındaki genetik ilişkilerin markerlerden elde edilen ilişki matrisi ($\mathbf{G}$) ile tahmin edilebilmesi mümkündür. Tek aşamalı tahmin yöntemleri geleneksel BLUP yönteminde olduğu gibi aşağıdaki karışık model eşitliğinin çözülmesine dayalıdır:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e} \quad (2.13)$$

Burada, $\mathbf{u}$, bireylere ait şansa bağlı etkiler vektörüdür ve varyansı $\mathbf{G}\sigma_u^2$ şeklindedir. Söz konusu modele ilişkin karışık model eşitlikleri,

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{G}^{-1}\sigma_u^2 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix} \quad (2.14)$$

şeklinde verilmektedir. Bu yöntem van Raden (2008) tarafından tanıtılmış olup

GBLUP (genomik BLUP) olarak adlandırılmaktadır. GBLUP yönteminin pek çok alanda popüler olmasının 3 temel nedeni vardır: (i) Genetik değer tahminleri uzun yıllardır PBLUP yöntemi ile yapılmaktadır ve GBLUP yöntemi ile genetik değer tahmini için yalnızca $\mathbf{A}$ matrisinin $\mathbf{G}$ matrisi ile değiştirilmesi gerekmektedir. Bu durum, araştırmacıların mevcut programlarda bir değişiklik yapmadan GBLUP yöntemini kolayca kullanabilmesini sağlamaktadır. (ii) Tahmin hatalarının varyansı PBLUP yönteminde olduğu gibi karışık model eşitliklerinden hesaplanabilir. (iii) Çiftlik hayvanlarında, insanlarda ya da bitkilerde üzerinde durulan özellikler çok sayıda gen tarafından belirlendiğinden, PBLUP yöntemi ile benzer varsayımlara sahip GBLUP yöntemi oldukça isabetli tahminler sağlamaktadır.

Misztal vd (2010), PBLUP ve GBLUP yöntemlerini bir araya getiren ve SS-GBLUP (single step GBLUP) olarak adlandırılan tahmin yöntemini tanıtmışlardır. Bu yöntem, karışık model eşitliğindeki $\mathbf{A}$ matrisinin,

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{pmatrix} \quad (2.15)$$

şeklinde verilen $\mathbf{H}$ matrisi ile değiştirilmesi yoluyla uygulanmaktadır (Aguilar vd 2010; Christensen ve Lund 2010). $\mathbf{A}_{22}^{-1}$ genotip bilgisi mevcut olan bireylere ait akrabalık matrisinin tersidir. SS-GBLUP yönteminde kullanılan karışık model eşitlikleri,

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{H}^{-1}a \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix} \quad (2.16)$$

şeklindedir.

Fernando vd (2014), bireylere ait fenotip, genotip ve akrabalık bilgilerini kullanarak genetik değer tahmininde bulunan ve SSBR (single step Bayesian regression) olarak adlandırdıkları yöntemi tanıtmışlardır. SS-BLUP yönteminin aksine, söz konusu yöntem marker etkilerinin normal dağılımdan gelme varsayımına bağımlı değildir ve aynı zamanda da büyük matrislerin tersinin alınmasını gerektirmemektedir. Söz konusu yöntem şu an için geliştiricileri tarafından yoğun bir şekilde kullanılmaktadır.

2.1.5.2. İki aşamalı yöntemler

Dolaylı yöntemlerde eklemeli genetik değer tahmini iki aşamalı bir şekilde gerçekleştirilmektedir. İlk aşamada marker etkileri genetik ve fenotipik verileri mevcut olan bir RP yardımıyla elde edilmekte ve ardından da tahmin edilen marker etkileri yardımı ile yalnızca genetik verileri mevcut olan TP bireylerine ait eklemeli genetik değerler tahmin edilmektedir. Her ne kadar yaygın kullanımı olmasa da, marker etkilerinin tahmininde EKK yönteminin de kullanılması mümkündür. Bu

amaçla, sayıları on binleri bulan SNP'lerin bir alt seti seçilerek çoklu doğrusal regresyon modeli uydurulabilir. Etkisi tahmin edilecek markerlerin seçimi amacıyla adimsal regresyon prosedürleri ya da temel bileşenler analizi gibi veri indirgeme yöntemlerinden yararlanılabilir (Meuwissen vd 2001).

Marker etkilerini sabit varsayan EKK yönteminin aksine, söz konusu etkiler modele şansa bağlı etki olarak da dahil edilebilir. Bu yaklaşım ilk kez Meuwissen vd (2001) tarafından yayınlanan ve genomik değerlendirme çağının başlangıcı olarak kabul edilen çalışma ile sunulmuştur. Söz konusu çalışmada, genomik değerlendirme amacıyla kullanılabilecek ve marker etkilerine dair varsayımları bakımından farklılık gösteren 3 farklı yöntem tanıtılmıştır:

$$\begin{aligned} \text{SNP-BLUP: } \beta_j &\sim N(0, \sigma_\beta^2) \\ \text{BayesA: } \beta_j &\sim N(0, \sigma_j^2) & \sigma_j^2 &\sim \chi^{-2}(v_\beta, S_\beta) \\ \text{BayesB: } \beta_j &\sim N(0, \sigma_j^2) & \sigma_j^2 &\sim \begin{cases} 0 & \pi \\ \chi^{-2}(v_\beta, S_\beta) & 1 - \pi \end{cases} \end{aligned}$$

2.1.5.2.1 Bayesci regresyon (BR) Aşağıda verilen doğrusal modeli dikkate alalım:

$$y_i = \mu + \sum_{j=1}^k x_{ij}\beta_j + e_i \quad (2.17)$$

Burada, y_i fenotipik değer, μ sabit terim, x_{ij} i 'inci bireyin j 'inci lokustaki genotipi, β_j j marker lokusunun etkisi ve e_i hata terimidir. Modelin olabilirlik fonksiyonu,

$$\begin{aligned} p(\mathbf{y} \mid \mu, \boldsymbol{\beta}, \sigma_\beta^2, \sigma_e^2) &\propto (\sigma_e^2)^{-\frac{n}{2}} \\ &\times \exp \left\{ -\frac{1}{2\sigma_e^2} \left(\mathbf{y} - \mathbf{1}\mu - \sum \mathbf{x}_j\beta_j \right)' \left(\mathbf{y} - \mathbf{1}\mu - \sum \mathbf{x}_j\beta_j \right) \right\} \end{aligned}$$

şeklindedir. Model parametreleri için önsel dağılımları

$$p(\mu) \propto \text{sabit}$$

$$p(\beta_j \mid \sigma_\beta^2) \propto (\sigma_\beta^2)^{-\frac{1}{2}} \exp \left\{ -\frac{\beta_j^2}{2\sigma_\beta^2} \right\}$$

$$p(\sigma_\beta^2 \mid v_\beta, S_\beta^2) \propto (\sigma_\beta^2)^{-\frac{1}{2}(v_\beta+2)} \exp \left\{ -\frac{v_\beta S_\beta^2}{2\sigma_\beta^2} \right\}$$

$$p(\sigma_e^2 | v_e, S_e^2) \propto (\sigma_e^2)^{-\frac{1}{2}(v_e+2)} \exp \left\{ -\frac{v_e S_e^2}{2\sigma_e^2} \right\}$$

şeklinde tanımlayalım. Bu durumda müşterek sonsal dağılımı,

$$\begin{aligned} p(\mu, \boldsymbol{\beta}, \sigma_\beta^2, \sigma_e^2 | \mathbf{y}) &\propto p(\mathbf{y} | \mu, \boldsymbol{\beta}, \sigma_\beta^2, \sigma_e^2) p(\mu) p(\boldsymbol{\beta} | \sigma_\beta^2) p(\sigma_\beta^2 | v_\beta, S_\beta^2) p(\sigma_e^2 | v_e, S_e^2) \\ &\propto (\sigma_e^2)^{-\frac{n}{2}} \\ &\times \exp \left\{ -\frac{1}{2\sigma_e^2} \left(\mathbf{y} - \mathbf{1}\mu - \sum \mathbf{x}_j \beta_j \right)' \left(\mathbf{y} - \mathbf{1}\mu - \sum \mathbf{x}_j \beta_j \right) \right\} \\ &\times \prod_{j=1}^k (\sigma_\beta^2)^{-\frac{1}{2}} \exp \left\{ -\frac{\beta_j^2}{2\sigma_\beta^2} \right\} \\ &\times (\sigma_\beta^2)^{-\frac{1}{2}(v_\beta+2)} \exp \left\{ -\frac{v_\beta S_\beta^2}{2\sigma_\beta^2} \right\} \\ &\times (\sigma_e^2)^{-\frac{1}{2}(v_e+2)} \exp \left\{ -\frac{v_e S_e^2}{2\sigma_e^2} \right\} \end{aligned}$$

şeklinde elde ederiz. Buradan marker etkileri, β_j , için koşullu sonsal dağılım $\mathbf{w}_j = \mathbf{y} - \mathbf{1}\mu - \sum_{l \neq j} \mathbf{x}_l \beta_l$ olmak üzere,

$$p(\beta_j | \mu, \sigma_\beta^2, \sigma_e^2, \mathbf{y}) \sim N \left[(\mathbf{x}'_j \mathbf{x}_j + \frac{\sigma_e^2}{\sigma_\beta^2})^{-1} \mathbf{x}'_j \mathbf{w}_j, (\mathbf{x}'_j \mathbf{x}_j + \frac{\sigma_e^2}{\sigma_\beta^2})^{-1} \sigma_e^2 \right]$$

şeklinde elde edilir.

Meuwissen vd (2001), çalışmalarında $\boldsymbol{\beta}$ için ortalaması “0” ve varyans-kovaryans matrisi $\mathbf{I}\sigma_\beta^2$ olan bir normal dağılım varsaymışlardır. Model 2.17 için $\sigma_\beta^2/\sigma_e^2$ oranı bilindiğinde, $\boldsymbol{\beta}$ ’nin en iyi doğrusal yansız tahmin edicisi (best linear unbiased predictor-BLUP), Henderson (1984)’un karışık model eşitlikleri kullanılarak elde edilebilir. Bununla birlikte, yukarıda verilen model için $\boldsymbol{\beta} | \sigma_\beta^2 \sim N(\mathbf{0}, \mathbf{I}\sigma_\beta^2)$ varsayımı altında, $\boldsymbol{\beta}$ ’nin posterior dağılımına ait ortalamanın BLUP çözümünün aynısı olduğu ve BR ve Meuwissen vd (2001) tarafından kullanılan SNP-BLUP yönteminin aynı sonuçları verdiği görülecektir.

GBLUP (genomic best linear unbiased prediction) yöntemi temelinde SNP-BLUP yöntemine eşdeğerdir ve SNP-BLUP yöntemi genellikle GBLUP olarak da adlandırılmaktadır. Bu yöntemde SNP-BLUP’tan farklı olarak marker etkileri yerine doğrudan eklemeli genetik değerler, $\mathbf{u} = \mathbf{X}\boldsymbol{\beta}$, tahmin edilmektedir. $\mathbf{P}$ allel frekanslarını içeren vektör, p_j j ’inci markerin allel frekansı ve $\mathbf{X}_c = \mathbf{X} - \mathbf{P}$ olmak üzere, genomik ilişki matrisi, $\mathbf{G}$,

$$\mathbf{G} = \frac{\mathbf{X}_c \mathbf{X}_c'}{2 \sum p_j (1 - p_j)} \quad (2.18)$$

şeklinde elde edilir (van Raden 2008). GBLUP yöntemi, Henderson (1984)'un karışık model eşitliğinde pedigrî temelli akrabalık matrisi $\mathbf{A}$ yerine genomik ilişki matrisi $\mathbf{G}$ 'nin kullanılması yoluyla kolayca uygulanabilir olduğundan yaygın olarak kullanılmaktadır. GBLUP yöntemi için kullanılacak temel model,

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad (2.19)$$

şeklindedir. Burada, $\mathbf{u}$ genetik değerleri içeren $n \times 1$ boyutlu vektör ve $\mathbf{Z}$ de desen matrisidir. Genetik değerlerin $N(\mathbf{0}, \mathbf{G}\sigma_u^2)$ şeklinde bir dağılıma sahip olduğu varsayılmaktadır.

2.1.5.2.2 BayesA Meuwissen vd (2001) tarafından tanıtılan BayesA yönteminde, SNP-BLUP yönteminden farklı olarak her bir lokusun toplam fenotipik varyasyona farklı miktarlarda katkıda bulunduğu varsayılmaktadır. BayesA yönteminde marker etkilerinin bağımsız tek değişkenli t -dağılımına sahip olduğu varsayılmaktadır. Gianola vd (2009), bu şekilde atanacak bir önsel dağılımın, j 'inci lokus için ortalaması "0" varyansı σ_j^2 olan tek değişkenli normal dağılım ve varyans σ_j^2 için de parametreleri v_β ve S_β^2 olan ters ki-kare dağılım atamakla aynı olduğunu göstermiştir.

Modelin olabilirlik fonksiyonu $\boldsymbol{\xi} = [\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2]$ olmak üzere,

$$p(\mathbf{y} \mid \mu, \boldsymbol{\beta}, \boldsymbol{\xi}, \sigma_e^2) \propto (\sigma_e^2)^{-\frac{n}{2}} \times \exp \left\{ -\frac{1}{2\sigma_e^2} (\mathbf{y} - \mathbf{1}\mu - \sum_{j=1}^k \mathbf{x}_j \beta_j)' (\mathbf{y} - \mathbf{1}\mu - \sum_{j=1}^k \mathbf{x}_j \beta_j) \right\}$$

şeklindedir. Model parametreleri için önsel dağılımları

$$p(\mu) \propto \text{sabit}$$

$$p(\beta_j \mid \sigma_j^2) \propto (\sigma_j^2)^{-\frac{1}{2}} \exp \left\{ -\frac{\beta_j^2}{2\sigma_j^2} \right\}$$

$$p(\sigma_j^2 \mid v_\beta, S_\beta^2) \propto (\sigma_j^2)^{-\frac{1}{2}(v_\beta+2)} \exp \left\{ -\frac{v_\beta S_\beta^2}{2\sigma_j^2} \right\}$$

$$p(\sigma_e^2 \mid v_e, S_e^2) \propto (\sigma_e^2)^{-\frac{1}{2}(v_e+2)} \exp \left\{ -\frac{v_e S_e^2}{2\sigma_e^2} \right\}$$

şeklinde tanımlayalım. Bu durumda müşterek sonsal dağılımı,

$$\begin{aligned}
p(\mu, \beta, \xi, \sigma_e^2 \mid \mathbf{y}) &\propto p(\mathbf{y} \mid \mu, \beta, \sigma_e^2) p(\mu) p(\beta \mid \xi) p(\xi \mid v_\beta, S_\beta^2) p(\sigma_e^2 \mid v_e, S_e^2) \\
&\propto (\sigma_e^2)^{-\frac{n}{2}} \\
&\times \exp \left\{ -\frac{1}{2\sigma_e^2} (\mathbf{y} - \mathbf{1}\mu - \sum_{j=1}^k \mathbf{x}_j \beta_j)' (\mathbf{y} - \mathbf{1}\mu - \sum_{j=1}^k \mathbf{x}_j \beta_j) \right\} \\
&\times \prod_{j=1}^k (\sigma_j^2)^{-\frac{1}{2}} \exp \left\{ -\frac{\beta_j^2}{2\sigma_j^2} \right\} \\
&\times \prod_{j=1}^k (\sigma_j^2)^{-\frac{1}{2}(v_\beta+2)} \exp \left\{ -\frac{v_\beta S_\beta^2}{2\sigma_j^2} \right\} \\
&\times (\sigma_e^2)^{-\frac{1}{2}(v_e+2)} \exp \left\{ -\frac{v_e S_e^2}{2\sigma_e^2} \right\}
\end{aligned}$$

şeklinde elde ederiz. Buradan marker etkileri, β_j , için koşullu sonsal dağılım,

$$p(\beta_j \mid \mu, \xi, \sigma_e^2, \mathbf{y}) \sim N \left[(\mathbf{x}'_j \mathbf{x}_j + \frac{\sigma_e^2}{\sigma_j^2})^{-1} \mathbf{x}'_j \mathbf{w}_j, (\mathbf{x}'_j \mathbf{x}_j + \frac{\sigma_e^2}{\sigma_j^2})^{-1} \sigma_e^2 \right]$$

olarak elde edilir.

2.1.5.2.3 BayesB BayesB yöntemi bir karışım ya da değişken seçim yöntemi olarak değerlendirilmektedir. Karışım modellerinde incelenen değişken iki ya da daha fazla dağılım karışımının birinden gelmektedir. Bu bağlamda BayesB yöntemi Bayesçi karışım modellerinin özel bir durumu olup dağılımlardan biri etkisi “0” olan lokuslar ile ilişkili iken, diğer dağılım BayesA yönteminde olduğu gibi etkisi sıfırdan farklı olan ve kendine ait bir varyansa sahip lokuslarla ilişkilidir (Meuwissen vd 2001). Bayesçi yaklaşımda araştırmacı bir lokusun etkisinin “0” olduğu inancını yansıtan ve değeri bilindiği varsayılan bir π parametresi ile bu inancını analiz sürecine dahil eder. Bu inanç veri setinden gelen bilgi ile birleşerek bir lokusun etkisinin olmadığına dair sonsal olasılığa katkıda bulunur. Bu nedenle her bir Markov chain Monte Carlo (MCMC) iterasyonunda her bir lokusun “0” ya da sıfırdan farklı bir etkisi vardır. Bu yöntem daha önce de bahsedildiği üzere bir çeşit değişken seçim yöntemidir, çünkü belirli bir iterasyonda yalnızca seçilen markerlar modele dahil edilmektedir (Garrick vd 2014).

δ_j π olasılıkla “0” ve $1 - \pi$ olasılıkla 1 değerini alan Bernoulli değişken ve $\beta_j = \gamma_j \delta_j$ olmak üzere model (2.17) aşağıdaki gibi yazılsın:

$$y_i = \mu + \sum_{j=1}^k x_{ij} \gamma_j \delta_j + e_i \quad (2.20)$$

Ayrıca, γ_j ortalaması “0” varyansı σ_j^2 olan tek değişkenli normal dağılıma sahip

olsun ve varyans σ_j^2 için de parametreleri v_β ve S_β^2 olan ters ki-kare dağılışı atayalım. Bu durumda, modelin olabilirlik fonksiyonu $\boldsymbol{\xi} = [\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2]$ olmak üzere,

$$p(\mathbf{y} \mid \mu, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\xi}, \sigma_e^2) \propto (\sigma_e^2)^{-\frac{n}{2}} \times \exp \left\{ -\frac{1}{2\sigma_e^2} (\mathbf{y} - \mathbf{1}\mu - \sum_{j=1}^k \mathbf{x}_j \gamma_j \delta_j)' (\mathbf{y} - \mathbf{1}\mu - \sum_{j=1}^k \mathbf{x}_j \gamma_j \delta_j) \right\}$$

şeklinde olacaktır. Model parametreleri için önsel dağılımları,

$$p(\mu) \propto \text{sabit}$$

$$p(\gamma_j \mid \sigma_j^2) \propto (\sigma_j^2)^{-\frac{1}{2}} \exp \left\{ -\frac{\gamma_j^2}{2\sigma_j^2} \right\}$$

$$\delta_j \mid \pi \begin{cases} = 0 & \pi \\ = 1 & 1-\pi \end{cases}$$

$$p(\sigma_j^2 \mid v_\beta, S_\beta^2) \propto (\sigma_j^2)^{-\frac{1}{2}(v_\beta+2)} \exp \left\{ -\frac{v_\beta S_\beta^2}{2\sigma_j^2} \right\}$$

$$p(\sigma_e^2 \mid v_e, S_e^2) \propto (\sigma_e^2)^{-\frac{1}{2}(v_e+2)} \exp \left\{ -\frac{v_e S_e^2}{2\sigma_e^2} \right\}$$

şeklinde tanımlayalım. Bu durumda müşterek sonsal dağılımı,

$$\begin{aligned} p(\mu, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\xi}, \sigma_e^2 \mid \mathbf{y}) &\propto p(\mathbf{y} \mid \mu, \boldsymbol{\gamma}, \boldsymbol{\delta}, \sigma_e^2) p(\mu) p(\boldsymbol{\gamma} \mid \boldsymbol{\xi}) \\ &\quad \times p(\boldsymbol{\xi} \mid v_\beta, S_\beta^2) p(\boldsymbol{\delta}) p(\sigma_e^2 \mid v_e, S_e^2) \\ &\propto (\sigma_e^2)^{-\frac{n}{2}} \\ &\quad \times \exp \left\{ -\frac{1}{2\sigma_e^2} (\mathbf{y} - \mathbf{1}\mu - \sum_{j=1}^k \mathbf{x}_j \gamma_j \delta_j)' (\mathbf{y} - \mathbf{1}\mu - \sum_{j=1}^k \mathbf{x}_j \gamma_j \delta_j) \right\} \\ &\quad \times \prod_{j=1}^k (\sigma_j^2)^{-\frac{1}{2}} \exp \left\{ -\frac{\gamma_j^2}{2\sigma_j^2} \right\} \\ &\quad \times \prod_{j=1}^k \pi^{(1-\delta_j)} (1-\pi)^{\delta_j} \\ &\quad \times \prod_{j=1}^k (\sigma_j^2)^{-\frac{1}{2}(v_\beta+2)} \exp \left\{ -\frac{v_\beta S_\beta^2}{2\sigma_j^2} \right\} \\ &\quad \times (\sigma_e^2)^{-\frac{1}{2}(v_e+2)} \exp \left\{ -\frac{v_e S_e^2}{2\sigma_e^2} \right\} \end{aligned}$$

şeklinde elde ederiz. Buradan marker etkileri, γ_j , için koşullu sonsal dağılım, $\mathbf{w}_j = \mathbf{y} - \mathbf{1}\mu - \sum_{l \neq j} \mathbf{x}_l \gamma_l \delta_l$ olmak üzere,

$$p(\gamma_j \mid \mu, \boldsymbol{\delta}, \boldsymbol{\xi}, \sigma_e^2, \mathbf{y}) \begin{cases} \sim N \left[(\mathbf{x}'_j \mathbf{x}_j + \frac{\sigma_e^2}{\sigma_j^2})^{-1} \mathbf{x}'_j \mathbf{w}_j, (\mathbf{x}'_j \mathbf{x}_j + \frac{\sigma_e^2}{\sigma_j^2})^{-1} \sigma_e^2 \right] & \delta_j = 1 \text{ ise} \\ \sim N(0, \sigma_j^2) & \delta_j = 0 \text{ ise} \end{cases}$$

olarak elde edilir. $\pi = 0$ olduğunda marker etkilerinin BayesB tahminleri ile BayesA tahminlerinin aynı olduğu açıktır.

2.1.5.2.4 BayesC Kızılkaya vd (2010), bir başka değişken seçim yöntemi olan; ancak tüm lokuslar için sabit bir varyans varsayan BayesC yöntemini tanıtmışlardır. Bu yöntemin özel bir hali, $\pi = 0$ olduğunda ortaya çıkar ve literatürde "BayesC0" olarak adlandırılır. $\pi = 0$ olduğunda BayesC'den elde edilecek genetik değer tahminleri GBLUP tahminleri ile aynı olacaktır.

Modelin olabilirlik fonksiyonu,

$$p(\mathbf{y} \mid \mu, \boldsymbol{\gamma}, \boldsymbol{\delta}, \sigma_\beta^2, \sigma_e^2) \propto (\sigma_e^2)^{-\frac{n}{2}} \times \exp \left\{ -\frac{1}{2\sigma_e^2} (\mathbf{y} - \mathbf{1}\mu - \sum_{j=1}^k \mathbf{x}_j \gamma_j \delta_j)' (\mathbf{y} - \mathbf{1}\mu - \sum_{j=1}^k \mathbf{x}_j \gamma_j \delta_j) \right\}$$

şeklinde. Model parametreleri için önsel dağılımları,

$$p(\mu) \propto \text{sabit}$$

$$p(\gamma_j \mid \sigma_\beta^2) \propto (\sigma_\beta^2)^{-\frac{1}{2}} \exp \left\{ -\frac{\gamma_j^2}{2\sigma_\beta^2} \right\}$$

$$\delta_j \mid \pi \begin{cases} = 0 & \pi \\ = 1 & 1-\pi \end{cases}$$

$$p(\sigma_\beta^2 \mid v_\beta, S_\beta^2) \propto (\sigma_\beta^2)^{-\frac{1}{2}(v_\beta+2)} \exp \left\{ -\frac{v_\beta S_\beta^2}{2\sigma_\beta^2} \right\}$$

$$p(\sigma_e^2 \mid v_e, S_e^2) \propto (\sigma_e^2)^{-\frac{1}{2}(v_e+2)} \exp \left\{ -\frac{v_e S_e^2}{2\sigma_e^2} \right\}$$

şeklinde tanımlayalım. Bu durumda müşterek sonsal dağılımı,

$$\begin{aligned}
p(\mu, \gamma, \delta, \sigma_\beta^2, \sigma_e^2 \mid \mathbf{y}) &\propto p(\mathbf{y} \mid \mu, \gamma, \delta, \sigma_e^2) p(\mu) p(\gamma, \mid \sigma_\beta^2) \\
&\times p(\sigma_\beta^2 \mid v_\beta, S_\beta^2) p(\delta) p(\sigma_e^2 \mid v_e, S_e^2) \\
&\propto (\sigma_e^2)^{-\frac{n}{2}} \\
&\times \exp \left\{ -\frac{1}{2\sigma_e^2} (\mathbf{y} - \mathbf{1}\mu - \sum_{j=1}^k \mathbf{x}_j \gamma_j \delta_j)' (\mathbf{y} - \mathbf{1}\mu - \sum_{j=1}^k \mathbf{x}_j \gamma_j \delta_j) \right\} \\
&\times \prod_{j=1}^k (\sigma_\beta^2)^{-\frac{1}{2}} \exp \left\{ -\frac{\gamma_j^2}{2\sigma_\beta^2} \right\} \\
&\times \prod_{j=1}^k \pi^{(1-\delta_j)} (1 - \pi)^{\delta_j} \\
&\times (\sigma_\beta^2)^{-\frac{1}{2}(v_\beta+2)} \exp \left\{ -\frac{v_\beta S_\beta^2}{2\sigma_\beta^2} \right\} \\
&\times (\sigma_e^2)^{-\frac{1}{2}(v_e+2)} \exp \left\{ -\frac{v_e S_e^2}{2\sigma_e^2} \right\}
\end{aligned}$$

şeklinde elde ederiz. Buradan marker etkileri, γ_j , için koşullu sonsal dağılım, $\mathbf{w}_j = \mathbf{y} - \mathbf{1}\mu - \sum_{l \neq j} \mathbf{x}_l \gamma_l \delta_l$ olmak üzere,

$$p(\gamma_j \mid \delta, \sigma_\beta^2, \sigma_e^2, \mathbf{y}) \begin{cases} \sim N \left[(\mathbf{x}'_j \mathbf{x}_j + \frac{\sigma_e^2}{\sigma_\beta^2})^{-1} \mathbf{x}'_j \mathbf{w}_j, (\mathbf{x}'_j \mathbf{x}_j + \frac{\sigma_e^2}{\sigma_\beta^2})^{-1} \sigma_e^2 \right] & \delta_j = 1 \text{ ise} \\ \sim N(0, \sigma_\beta^2) & \delta_j = 0 \text{ ise} \end{cases}$$

$\pi = 0$ iken $\delta_j = 1$ olduğundan genetik değerlerin BayesC tahminleri ile BR, RR-BLUP ve GBLUP tahminleri aynı olmaktadır (Nejati-Javaremi 1997; Fernando 1998; Habier vd 2007; van Raden 2008; Strandén ve Garrick 2009).

2.1.6. Genomik değerlendirmede isabet

İsabet derecesi bireylerin gerçek eklemeli genetik değerleri (u) ile tahmin edilen eklemeli genetik değerleri ($\hat{u}$) arasındaki korelasyon olarak ifade edilmekte ve genellikle $r_{(u, \hat{u})}$ ile gösterilmektedir. $var(u)$ ve $var(\hat{u})$ sırası ile gerçek ve tahmin edilen eklemeli genetik değerlere ilişkin varyansları, $kov(u, \hat{u})$ ise her iki değişken arasındaki kovaryansı göstermek üzere genomik değerlendirmede isabetin bir ölçüsü olarak güvenilirlik,

$$r_{(\hat{u}, u)}^2 = \frac{kov(\hat{u}, u)^2}{var(\hat{u})var(u)} \quad (2.21)$$

kullanılabilir. Uygulamada ise eklemeli genetik değerler bilinmediğinden genomik değerlendirmede isabet, bireylerin tahmin edilen genetik değerleri, $\hat{u}$, ve fenotipik değerleri, y , arasındaki korelasyonun karesi şeklinde hesaplanmakta ve belirleme

katsayısı, R^2 , ile gösterilmektedir:

$$R^2_{(\hat{u},y)} = R^2 = \frac{kov(\hat{u},y)^2}{var(\hat{u})var(y)} \quad (2.22)$$

Genomik değerlendirmede isabet, yöntemin faydasını ortaya koymada veya genomik değerlendirme amacıyla kullanılan tahmin yöntemlerini karşılaştırmada ele alınan önemli bir faktördür. Genomik değerlendirmede isabetin veri toplama aşamasından önce tahmin edilmesi amacıyla geliştirilmiş çeşitli yakınsamalar mevcuttur.

Daetwyler vd (2008), en küçük kareler yöntemini esas alarak sürekli bir özellik için genetik değer tahmininde isabeti önceden tahmin etmede kullanılmak üzere aşağıdaki formülü önermiştir (Ek-2):

$$r^2_{(\hat{u},u)} \approx \frac{n_R h^2}{n_R h^2 + n_G} \quad (2.23)$$

Burada, n_R RP büyüklüğü, h^2 kalıtım derecesi ve n_G ise birbirinden bağımsız olduğu varsayılan lokusların sayısıdır. Bu formül markerler arasındaki LD'yi dikkate almadığından n_G arttıkça genomik değerlendirmede isabetin sıfıra yaklaşacağı açıktır. Goddard (2009), markerlerin birbirinden bağımsız olduğu varsayıldığında, n_G yerine etkin kromozom bölgelerinin sayısının, M_e , kullanılabileceğini ileri sürmüştür:

$$M_e = (2N_e Lk)/\log(N_e L) \quad (2.24)$$

Daetwyler vd (2008) tarafından önerilen yakınsama, M_e kullanılarak yeniden düzenlendiğinde,

$$r^2_{(\hat{u},u)} \approx \frac{n_R h^2}{n_R h^2 + M_e} \quad (2.25)$$

elde edilir. Diğer yandan, Goddard (2009) bu formüle ilave olarak GBLUP yönteminin isabetini tahmin etmek üzere kullanılmak amacıyla aşağıda verilen yakınsamayı önermiştir:

$$r^2_{(\hat{u},u)} \approx 1 - \frac{\lambda}{2n_R \sqrt{a}} \log \left(\frac{1 + a + 2\sqrt{a}}{1 + a - 2\sqrt{a}} \right) \quad (2.26)$$

burada, $a = 1 + 2\lambda/n_R$ ve $\lambda = [(1 - h^2)M_e]/[h^2 \log(2N_e)]$ şeklindedir. Söz konusu formüller, markerlerin uygun bir alt setini seçen ve bu alt set için parametre tahminleri gerçekleştiren BayesB, BayesC gibi yönteminin isabetini tahmin etmede ancak QTL sayısı, n_{QTL} , M_e 'ye yaklaştığı zaman uygundur. Aksi halde, örneğin

BayesB için tahmin edilecek isabet gerçekleşen isabet değerinden düşük olacaktır. Daetwyler vd (2010b), bu durumu dikkate alarak aşağıdaki eşitliği önermiştir:

$$r_{(\hat{u},u)}^2 \approx \frac{n_R h^2}{n_R h^2 + \min(n_{QTL}, M_e)} \quad (2.27)$$

$n_{QTL} < M_e$ olduğunda GBLUP yöntemi uygun bir yöntem olmayacaktır. Ancak, veri seti ve önsel bilgileri kullanarak markerlerin uygun bir alt setini seçen BayesB yöntemi tüm markerlara ait etkileri tahmin etmek yerine n_{QTL} sayıdaki marker etkilerini tahmin etme eğiliminde olacaktır. Diğer yandan, n_{QTL} yeteri kadar büyük olduğunda her iki yöntem de M_e sayıda parametre tahmin etme eğiliminde olduğundan, ilgili yakınsamadan elde edilecek tahminler ile gerçek isabet uyumlu olacaktır (Daetwyler vd 2010b).

Genomik değerlendirmede isabeti tahmin etmeye çalışan ve markerler ile QTL'ler arasında tam bir LD varsayan yaklaşımların aksine, Goddard vd (2011), markerların kusursuz olmayan LD sebebiyle tüm genetik varyasyonu kapsamayacağını hesaba katan farklı bir yakınsama önermiştir:

$$r_{(\hat{u},u)}^2 \approx q^2 \frac{\theta}{\theta + 1}. \quad (2.28)$$

Burada, q^2 daha önce de verildiği gibi genetik varyansın markerlar tarafından açıklanan oranı ve $\theta = n_R q^2 h^2 / M_e$ şeklindedir. Goddard vd (2011) tarafından önerilen yakınsama yeniden düzenlendiğinde, R^2 için aşağıdaki yakınsama elde edilir (1.1):

$$R^2 \approx h_M^2 \frac{n_R h_M^2}{n_R h_M^2 + M_e} \quad (2.29)$$

Yukarıda verilen formüller incelendiğinde, genomik değerlendirmede isabeti etkileyen faktörleri; (i) araştırmacının kontrolünde olmayan ve popülasyona özgü faktörler ve (ii) araştırmacının kontrolünde olan faktörler olarak iki ayrı grupta değerlendirmek mümkündür (Zhang vd 2011). Birinci grupta; ortalama kromozom uzunluğu (L), kromozom sayısı (k), her bir popülasyona özgü bir büyüklük olan ve popülasyonun evrilme sürecine bağlı olan efektif popülasyon büyüklüğü (N_e) ilgilenilen özelliği etkileyen QTL sayısı (n_{QTL}) ve özelliğin kalıtım derecesi (h^2) yer almaktadır. İkinci grupta ise; RP büyüklüğü ve yapısı, marker yoğunluğu ve kullanılan tahmin yöntemi yer almaktadır. Ortalama kromozom uzunluğu (L) ve etkin popülasyon büyüklüğü N_e 'in bir fonksiyonu olan M_e 'deki bir artışın genetik değer tahmininin isabetinde bir azalma yaratacağı görülmektedir (Zhang vd 2011).

2.1.7. Genomik değerlendirme çalışmaları ve isabeti etkileyen faktörler

Genomik değerlendirme, Meuwissen vd (2001) tarafından temelleri atılan bir yöntem olup, allel değişim etkisi, doğrusal regresyon, karışık etkiler modeli ve genetik ilişki konularının açık kavramlarını bir araya getiren yeni bir çalışma alanı yaratmıştır (Garrick vd 2014). Meuwissen vd (2001)'nin çalışmasını takiben, söz konusu çalışmada önerilen BLUP, BayesA ve BayesB yöntemlerine alternatifler sunan, genomik değerlendirmenin klasik yöntemlere göre üstünlüklerini ortaya koyan ya da genomik değerlendirmenin gerçek verilere uygulanması ile elde edilen sonuçları bildiren çok sayıda çalışma gerçekleştirilmiştir.

Gerçek veriler kullanılarak yapılan çalışmalar genomik değerlendirmenin klasik yöntemlere göre üstünlüğünü ortaya koymuştur. Siyah Alaca ırkı sığırlarda (Hayes vd 2009a; van Raden vd 2009), etçi sığır ırklarında (Saatchi vd 2011; Garrick 2011), etlik piliçlerde (Gonzalez-Recio vd 2008), yumurtacı tavuklarda (Wolc vd 2011a,b), koyunlarda (Daetwyler vd 2010a) ve çeşitli bitki türlerinde (Crossa vd 2010; Heslot vd 2012) genomik değerlendirmenin klasik yöntemlere göre üstün olduğunu bildiren çalışmalar mevcuttur.

Gerek simülasyonlar gerekse de gerçek veriler kullanılarak yürütülen çalışmalarda genomik değerlendirmede sağlanacak isabetin çok sayıda faktörden etkilendiği bildirilmiştir. Bunlar; marker etkilerinin tahmininde kullanılan yöntem (Habier vd 2007; Luan vd 2009), özelliğin kalıtım derecesi (h^2) ve genetik yapısı (Luan vd 2009; Daetwyler vd 2010b), referans popülasyonun büyüklüğü (n_R) ve yapısı (Goddard 2009; Daetwyler vd 2010b; Habier vd 2010b; Pszczola vd 2011; Makowsky vd 2011; Clark vd 2012) olarak verilebilir. Kullanılacak tahmin yöntemi ve referans popülasyonun yapısı ve büyüklüğü, üzerinde çalışılan özelliğin kalıtım derecesi ve genetik yapısı dikkate alınarak belirlenebilir.

Meuwissen (2009), RP büyüklüğü arttıkça GBLUP ve BayesB arasında eklemeli genetik değer tahmininde isabet bakımından bulunan farklılığın azaldığını bildirmiştir. Söz konusu çalışmada ayrıca, marker yoğunluğu arttıkça BayesB yöntemi ile GBLUP yönteminden elde edilen tahminlerin isabetleri arasındaki farkın BayesB lehine açıldığını bildirilmiştir. Meuwissen ve Goddard (2010) QTL'leri de kapsayan sekans verilerini kullanarak gerçekleştirdikleri çalışmalarında BayesB yönteminin GBLUP yöntemine göre daha isabetli tahminler verdiğini bildirmişlerdir. Daetwyler vd (2010b), simülasyonlar kullanarak yaptıkları çalışmada GBLUP yönteminin isabetinin QTL sayısından etkilenmediğini, buna karşın BayesC'nin tahmin performansının QTL sayısı ile doğrudan ilişkili olduğunu ve QTL sayısı arttıkça BayesC'nin isabetinin de azaldığını ortaya koymuşlardır. Clark vd (2011), allel frekansları bakımından farklılık gösteren farklı senaryolar altında BayesB'nin isabetinin GBLUP yönteminden daha yüksek olduğunu ortaya koymuşlardır. Söz konusu çalışmada, özellikle büyük etkili QTL'ler mevcut olduğunda BayesB'nin GBLUP'a göre üstün olduğu, ancak diğer durumlarda da BayesB'nin en az GBLUP kadar isabetli tahminler verdiğini bildirilmiştir. Diğer yandan, Ober vd (2011)

BayesB yönteminden elde edilen tahmin isabetinin GBLUP yönteminden elde edilen tahmin isabetine göre daha yüksek olmadığını bildirmişlerdir.

Simülasyon çalışmaları az sayıda QTL tarafından determine edilen özellikler için değişken seçimine dayalı Bayesçi yöntemlerin (BayesB, BayesC vb.) tüm SNP etkilerini tahmin eden diğer yöntemlere göre daha başarılı olduğunu ortaya koymuştur (Meuwissen vd 2001; Daetwyler vd 2010b; Zeng vd 2012). Ancak, simülasyon çalışmalarının aksine gerçek verilerin analiz edildiği çalışmalarda bu durum genelde doğrulanamamış ve farklı yöntemlerin isabet başarıları arasındaki farklılıklar kimi zaman oldukça düşük seviyelerde kalmış ya da yöntemlerin başarı sıralaması değişkenlik göstermiştir. Resende vd (2012), büyük etkiye sahip az sayıda gen tarafından etkilendiği bilinen bir özellik dışında üzerinde çalışılan tüm özellikler için tahmin yöntemleri arasında küçük farklılıklar tespit etmişlerdir. Bermingham vd (2015), insanlara ait verileri kullandıkları çalışmalarında çok sayıda gen tarafından determine edildiği bilinen özellikler için GBLUP ve BayesC yöntemi arasında genetik değer tahmin isabeti bakımından bir farklılık olmadığını, buna karşın belirli sayıda gen tarafından determine edilen bir özellik için BayesC yönteminin GBLUP'tan daha isabetli tahminler verdiğini bildirmişlerdir.

Calus ve Veerkamp (2007), kalıtım derecesi h^2 , yüksek (0.50) bir özellik için kalıtım derecesi düşük (0.10) olan özelliğe göre daha isabetli eklemeli genetik değer tahminleri elde ettikleri bir simülasyon çalışması gerçekleştirmişlerdir. Saatchi vd (2010), simülasyonlar kullanarak $h^2 = 0.10$ ve 0.50 olan iki özellikten kalıtım derecesi yüksek (0.50) olan özellik için elde edilen eklemeli genetik değer tahminlerinin, kalıtım derecesi düşük (0.10) olan özelliğe göre daha isabetli olduğunu bildirmişlerdir. Brito vd (2011), simülasyonla üretilmiş bir veri setinde $h^2 = 0.10, 0.25$ ve 0.40 olan üç farklı özelliği incelemiş ve özelliğin kalıtım derecesi arttıkça eklemeli genetik değer tahminlerinde isabetin arttığını bildirmiştir. Kapell (2012), farelerde $0.10-0.74$ arasında kalıtım derecesine sahip özellikleri incelediği çalışmasında kalıtım derecesi arttıkça eklemeli genetik değer tahmininde isabetin de arttığını bildirmiştir. Azizian vd (2016), simülasyonlardan yararlandıkları çalışmalarında artan kalıtım dereceleri için r^2 değerlerinin de arttığını bildirmişlerdir.

Marker yoğunluğunun genetik değer tahmininde isabet üzerindeki etkilerini inceleyen çok sayıda çalışma yapılmıştır. Yüksek yoğunluklu marker panelleri kullanılması ile genetik değer tahmininde isabetin artacağı düşünülmeye karşın, bazı çalışmalardan elde edilen sonuçlar bunu doğrular nitelikte olmamıştır. Vazquez vd (2010), Siyah Alaca ırkı sığırlarda gerçekleştirdikleri ve çok sayıda özelliği inceledikleri çalışmalarında 10,000'den fazla SNP kullanılmasının tahmin isabetinde çok küçük bir artışa neden olduğunu bildirmişlerdir. Ober vd (2011), meyve sineklerine ait verileri kullandıkları çalışmalarında 150,000 veya 2.5 milyon SNP kullanılarak elde edilen tahminlerin isabetleri bakımından farklı olmadığını bildirmiştir. Benzer şekilde Su vd (2012), bazı süt sığırı ırklarında 54K ve 777K marker panellerini kullanarak elde edilen tahminlerin isabet bakımından çok küçük bir farklılığa yol açtığını bildirmişlerdir. Wientjes vd (2015), sığırlarda

606,384 SNP ile 60,000 SNP kullanılarak elde edilen tahminler arasında isabet bakımından bir farklılık bulunmadığını bildirmişlerdir. Diğer yandan, Makowsky vd (2011), insanlardan elde edilen verileri kullanarak marker sayısı 100,000'lerin üzerine çıktığında dahi isabetin artmakta olduğunu ortaya koymuştur. Druet vd (2014), nadir görülen mutasyonlar söz konusu olduğunda sekans verilerinin kullanılmasının isabet derecelerini arttırdığını bildirmiştir. Simülasyon çalışmaları yüksek yoğunluklu marker panellerinin eklemeli genetik değer tahmininde isabet üzerindeki etkisinin ilgili özellik az sayıda gen tarafından determine edildiğinde yüksek, çok sayıda gen tarafından determine edildiğinde ise düşük olduğunu ortaya koymuştur (Meuwissen ve Goddard 2010).

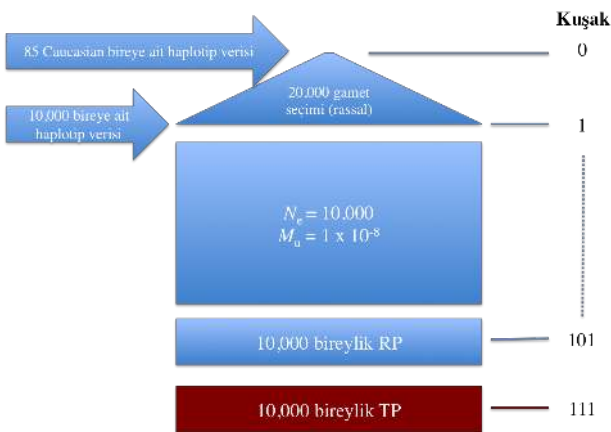
Genomik değerlendirmede karşılaşılan sorunlardan biri sayıları 10,000'leri bulan SNP etkilerini tahmin etmek için yeterli büyüklükte bir referans popülasyon bulunmasındaki zorluktur. Tahmin isabetinin RP büyüklüğü ile arttığını ortaya koyan çok sayıda çalışma bulunmaktadır (Hayes vd 2009b; Verbyla vd 2009; Lorenzana ve Bernardo 2009; van Raden ve Sullivan 2010; Bastiaansen vd 2010). Referans popülasyon büyüklüğündeki artışın genomik değerlendirmede isabet üzerine etkisi iki yolla gerçekleşmektedir: (1) marker etkilerinin tahmininde isabetin artması ve (2) TP bireylerinin RP bireyleri ile olan ilişkilerinin artması (de los Campos vd 2013a). Literatürde, yüksek isabetli genomik değerlendirme için en uygun referans popülasyonun yapısına ve TP ve RP bireyleri arasındaki ilişkinin tahminlerin isabeti üzerindeki etkisine dair çok sayıda çalışma bulunmaktadır (Habier vd 2007; Legarra vd 2008; Habier vd 2010b; Pszczola vd 2011; Makowsky vd 2011; Clark vd 2011). Meuwissen (2009), referans popülasyon bireyleri ile aralarında akrabalık bulunmayan bireylerin genetik değer tahmini için referans popülasyonda yakın akrabası bulunan bireyler için gerekenden daha yoğun marker panelleri ve daha büyük referans popülasyonları kullanılması gerektiğini bildirmiştir. Habier vd (2010b), referans ve test popülasyonu bireyleri arasındaki akrabalığın genetik değer tahmininde isabet üzerine etkisini incelemiş ve TP ve RP bireyleri arasındaki akrabalık azaldıkça tahminlerin isabetinin de düştüğünü bildirmişlerdir. Pszczola vd (2011), genomik değerlendirmede isabeti arttırmak için RP içerisindeki akrabalığın düşük tutulması gerektiğini, buna karşın TP bireyleri ile RP bireyleri arasındaki akrabalığın ise maksimize edilmesi gerektiğini bildirmişlerdir. Perez-Cabal vd (2012), TP bireylerinin yakın akrabalarının RP içerisinde olmasının genomik değerlendirmede isabeti arttırdığını ve özellikle düşük kalıtım derecesine sahip özellikler ile çalışıldığında bu durumun daha da önemli hale geldiğini bildirmişlerdir. Wu vd (2015), RP ve TP arasındaki ilişkinin yüksek olmasının r^2 'yi arttırdığını, buna karşın RP içerisindeki ilişkilerin yüksek olmasının r^2 'yi azalttığını bildirmiştir. Grinberg vd (2016), çok yıllık çimler üzerinde yapmış oldukları çalışmada, optimum referans popülasyonun yapısına dair net bir sonuca ulaşamamışlardır.

3. MATERİYAL VE METOD

3.1. Veri Setleri ve Simülasyonlar

Çalışmada simülasyon yolu ile üretilmiş veri setleri kullanılmıştır. Bu amaçla, 1,000 Genom Projesi (THE 1000 GENOMES PROJECT CONSORTIUM 2012) kapsamında elde edilen ve çeşitli ırklardan toplam 1,029 bireye ait haplotip bilgilerini barındıran veri seti içerisinde Caucasian ırkına (Beyaz ırk) mensup 85 bireye ait veriler kullanılmıştır (<ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20110521/>). Genomik tahmin R^2 değerinin asimptotik davranışının bireylere ait tüm genoma dağılan SNP verileri kullanılarak belirlenmesi daha önce de belirtildiği gibi yarım milyondan fazla birey gerektirecektir. Bu nedenle, çalışmanın mevcut bilgisayar olanakları ile yürütülebilmesi için yalnızca ilk 5 kromozomdan 0.1M uzunluğunda bir kromozom parçası seçilmiş ve kromozom üzerindeki pozisyonu verilen SNP'ler dikkate alınmıştır.

Söz konusu bireylere ait maternal ve paternal haplotip verilerinden 10,000 bireylik kurucu bir popülasyon oluşturabilmek amacıyla, tamamen rassal olarak eşleştirilen bireylerden, krossing-over ve mutasyon da dikkate alınarak, gametler oluşturulmuştur. $N_e = 10,000$ ve mutasyon oranı $M_u = 1 \times 10^{-8}$ olmak üzere, ebeveynler 100 kuşak boyunca rassal olarak eşleştirilmiştir (Şekil 3.1). Referans ve test popülasyonları arasındaki akrabalığı minimuma indirmek ve TP bireyleri için yakın akrabaların RP'nda bulunmamasını garanti etmek amacıyla referans ve test popülasyonlarının birbirlerinden 10 kuşak uzakta olmalarına karar verilmiş ve 101. kuşak RP ve 111. kuşak TP olarak kullanılmak üzere simülasyonlara 11 kuşak daha devam edilmiştir. Simülasyonlar XSim yazılımı kullanılarak gerçekleştirilmiştir (Cheng vd 2015).



Şekil 3.1. Simülasyon şeması

N_e : etkin popülasyon büyüklüğü; M_u : mutasyon oranı; RP ve TP sırası ile referans ve test popülasyonları

Çizelge 3.1. Her bir senaryodaki MAF dağılımları

	MAF	Tip	<3%	3%-5%	5%-10%	10%-15%	>15%
S1	Yüksek	Marker	0.061(256)*	0.046(192)	0.117(492)	0.114(477)	0.663(2,783)
	Düşük	QTL	0.314(22)	0.243(17)	0.229(16)	0.200(14)	0.014(1)
S2	Yüksek	Marker	0.065(273)	0.049(206)	0.119(500)	0.115(483)	0.652(2,738)
	Rassal	QTL	0.029(2)	0.057(4)	0.100(7)	0.100(7)	0.714(50)
S3	Rassal	Marker	0.082(346)	0.050(211)	0.096(401)	0.104(436)	0.668(2,806)
	Rassal	QTL	0.100(7)	0.029(2)	0.114(8)	0.114(8)	0.643(45)

*Verilen sınıftaki SNP'lerin payı ve (sayısı); MAF: minör allel frekansı

RP ve TP kullanılarak hesaplanan minör allel frekansları (MAF) 0.005'ten düşük olan SNP'ler veri setinden çıkartılmıştır. Nihai veri seti toplamda 36,242 SNP içermektedir. Çalışmada, marker ve QTL'lerin allel frekansları bakımından 3 farklı senaryo üzerinde durulmuştur:

Senaryo 1 (S1) - markerler yüksek, QTL'ler düşük allel frekansları bakımından seçilmiştir

Senaryo 2 (S2) - markerler yüksek allel frekansı, QTL'ler ise rassal olarak seçilmiştir

Senaryo 3 (S3) - marker ve QTL'ler rassal olarak seçilmiştir

Veri setinde yer alan 36,242 SNP arasından yukarıda bahsedilen senaryolara uygun olacak şekilde $n_{marker} = 4,200$ marker ve $n_{QTL} = 70$ QTL seçilmiştir. Bu sayede marker yoğunluğunun (84/cM) ve n_{marker}/n_{QTL} oranının (60/1) de los Campos vd (2013b) yapmış oldukları çalışmada kullanılan değerlerle (300,000 marker ve 5,000 QTL) benzerlik teşkil etmesi amaçlanmıştır. Çalışmada ele alınan senaryolarda, S1-S3, yer alan SNP'lerin minör allel frekansı dağılımları Çizelge 3.1'de verilmiştir.

3.2. Referans ve Test Popülasyonları

Bu çalışmada, daha önce bahsedildiği üzere referans ve test popülasyonları aralarında 10 kuşak fark olacak şekilde sırası ile 101 ve 111. kuşaktaki bireyler arasından seçilmiştir. Referans popülasyonunda yer alacak birey sayısı, n_R , 75-150-500-1,000-2,000-4,000 ve 8,000 olacak şekilde 101. kuşaktaki 10,000 birey arasından tamamen rassal olarak seçilmiştir. Benzer şekilde, test popülasyonunda yer alacak 2,000 birey 111. kuşaktaki 10,000 birey arasından tamamen rassal olarak seçilmiştir.

3.3. Fenotipik Değerlerin Simülasyonu

de los Campos vd (2013b)'nin çalışmasına benzer şekilde yüksek kalıtım derecesine sahip bir özellik, $h^2 = 0.8$, için fenotipler simülasyon yolu ile üretilmiştir.

QTL etkileri, α_j 'ler $\alpha_j \sim N(0, 1)$ olan normal dağılıştan üretilmiştir. Kalıtım derecesinin tekerrürler arasında değişmemesini temin etmek amacıyla QTL etkileri her bir tekerrürde ölçeklendirilmiştir. QTL etkileri ve QTL genotipleri kullanılarak i bireyine ait genetik değer,

$$u_i = \sum_{j=1}^{n_{QTL}} \alpha_j \times Q_{ij},$$

şeklinde elde edilmektedir. Burada, n_{QTL} QTL sayısı, α_j j 'inci QTL'in eklemeli etkisi ve Q_{ij} de i 'inci bireyin j 'inci QTL'deki 0,1 veya 2 olarak kodlanan genotipidir. Her bir senaryonun, S1-S3, tüm tekerrürlerinde aynı SNP'ler marker veya QTL olarak kullanılmış olup QTL etkileri ise tekerrürler arasında farklılık göstermektedir. Bireylerin fenotipik değerlerini, y_i , oluşturabilmek için her bir bireyin genetik değerine, u_i , standart normal dağılıştan üretilen bir e_i değeri ilave edilmiştir.

3.4. Marker Etkilerinin Tahmini

Veri setine uydurulan istatistiksel model,

$$y_i = \mu + \sum_{j=1}^p x_{ij}\beta_j + e_i,$$

şeklinde. Burada, y_i RP'deki i bireyinin fenotipik değeri, μ genel ortalama, p lokus sayısı, x_{ij} i 'inci bireyin j 'inci lokustaki genotipi, β_j j 'inci markerin etkisi ve e_{ij} şansa bağlı çevre etkisidir ($e_i \sim N(0, \sigma_e^2)$). Bireylere ait eklemeli genetik değerlerin tahmini için ilk olarak marker etkileri BayesB ve BayesC yöntemleri kullanılarak tahmin edilmiştir. Eklemeli genetik değerlerin BayesB ve BayesC tahminleri elde edilirken $\pi = 0.98$ olarak alınmış, GBLUP tahminleri elde edilirken ise BayesC yönteminden yararlanılmış ve $\pi = 0$ olarak alınmıştır (BayesC0). Zincir uzunluğu 11,000 olarak belirlenmiş ve bunun ilk 1,000 örneklik kısmı burn-in olarak atılmıştır. Analizler GenSel yazılımı (Fernando ve Garrick 2013; Garrick ve Fernando 2013) kullanılarak gerçekleştirilmiştir.

3.5. Eklemeli Genetik Değerlerin Tahmini ve İsabet

Marker etkileri RP'daki bireylere ait genotip ve fenotipik bigiler kullanılarak tahmin edildikten sonra i 'inci bireyin eklemeli genetik değeri, u ,

$$\hat{u}_i = \sum_{j=1}^p x_{ij}\hat{\beta}_j$$

şeklinde elde edilir. Burada $\hat{\beta}_j$ j 'inci lokusun tahmin edilen etkisi ve x_{ij} de i 'inci bireyin j 'inci lokustaki genotipidir. Genomik değerlendirmede isabetin ölçüsü olarak

fenotipik değerler ve tahmin edilen eklemeli genetik değerler arasındaki korelasyon katsayısının karesi, R^2 , kullanılmıştır.

$\mathbf{g}_{M,i}$ ve $\mathbf{g}_{Q,i}$ sırası ile test popülasyonundaki bir bireyin referans popülasyonundaki bireylerle marker ve QTL düzeylerindeki ilişkisini gösteren vektörler ve ϵ_i artıkları içeren vektör olmak üzere,

$$\mathbf{g}_{M,i} = b_i \times \mathbf{g}_{Q,i} + \epsilon_i \quad (3.1)$$

modeli kullanılarak test popülasyonundaki her bir birey için b_i değerleri tahmin edilmiştir. Gerek marker gerekse QTL düzeylerindeki ilişkilerin hesaplanmasında, gerçekleşen ilişki matrisi, $\mathbf{G}$, kullanılmıştır:

$$\mathbf{G} = \frac{\mathbf{Z}\mathbf{Z}'}{p}$$

Burada $\mathbf{Z}$ referans ve test popülasyonlarındaki bireyler için standardize edilmiş genotipleri (marker veya QTL) içeren matris ve p lokus sayısıdır (marker veya QTL). $\mathbf{x}_j$ tüm bireylerin j 'inci lokustaki genotiplerini içeren vektör ve q_j de j 'inci lokus için allel frekansı olmak üzere standardizasyon işlemi aşağıdaki şekilde yapılmıştır:

$$\mathbf{z}_j = \frac{\mathbf{x}_j - 2q_j}{\sqrt{2q_j(1 - q_j)}}.$$

Her senaryo, S1-S3, 10 tekerrürlü olarak ele alınmış ve sonuçlar tüm tekerrürlerin ortalaması alınarak sunulmuştur.

3.6. Tahmin İsabedi ile Sabit Etkiler Modelinde Tahmin Edilebilirlik Arasındaki Bağlantı

Aşağıdaki gibi verilen ve β 'nin sabit varsayıldığı doğrusal modeli ele alalım:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{e}$$

Böyle bir modelde, $\mathbf{k}'\beta$ olarak verilen bir doğrusal fonksiyonun yalnızca $\mathbf{k}'\hat{\beta}$ tahmincisinin beklenen değerinin $\mathbf{k}'\beta$ olması durumunda tahmin edilebilir olduğu varsayılmaktadır. $\hat{\beta}$ en küçük kareler tahmincisi olduğunda $\mathbf{k}'\beta$ yalnızca $\mathbf{k}'$ olarak verilen satır vektörünün $\mathbf{X}$ matrisinin satır uzayında olması durumunda tahmin edilebilir (Scheffe 1989).

RR-BLUP yönteminde β 'nin ortalamasının sıfır olduğu varsayılır ve $\mathbf{x}'_T\beta$ 'nin BLUP değeri

$$E(\mathbf{x}'_T \tilde{\boldsymbol{\beta}}) = E(\mathbf{x}'_T \boldsymbol{\beta}) = 0$$

olması dolayısıyla yansızdır. Burada $\tilde{\boldsymbol{\beta}}$ $\boldsymbol{\beta}$ 'nin BLUP değeridir. Tahmin isabeti ve tahmin edilebilirlik arasındaki bağlantıyı kurabilmek için $\mathbf{X}_R$ 'nin RP bireylerinin genotiplerini içeren matris ve $\mathbf{x}'_T$ 'nin de TP'ndaki bir bireyin genotiplerini içeren bir vektör olduğunu varsayalım. Ayrıca, $\Re(\mathbf{X}_R)$, $\mathbf{X}_R$ matrisinin satır uzayını gösterebilir. Bu durumda herhangi bir $\mathbf{x}_T$ vektörünü $\mathbf{x}_{T_1} = \mathbf{Q}_{X'_R} \mathbf{x}_T \in \Re(\mathbf{X}_R)$ ve $\mathbf{x}_{T_2} = (\mathbf{I} - \mathbf{Q}_{X'_R}) \mathbf{x}_T \perp \Re(\mathbf{X}_R)$ olacak şekilde aşağıdaki gibi yazabiliriz:

$$\mathbf{x}_T = \mathbf{x}_{T_1} + \mathbf{x}_{T_2}. \quad (3.2)$$

Burada, $\mathbf{Q}_{X'_R} = \mathbf{X}'_R (\mathbf{X}_R \mathbf{X}'_R)^{-1} \mathbf{X}_R$ şeklindedir (Scheffe 1989). Eşitlik (3.2)'den, test popülasyonundaki bir bireyin eklemeli genetik değerini, $\mathbf{x}'_T \tilde{\boldsymbol{\beta}}$,

$$\mathbf{x}'_T \tilde{\boldsymbol{\beta}} = \mathbf{x}'_{T_1} \tilde{\boldsymbol{\beta}} + \mathbf{x}'_{T_2} \tilde{\boldsymbol{\beta}} \quad (3.3)$$

şeklinde yazabiliriz. Ek-3'te $\tilde{\boldsymbol{\beta}} \in \Re(\mathbf{X}_R)$ olduğu ve $\mathbf{x}'_T \boldsymbol{\beta}$ 'nin BLUP değerinin

$$\mathbf{x}'_T \tilde{\boldsymbol{\beta}} = \mathbf{x}'_{T_1} \tilde{\boldsymbol{\beta}} \quad (3.4)$$

olduğu gösterilmiştir, görüleceği gibi veri seti hiçbir şekilde $\mathbf{x}'_{T_2} \tilde{\boldsymbol{\beta}}$ 'nin tahminine katkıda bulunmaz ve prior dağılımı atanan sıfır tahminini alır. $u_i = \mathbf{x}'_T \boldsymbol{\beta}$ ve $\hat{u}_i = \mathbf{x}'_T \tilde{\boldsymbol{\beta}}$ arasındaki korelasyonu Eşitlik (3.4)'ten faydalanarak yazacak olursak, BLUP varsayımları altında (Henderson 1984),

$$Cor(u_i, \hat{u}_i) = \sqrt{\frac{Var(\hat{u}_i)}{Var(u_i)}} \quad (3.5)$$

ve eşitlik (3.4)'ten eşitlik (3.5)'in payı $Var(\mathbf{x}'_{T_1} \tilde{\boldsymbol{\beta}})$ olacaktır. Buradan, Eşitlik (3.5) aşağıdaki gibi yazılabilir:

$$Cor(u_i, \hat{u}_i) = \sqrt{\frac{Var(\mathbf{x}'_{T_1} \tilde{\boldsymbol{\beta}})}{Var(\mathbf{x}'_T \boldsymbol{\beta})}}.$$

Açıkça görüldüğü gibi $Var(\mathbf{x}'_{T_2} \tilde{\boldsymbol{\beta}}) = 0$ olup $Cor(u_i, \hat{u}_i)$ tahminine herhangi bir katkıda bulunmaz. Buradan hareketle, bir TP bireyinin genotip vektörü RP bireylerinin genotip matrisine ortogonal olduğunda ($\mathbf{x}_T = \mathbf{x}_{T_2}$) söz konusu bireyin RP ile genomik seviyede ilişkisiz olduğu söylenebilir. Böyle bir birey için $Cor(u_i, \hat{u}_i) = 0$ olacaktır. Diğer yandan, genotip vektörü RP bireylerinin genotiplerini içeren matrisin satır uzayında olan bir TP bireyinin ($\mathbf{x}_T = \mathbf{x}_{T_1}$) ise, RP bireyleri ile kusursuz bir ilişki içerisinde olduğu söylenebilir. Böyle bir

birey için $Var(\mathbf{x}'_{T_1}\tilde{\beta})$ $Var(\mathbf{x}'_T\beta)$ 'ye yaklařırken $Cor(u_i, \hat{u}_i)$ de 1'e yaklařacaktır. Genelleřtirmek olursak, $Cor(u_i, \hat{u}_i)$ 'nin maksimum deęeri $\sqrt{\frac{Var(\mathbf{x}'_{T_1}\beta)}{Var(\mathbf{x}'_T\beta)}}$ 'dir ve bu durumda $Cor^2(u_i, \hat{u}_i)$ 'in, yani r^2 'nin, maksimum deęeri (upper bound-UP)

$$UP_i = \frac{\mathbf{x}'_{T_1}\mathbf{x}_{T_1}}{\mathbf{x}'_T\mathbf{x}_T} \quad (3.6)$$

olacaktır (Ek-4). $UP_i = 0$ olduęunda r^2 deęeri RP'un byklęnden baęımsız olarak «0» olacaktır. Bununla birlikte, $UP_i = 1$ olduęunda RP'daki birey sayısı arttıka r^2 de 1'e yaklařacaktır. Dięer yandan, $UP_i < a$, olduęunda r^2 referans poplasyonun byklęnden baęımsız olarak a 'dan daha kkk olacaktır. Buradan hareketle, i 'inci birey iin R^2 'nin st limiti h^2UP_i olacaktır.

Teorik olarak ortaya konan bu st limiti test etmek iin 36,242 SNP arasından 5,000 tanesi rassal olarak seilmiř ve tamamı QTL olarak deęerlendirilerek $h^2 = 0.8$ ve 0.999 olmak zere iki farklı kantitatif zellik simle edilmiřtir. Referans poplasyonu oluřturmak zere 101. kuřaktaki 10,000 birey arasından 500, 1,000, 2,000 ve 5,000 birey rassal olarak seilirken, test poplasyonunu oluřturmak iin 111. kuřaktaki 10,000 birey arasından 2,000 tanesi rassal olarak seilmiřtir. Genetik deęerlerin tahmininde GBLUP yntemi (BayesC, $\pi = 0$) kullanılmıřtır. Analizin dięer adımları Blm 3.4 ve 3.5'te verildięi gibidir. $h^2 = 0.999$ olan bir kantitatif zellik simle edilmesinin nedeni marker etkilerinin tahminindeki hatayı minimumda tutarak tahmin isabetinin neredeyse btnyle TP bireylerinin tahmin edilebilirlikleri tarafından belirlenmesini saęlamaktır. Test poplasyonu bireylerinin RP ile olan iliřkileri her bir TP bireyinin RP bireyleri ile QTL dzeyindeki maksimum iliřki katsayısı kullanılarak, $max(\mathbf{g}_{T_i})$, incelenmiřtir. Bireyler soz konusu iliřki katsayısı bakımından iki gruba, dřk iliřki (DI) ve yksek iliřki (YI), ayrılmıřtır. Sz konusu gruplandırma yapılırken $max(\mathbf{g}_{T_i}) < 0.15$ olan bireyler DI grubuna ve $max(\mathbf{g}_{T_i}) \geq 0.25$ olan bireyler ise YI grubuna atanmıřtır.

4. BULGULAR

4.1. Panelde Yalnızca Markerlerin Bulunduğu Durum

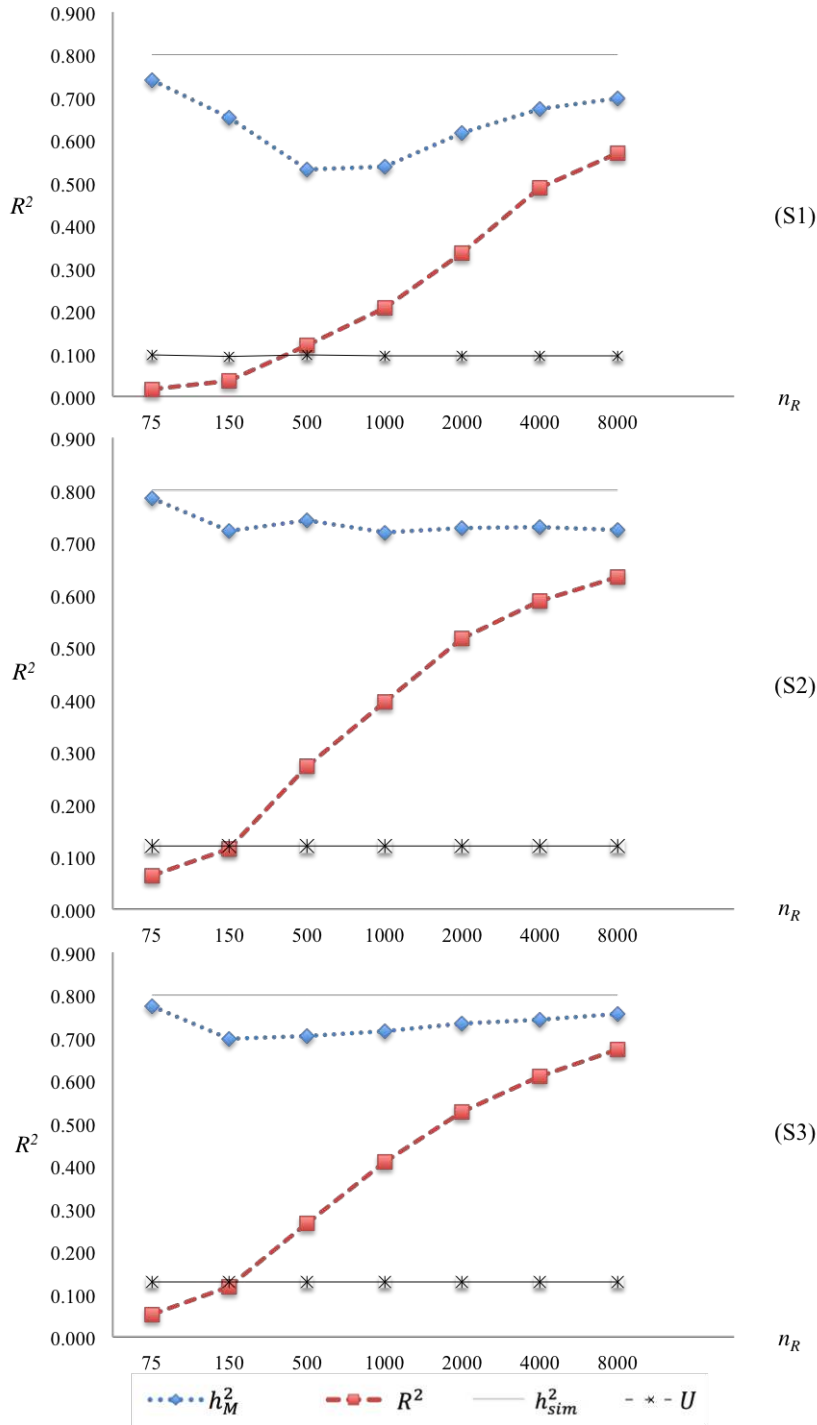
4.1.1. GBLUP yöntemine ilişkin tahminler

GBLUP yöntemi ile yalnızca marker genotipleri kullanılarak yapılan tahminlerin isabeti Şekil 4.1'de grafiksel olarak özetlenmiştir. Açıkça görüleceği gibi çalışmada ele alınan tüm senaryolar için, S1-S3, n_R 'daki bir artış tahminlerin isabet derecelerinde bir artışa neden olmaktadır. Isabet derecelerindeki bu artış, markerlerin yüksek, buna karşın QTL'lerin düşük allel frekansları için secildiği ve bu yönüyle gerçek verilerin kullanıldığı çalışmalara daha çok benzeyen Senaryo 1 için de geçerlidir.

Senaryo 1'de $n_R = 75-150-500-1,000-2,000-4,000$ veya 8,000 için elde edilen R^2 değerleri sırası ile 0.017, 0.035, 0.119, 0.207, 0.336, 0.489 ve 0.569'dur (Çizelge 4.1). Markerlerin yüksek allel frekansı için, buna karşılık QTL'lerin şansa bağlı olarak seçildiği Senaryo 2'de ise 75-8,000 aralığında değişen RP büyüklükleri için elde edilen R^2 değerleri sırası ile 0.064, 0.115, 0.272, 0.395, 0.516, 0.588 ve 0.635'tir (Çizelge 4.2). Marker ve QTL'lerin şansa bağlı olarak seçildiği Senaryo 3'te 75-8,000 aralığında değişen RP büyüklükleri için elde edilen R^2 değerleri sırası ile 0.053, 0.117, 0.266, 0.409, 0.528, 0.609 ve 0.673 olmuştur (Çizelge 4.3).

Kalıtım derecesi tahminleri, h_M^2 , Senaryo 1'de artan RP büyüklükleri için değişken bir seyir izlerken, Senaryo 2 ve 3'te daha stabil bir seyir izlemektedir (Şekil 4.1). Senaryo 1'de $n_R = 75-150-500-1,000-2,000-4,000$ veya 8,000 için elde edilen h_M^2 değerleri sırası ile 0.741, 0.654, 0.531, 0.539, 0.617, 0.673 ve 0.698'dir (Çizelge 4.1). Markerlerin yüksek allel frekansı için, buna karşılık QTL'lerin şansa bağlı olarak seçildiği Senaryo 2'de ise 75-8,000 aralığında değişen RP büyüklükleri için elde edilen h_M^2 değerleri sırası ile 0.784, 0.722, 0.743, 0.719, 0.728, 0.730 ve 0.724'dir (Çizelge 4.2). Marker ve QTL'lerin şansa bağlı olarak seçildiği Senaryo 3'te 75-8,000 aralığında değişen RP büyüklükleri için elde edilen h_M^2 değerleri sırası ile 0.775, 0.698, 0.705, 0.716, 0.735, 0.743 ve 0.757 olmuştur (Çizelge 4.3).

Tüm senaryolarda RP'daki birey sayısı arttıkça GBLUP yönteminden elde edilen tahminlerin isabeti tahmin edilen kalıtım derecesine, h_M^2 , yaklaşma eğilimindedir. Şekil 4.1'de test popülasyonundaki her bir birey için tahmin edilen regresyon katsayılarının, b_i , ortalama değeri kullanılarak hesaplanan ve de los Campos vd (2013b) tarafından önerilen üst limit, $[1 - (1 - b)^2]h^2$, gösterilmiştir. Söz konusu regresyon katsayılarının değişen referans popülasyonu büyüklüklerindeki ortalama değerleri yaklaşık aynı olup, yalnızca 2. ondalık basamaktan sonra değiştiğinden (sonuçlar verilmemiştir), $[1 - (1 - b)^2]h^2$ olarak hesaplanan üst limit de benzer şekilde değişen RP büyüklükleri için stabil bir görüntü ortaya koymuştur.



Şekil 4.1. Panelde yalnızca markerler bulunduğunda GBLUP yönteminin isabeti n_R : referans popülasyondaki birey sayısı; h_M^2 : genomik kalıtım derecesi; R^2 : y ve $\hat{u}$ arasındaki korelasyonun karesi; h_{sim}^2 : simüle edilen özelliğin kalıtım derecesi; U (de los Campos vd 2013b) tarafından önerilen üst sınır, $[1 - (1 - b)^2]h^2$, S1: markerler ve QTL sırası ile yüksek ve düşük MAF bakımından seçilmiştir; S2: markerler yüksek MAF bakımından seçilirken, QTL tamamen rassal olarak seçilmiştir; S3: markerler ve QTL tamamen rassal olarak seçilmiştir.

Çizelge 4.1. Senaryo 1'e ait sonuçlar

	n_R	n_T	R^2	h_M^2	R_{M+QTL}^2	h_{M+QTL}^2
GBLUP	75	2,000	0.017(0.004)	0.741(0.021)	0.023(0.005)	0.686(0.027)
BayesB	75	2,000	0.032(0.007)	0.731(0.013)	0.069(0.016)	0.770(0.015)
BayesC	75	2,000	0.027(0.005)	0.708(0.027)	0.052(0.012)	0.722(0.030)
GBLUP	150	2,000	0.035(0.004)	0.654(0.030)	0.046(0.005)	0.661(0.029)
BayesB	150	2,000	0.082(0.015)	0.664(0.021)	0.174(0.022)	0.730(0.019)
BayesC	150	2,000	0.062(0.011)	0.650(0.032)	0.130(0.016)	0.721(0.027)
GBLUP	500	2,000	0.119(0.011)	0.531(0.037)	0.159(0.011)	0.618(0.035)
BayesB	500	2,000	0.371(0.021)	0.605(0.018)	0.660(0.009)	0.788(0.007)
BayesC	500	2,000	0.382(0.026)	0.635(0.019)	0.683(0.008)	0.801(0.007)
GBLUP	1,000	2,000	0.207(0.010)	0.539(0.021)	0.283(0.009)	0.640(0.017)
BayesB	1,000	2,000	0.486(0.018)	0.604(0.017)	0.749(0.004)	0.794(0.003)
BayesC	1,000	2,000	0.492(0.019)	0.610(0.015)	0.755(0.003)	0.795(0.003)
GBLUP	2,000	2,000	0.336(0.012)	0.617(0.010)	0.465(0.008)	0.734(0.008)
BayesB	2,000	2,000	0.571(0.011)	0.628(0.008)	0.784(0.002)	0.801(0.002)
BayesC	2,000	2,000	0.573(0.011)	0.629(0.007)	0.786(0.002)	0.801(0.002)
GBLUP	4,000	2,000	0.489(0.008)	0.673(0.006)	0.619(0.002)	0.773(0.002)
BayesB	4,000	2,000	0.602(0.009)	0.652(0.007)	0.789(0.003)	0.802(0.001)
BayesC	4,000	2,000	0.599(0.009)	0.649(0.007)	0.789(0.003)	0.801(0.001)
GBLUP	8,000	2,000	0.569(0.008)	0.698(0.005)	0.702(0.002)	0.792(0.001)
BayesB	8,000	2,000	0.624(0.008)	0.659(0.006)	0.794(0.002)	0.801(0.001)
BayesC	8,000	2,000	0.622(0.008)	0.656(0.006)	0.794(0.002)	0.801(0.001)

n_R : referans popülasyondaki birey sayısı; n_T : test popülasyonundaki birey sayısı; R^2 ve h_M^2 panelde yalnızca markerler bulunduğu durumda tahmin isabeti ve genomik kalıtım dereceleri; R_{M+QTL}^2 ve h_{M+QTL}^2 panelde markerler ve QTL'lerin bulunduğu durumda tahmin isabeti ve genomik kalıtım dereceleri

Çizelge 4.2. Senaryo 2'ye ait sonuçlar

	n_R	n_T	R^2	h_M^2	R_{M+QTL}^2	h_{M+QTL}^2
GBLUP	75	2,000	0.064(0.008)	0.784(0.011)	0.087(0.009)	0.757(0.013)
BayesB	75	2,000	0.051(0.005)	0.763(0.014)	0.092(0.014)	0.790(0.013)
BayesC	75	2,000	0.065(0.008)	0.754(0.015)	0.106(0.015)	0.775(0.014)
GBLUP	150	2,000	0.115(0.012)	0.722(0.028)	0.155(0.014)	0.750(0.023)
BayesB	150	2,000	0.207(0.020)	0.708(0.019)	0.298(0.021)	0.775(0.018)
BayesC	150	2,000	0.209(0.020)	0.729(0.027)	0.296(0.023)	0.793(0.022)
GBLUP	500	2,000	0.272(0.011)	0.743(0.013)	0.353(0.010)	0.819(0.009)
BayesB	500	2,000	0.473(0.014)	0.681(0.005)	0.682(0.009)	0.805(0.007)
BayesC	500	2,000	0.473(0.014)	0.692(0.005)	0.683(0.009)	0.808(0.007)
GBLUP	1,000	2,000	0.395(0.007)	0.719(0.012)	0.494(0.006)	0.803(0.008)
BayesB	1,000	2,000	0.569(0.009)	0.673(0.010)	0.754(0.003)	0.804(0.003)
BayesC	1,000	2,000	0.571(0.009)	0.675(0.010)	0.753(0.003)	0.803(0.003)
GBLUP	2,000	2,000	0.516(0.013)	0.728(0.008)	0.616(0.008)	0.802(0.004)
BayesB	2,000	2,000	0.629(0.011)	0.691(0.008)	0.780(0.003)	0.804(0.002)
BayesC	2,000	2,000	0.629(0.011)	0.691(0.008)	0.780(0.003)	0.803(0.002)
GBLUP	4,000	2,000	0.588(0.007)	0.730(0.005)	0.686(0.003)	0.800(0.002)
BayesB	4,000	2,000	0.655(0.009)	0.697(0.006)	0.789(0.002)	0.802(0.001)
BayesC	4,000	2,000	0.654(0.008)	0.696(0.006)	0.789(0.002)	0.802(0.001)
GBLUP	8,000	2,000	0.635(0.016)	0.724(0.011)	0.739(0.003)	0.800(0.001)
BayesB	8,000	2,000	0.665(0.018)	0.691(0.013)	0.796(0.002)	0.800(0.001)
BayesC	8,000	2,000	0.665(0.018)	0.690(0.013)	0.796(0.002)	0.800(0.001)

n_R : referans popülasyondaki birey sayısı; n_T : test popülasyonundaki birey sayısı; R^2 ve h_M^2 panelde yalnızca markerler bulunduğu durumda tahmin isabeti ve genomik kalıtım dereceleri; R_{M+QTL}^2 ve h_{M+QTL}^2 panelde markerler ve QTL'lerin bulunduğu durumda tahmin isabeti ve genomik kalıtım dereceleri

Çizelge 4.3. Senaryo 3'e ait sonuçlar

	n_R	n_T	R^2	h_M^2	R_{M+QTL}^2	h_{M+QTL}^2
GBLUP	75	2,000	0.053(0.008)	0.775(0.009)	0.072(0.010)	0.748(0.009)
BayesB	75	2,000	0.073(0.016)	0.768(0.009)	0.124(0.021)	0.789(0.008)
BayesC	75	2,000	0.077(0.015)	0.753(0.010)	0.123(0.021)	0.774(0.009)
GBLUP	150	2,000	0.117(0.012)	0.698(0.041)	0.157(0.013)	0.724(0.037)
BayesB	150	2,000	0.187(0.027)	0.724(0.028)	0.304(0.024)	0.799(0.018)
BayesC	150	2,000	0.190(0.026)	0.731(0.044)	0.307(0.024)	0.814(0.023)
GBLUP	500	2,000	0.266(0.011)	0.705(0.014)	0.345(0.011)	0.790(0.009)
BayesB	500	2,000	0.501(0.017)	0.672(0.010)	0.691(0.006)	0.796(0.006)
BayesC	500	2,000	0.499(0.016)	0.682(0.010)	0.691(0.007)	0.800(0.006)
GBLUP	1,000	2,000	0.409(0.009)	0.716(0.017)	0.505(0.008)	0.801(0.011)
BayesB	1,000	2,000	0.588(0.008)	0.679(0.013)	0.760(0.004)	0.803(0.007)
BayesC	1,000	2,000	0.587(0.008)	0.682(0.013)	0.761(0.004)	0.803(0.007)
GBLUP	2,000	2,000	0.528(0.011)	0.735(0.009)	0.624(0.006)	0.803(0.004)
BayesB	2,000	2,000	0.647(0.013)	0.699(0.010)	0.784(0.002)	0.803(0.002)
BayesC	2,000	2,000	0.647(0.013)	0.698(0.010)	0.784(0.002)	0.802(0.002)
GBLUP	4,000	2,000	0.609(0.007)	0.743(0.005)	0.693(0.004)	0.800(0.001)
BayesB	4,000	2,000	0.678(0.007)	0.709(0.006)	0.792(0.002)	0.801(0.001)
BayesC	4,000	2,000	0.678(0.007)	0.708(0.005)	0.792(0.002)	0.801(0.001)
GBLUP	8,000	2,000	0.673(0.005)	0.757(0.003)	0.740(0.003)	0.801(0.001)
BayesB	8,000	2,000	0.706(0.006)	0.729(0.004)	0.795(0.002)	0.801(0.001)
BayesC	8,000	2,000	0.705(0.005)	0.728(0.004)	0.795(0.002)	0.801(0.001)

n_R : referans popülasyondaki birey sayısı; n_T : test popülasyonundaki birey sayısı; R^2 ve h_M^2 panelde yalnızca markerler bulunduğu durumda tahmin isabeti ve genomik kalıtım dereceleri; R_{M+QTL}^2 ve h_{M+QTL}^2 panelde markerler ve QTL'lerin bulunduğu durumda tahmin isabeti ve genomik kalıtım dereceleri

4.1.2. Tahmin yöntemlerinin karşılaştırılması

Şekil 4.2’de, çalışmada ele alınan GBLUP, BayesB ve BayesC yöntemleri için değişen RP büyüklüklerindeki R^2 değerleri grafiksel olarak gösterilmiştir. Senaryo 1’de GBLUP, BayesB ve BayesC yöntemlerinden elde edilen R^2 değerleri $n_R = 75$ için 0.017, 0.032 ve 0.027, $n_R = 150$ için 0.035, 0.082 ve 0.062, $n_R = 500$ için 0.119, 0.371 ve 0.382, $n_R = 1,000$ için 0.207, 0.486 ve 0.492, $n_R = 2,000$ için 0.336, 0.571 ve 0.573, $n_R = 4,000$ için 0.489, 0.602 ve 0.599, $n_R = 8,000$ için ise 0.569, 0.624 ve 0.622’dir (Çizelge 4.1).

Senaryo 2’de GBLUP, BayesB ve BayesC yöntemlerinden elde edilen R^2 değerleri $n_R = 75$ için 0.064, 0.051 ve 0.065, $n_R = 150$ için 0.115, 0.207 ve 0.209, $n_R = 500$ için 0.272, 0.473 ve 0.473, $n_R = 1,000$ için 0.395, 0.569 ve 0.571, $n_R = 2,000$ için 0.516, 0.629 ve 0.629, $n_R = 4,000$ için 0.588, 0.655 ve 0.654, $n_R = 8,000$ için ise 0.635, 0.665 ve 0.665’dir (Çizelge 4.2).

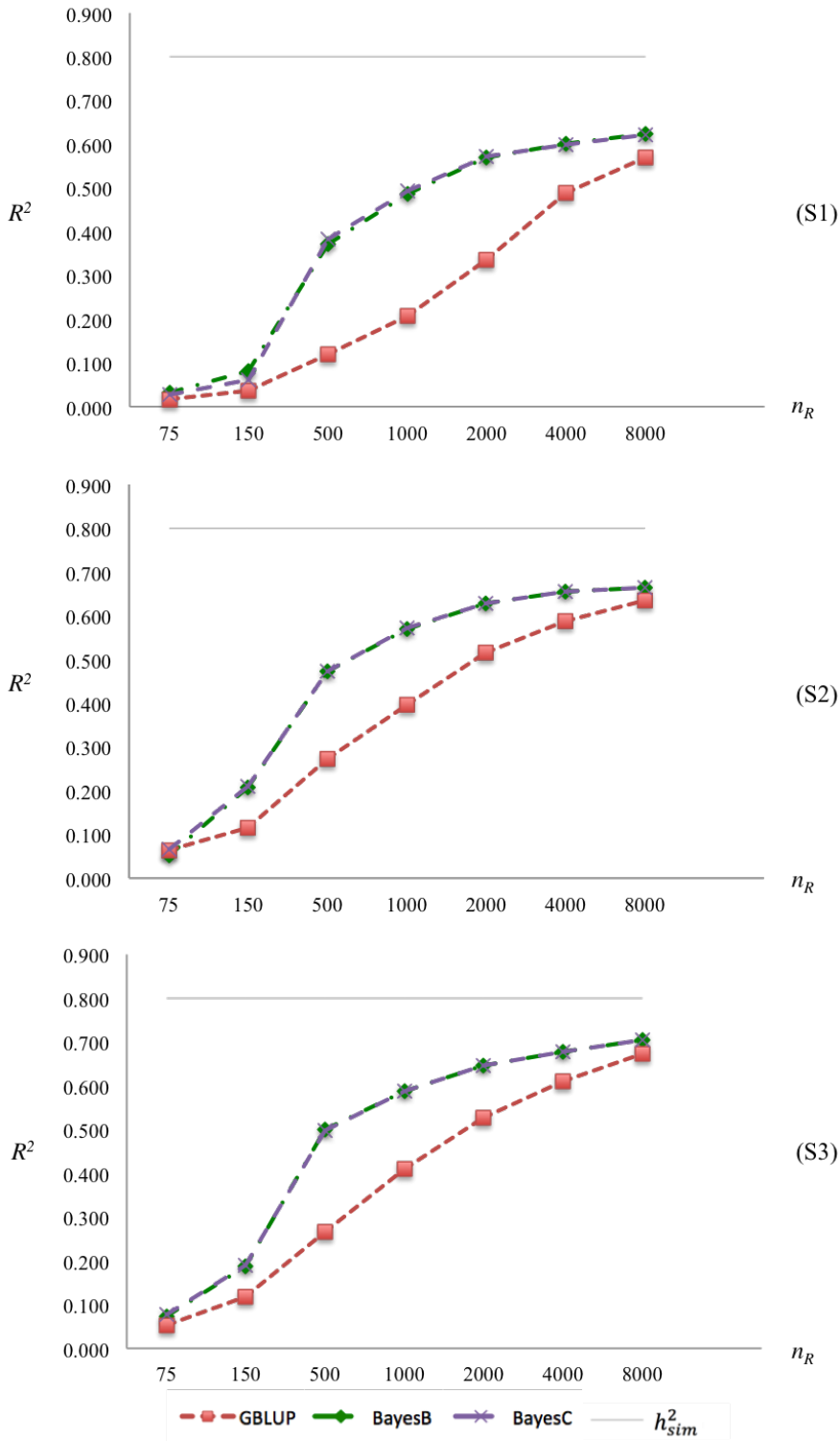
Senaryo 3’de GBLUP, BayesB ve BayesC yöntemlerinden elde edilen R^2 değerleri $n_R = 75$ için 0.053, 0.073 ve 0.077, $n_R = 150$ için 0.117, 0.187 ve 0.190, $n_R = 500$ için 0.266, 0.501 ve 0.499, $n_R = 1,000$ için 0.409, 0.588 ve 0.587, $n_R = 2,000$ için 0.528, 0.647 ve 0.647, $n_R = 4,000$ için 0.609, 0.678 ve 0.678, $n_R = 8,000$ için ise 0.673, 0.706 ve 0.705’dir (Çizelge 4.3).

4.2. Panelde Markerler ve QTL’lerin Bulunduğu Durum

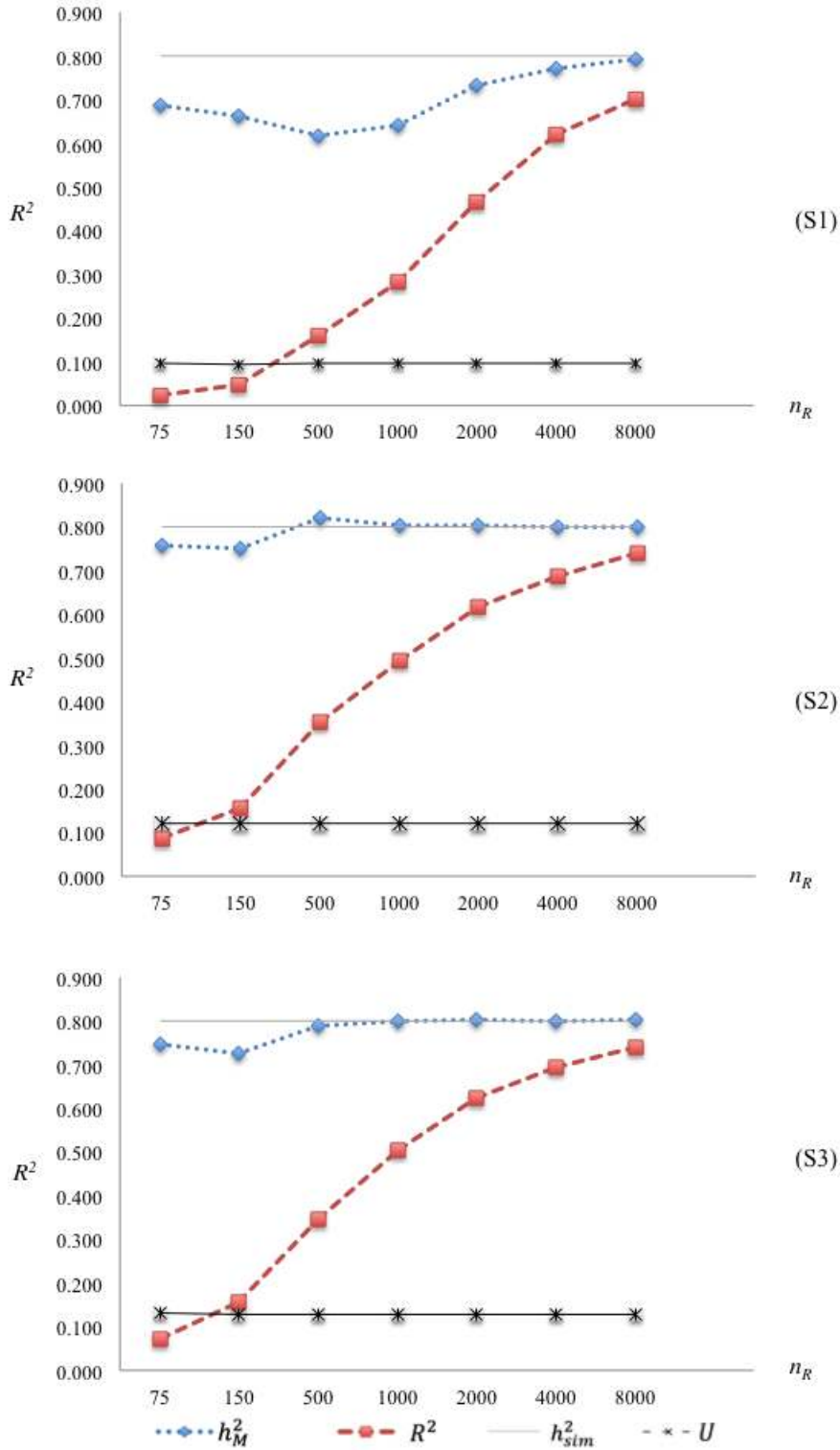
4.2.1. GBLUP yöntemine ilişkin tahminler

GBLUP yöntemi ile bireylere ait marker ve QTL genotipleri kullanılarak elde edilen tahminlerin isabeti Şekil 4.3’de grafiksel olarak özetlenmiştir. Açıkça görüleceği gibi çalışmada ele alınan tüm senaryolar için, S1-S3, n_R ’daki bir artış tahminlerin isabet derecelerinde bir artışa neden olmaktadır. İsabet derecelerindeki bu artış, markerların yüksek, buna karşın QTL’lerin düşük allel frekansları için seçildiği ve bu yönüyle gerçek bireylerden elde edilen verilerin kullanıldığı çalışmalara daha çok benzeyen Senaryo 1 için de geçerlidir. Senaryo 1’de $n_R = 75-150-500-1,000-2,000-4,000$ veya 8,000 için elde edilen R^2 değerleri sırası ile 0.023, 0.046, 0.159, 0.283, 0.465, 0.619 ve 0.702’dir (Çizelge 4.1). Markerların yüksek allel frekansı için, buna karşılık QTL’lerin şansa bağlı olarak seçildiği Senaryo 2’de ise 75-8,000 aralığında değişen RP büyüklükleri için elde edilen R^2 değerleri sırası ile 0.087, 0.155, 0.353, 0.494, 0.616, 0.686 ve 0.739’dur (Çizelge 4.2). Marker ve QTL’lerin şansa bağlı olarak seçildiği Senaryo 3’te 75-8,000 aralığında değişen RP büyüklükleri için elde edilen R^2 değerleri sırası ile 0.072, 0.157, 0.345, 0.505, 0.624, 0.693 ve 0.740 olmuştur (Çizelge 4.3).

Kalıtım derecesi tahminleri, h_M^2 , Senaryo 1’de artan RP büyüklükleri için değişken bir seyir izlerken, Senaryo 2 ve 3’te daha stabil bir seyir izlemektedir (Şekil 4.1). Senaryo 1’de $n_R = 75-150-500-1,000-2,000-4,000$ veya 8,000 için elde edilen



Şekil 4.2. Panelde yalnız markerler bulunduğunda tahmin yöntemlerinin isabeti n_R : referans popülasyondaki birey sayısı; R^2 : y ve $\hat{u}$ arasındaki korelasyonun karesi; h^2_{sim} : simüle edilen özelliğin kalıtım derecesi; S1: markerler ve QTL sırası ile yüksek ve düşük MAF bakımından seçilmiştir; S2: markerler yüksek MAF bakımından seçilirken, QTL tamamen rassal olarak seçilmiştir; S3: markerler ve QTL tamamen rassal olarak seçilmiştir.



Şekil 4.3. Panelde markerlar ve QTL'ler bulunduğu GBLUP yönteminin isabeti n_R : referans popülasyondaki birey sayısı; h_M^2 : genomik kalıtım derecesi; R^2 : y ve $\hat{u}$ arasındaki korelasyonun karesi; h_{sim}^2 : simüle edilen özelliğin kalıtım derecesi; U (de los Campos vd 2013b) tarafından önerilen üst sınır, $[1 - (1 - b)^2]h^2$, S1: markerlar ve QTL sırası ile yüksek ve düşük MAF bakımından seçilmiştir; S2: markerlar yüksek MAF bakımından seçilirken, QTL tamamen rassal olarak seçilmiştir; S3: markerlar ve QTL tamamen rassal olarak seçilmiştir.

h_M^2 değerleri sırası ile 0.686, 0.661, 0.618, 0.640, 0.734, 0.773 ve 0.792'dir (Çizelge 4.1). Markerların yüksek allel frekansı için, buna karşılık QTL'lerin şansa bağlı olarak seçildiği Senaryo 2'de ise 75-8,000 aralığında değişen RP büyüklükleri için elde edilen h_M^2 değerleri sırası ile 0.757, 0.750, 0.819, 0.803, 0.802, 0.800 ve 0.800'dir (Çizelge 4.2). Marker ve QTL'lerin şansa bağlı olarak seçildiği Senaryo 3'te 75-8,000 aralığında değişen RP büyüklükleri için elde edilen h_M^2 değerleri sırası ile 0.748, 0.724, 0.790, 0.801, 0.803, 0.800 ve 0.801 olmuştur (Çizelge 4.3).

4.2.2. Tahmin yöntemlerinin karşılaştırılması

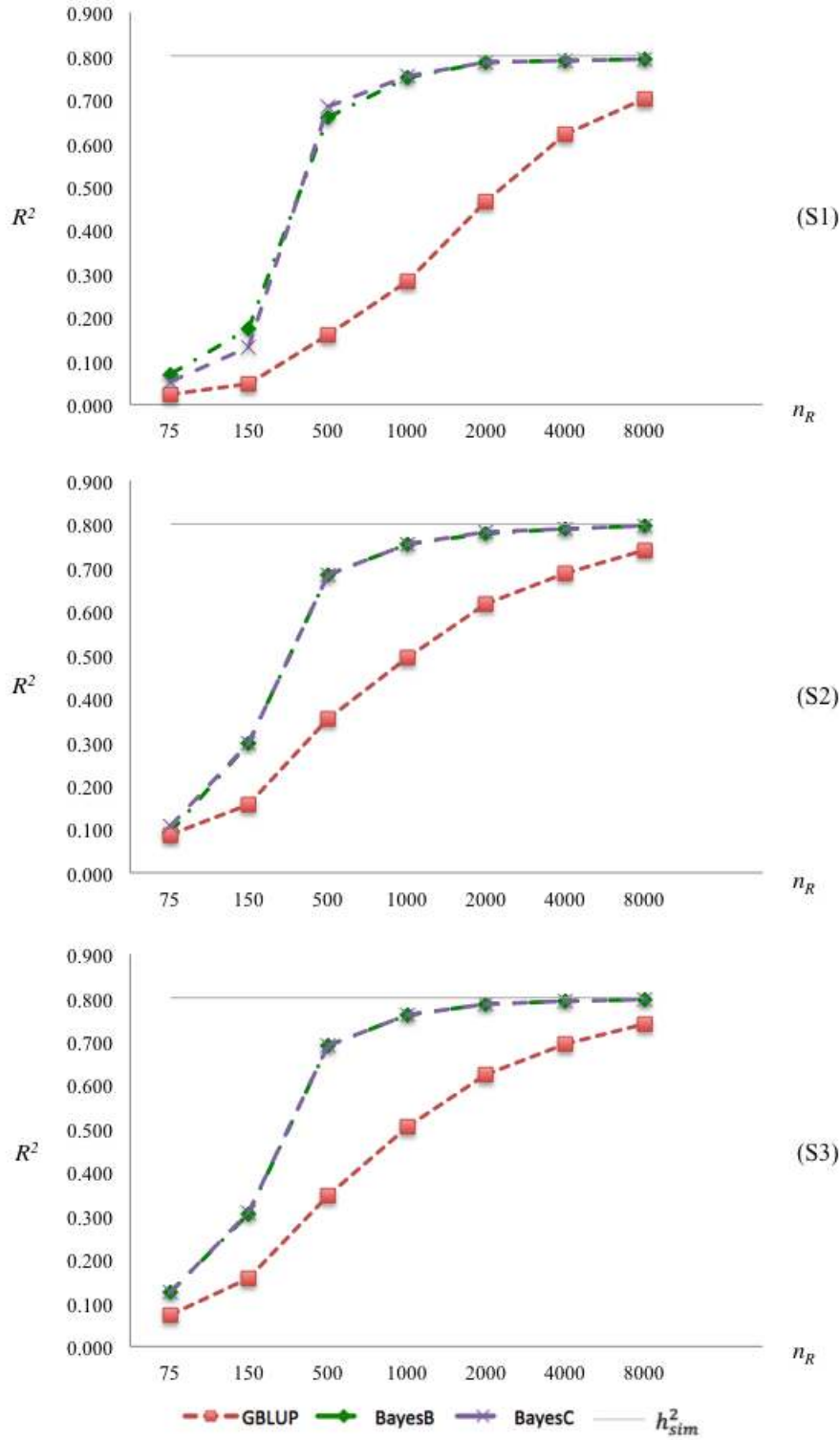
Şekil 4.4, çalışmada ele alınan GBLUP, BayesB ve BayesC yöntemleri için değişen RP büyüklüklerindeki R^2 değerleri grafiksel olarak gösterilmiştir. Senaryo 1'de GBLUP, BayesB ve BayesC yöntemlerinden elde edilen R^2 değerleri $n_R = 75$ için 0.023, 0.069 ve 0.052, $n_R = 150$ için 0.046, 0.174 ve 0.130, $n_R = 500$ için 0.159, 0.660 ve 0.683, $n_R = 1,000$ için 0.283, 0.749 ve 0.755, $n_R = 2,000$ için 0.465, 0.784 ve 0.786, $n_R = 4,000$ için 0.619, 0.789 ve 0.789, $n_R = 8,000$ için ise 0.702, 0.794 ve 0.794'dir (Çizelge 4.1).

Senaryo 2'de GBLUP, BayesB ve BayesC yöntemlerinden elde edilen R^2 değerleri $n_R = 75$ için 0.087, 0.092 ve 0.106, $n_R = 150$ için 0.155, 0.298 ve 0.296, $n_R = 500$ için 0.353, 0.682 ve 0.683, $n_R = 1,000$ için 0.494, 0.754 ve 0.753, $n_R = 2,000$ için 0.616, 0.780 ve 0.780, $n_R = 4,000$ için 0.686, 0.789 ve 0.789, $n_R = 8,000$ için ise 0.739, 0.796 ve 0.796'dir (Çizelge 4.2).

Senaryo 3'de GBLUP, BayesB ve BayesC yöntemlerinden elde edilen R^2 değerleri $n_R = 75$ için 0.072, 0.124 ve 0.123, $n_R = 150$ için 0.157, 0.304 ve 0.307, $n_R = 500$ için 0.345, 0.691 ve 0.691, $n_R = 1,000$ için 0.505, 0.760 ve 0.761, $n_R = 2,000$ için 0.624, 0.784 ve 0.784, $n_R = 4,000$ için 0.693, 0.792 ve 0.792, $n_R = 8,000$ için ise 0.740, 0.795 ve 0.795'dir (Çizelge 4.3).

4.3. Gerçekleşen İlişkiler ve Tahmin Edilebilirlik

Bu çalışmada test popülasyonu bireylerinin referans popülasyonu bireyleri ile olan ilişkilerinin genetik değer tahmini üzerine etkisini incelemek amacıyla 2,000 TP bireyi için iki farklı gruplandırma yapılmıştır. İlk olarak, her bir TP bireyinin $n_R = 500, 1,000, 2,000$ ve $5,000$ olduğu durumlarda RP bireyleri ile QTL düzeyindeki ilişkileri saptanmış, söz konusu bireyin en yüksek ilişki katsayısı, $\max(\mathbf{g}_{V_i})$, 0.15'ten düşük ise birey DI, 0.25 ve üzerindeyse YI grubuna atanmıştır. Sonuçlar kalıtım derecesinden bağımsız olmakla birlikte, $h^2 = 0.8$ olduğu ilk senaryoda 500-5,000 aralığında değişen RP büyüklükleri için DI grubunda sırası ile 1,347, 972, 546 ve 163, YI grubunda ise sırası ile 8, 17, 31 ve 69 birey bulunmaktadır (Sonuçlar 10 tekerrürün ortalaması olarak verilmiştir) (Çizelge 4.4). Benzer şekilde, kalıtım derecesinin 0.999 olarak simüle edildiği bir diğer senaryoda da 500-5,000 arasında değişen RP büyüklükleri için D grubunda sırası ile 1,342, 963, 558 ve 166, Y grubunda ise sırası ile 8, 15, 29 ve 64 birey bulunmaktadır.



Şekil 4.4. Panelde marker ve QTL'ler bulunduğu tahmin yöntemlerinin isabeti n_R : referans popülasyondaki birey sayısı; R^2 : y ve $\hat{u}$ arasındaki korelasyonun karesi; h^2_{sim} : simüle edilen özelliğin kalıtım derecesi; S1: markerler ve QTL sırası ile yüksek ve düşük MAF bakımından seçilmiştir; S2: markerler yüksek MAF bakımından seçilirken, QTL tamamen rassal olarak seçilmiştir; S3: markerler ve QTL tamamen rassal olarak seçilmiştir.

Özellik'in kalıtım derecesinin 0.8 olduğu durumda elde edilen R^2 değerleri $n_R = 500, 1,000, 2,000$ ve $5,000$ olduğu durumlarda DI grubu bireyleri için sırası ile 0.352, 0.477, 0.614 ve 0.699 iken, YI grubu bireyleri için sırası ile 0.439, 0.469, 0.614 ve 0.733 olarak elde edilmiştir. Kalıtım derecesi 0.999 olan ikinci özellik için elde edilen isabet dereceleri $n_R = 500, 1,000, 2,000$ ve $5,000$ olduğu durumlarda DI grubu bireyleri için sırası ile 0.542, 0.756, 0.932 ve 0.995 iken, YI grubu bireyleri için sırası ile 0.533, 0.704, 0.939 ve 0.996 olarak elde edilmiştir.

Çizelge 4.4'in son iki sütunu sırası ile DI ve YI bireylerine ait $r_{(\hat{u},u)}^2$ değerlerinin üst sınırlarını, $\bar{U}P$, vermektedir. Söz konusu değerler $h^2 = 0.8$ olduğu durumda 500, 1,000, 2,000 ve 5,000 bireylik RP büyüklüklerinde DI grubu için sırası ile 0.549, 0.777, 0.942 ve 1.000, YI bireyleri için ise sırası ile 0.569, 0.791, 0.947 ve 1.000 olarak elde edilmiştir. Kalıtım derecesi 0.999 olan özellik için elde edilen $\bar{U}P$ değerleri 500, 1,000, 2,000 ve 5,000 bireylik RP büyüklüklerinde DI grubu için sırası ile 0.550, 0.777, 0.942 ve 1.000, YI bireyleri için ise sırası ile 0.575, 0.792, 0.948 ve 1.000 olarak elde edilmiştir.

Çalışmamızda genetik değer tahmininde isabetin ölçüsü olarak bireylerin tahmin edilen genetik değerleri, $\hat{u}$, ve fenotipik değerleri, y , arasındaki korelasyonun karesi, $R_{(\hat{u},y)}^2$, kullanıldığından, DI ve YI gruplarındaki bireyler için hesaplanan $\bar{U}P_{DI}$ ve $\bar{U}P_{YI}$ değerleri h^2 ile çarpıldığında, sırası ile R_{DI}^2 ve R_{YI}^2 değerleri için üst sınırlar elde edilecektir. Kalıtım derecesi 0.8 olduğu durumda 500-5,000 aralığındaki n_R değerleri için R_{DI}^2 'nin üst sınırları 0.439, 0.622, 0.754 ve 0.800, R_{YI}^2 'nin üst sınırları ise 0.455, 0.633, 0.758 ve 0.800 olarak elde edilmiştir. Kalıtım derecesi 0.999 olduğu durumda 500-5,000 aralığındaki n_R değerleri için R_{DI}^2 'nin üst sınırları 0.549, 0.776, 0.941 ve 0.999, R_{YI}^2 'nin üst sınırları ise 0.574, 0.791, 0.947 ve 0.999 olarak elde edilmiştir.

Bu çalışmada TP ve RP bireyleri arasında genom düzeyindeki ikili ilişkilerin tahmin performansı üzerine etkisi en yüksek genomik ilişki katsayısı, $\max(\mathbf{g}_{V_i})$, bakımından iki grup, DI ve YI, üzerinden incelenmiştir (Çizelge 4.4). RP büyüklüğü arttıkça, $\max(\mathbf{g}_{T_i}) < 0.15$ olan birey sayısının, n_{DI} , azaldığı, buna karşın $\max(\mathbf{g}_{T_i}) \geq 0.25$ olan birey sayısının, n_Y , arttığı görülmüştür. $h^2 = 0.8$ olduğu durumda $n_R = 500$ ve $5,000$ için DI grubunda sırası ile 1,347 ve 163, YI grubunda ise sırası ile 8 and 69 birey yer almaktadır. Her iki grupta yer alan birey sayılarının artan RP büyüklükleri için izledikleri seyirler, RP büyüklüğü arttıkça TP bireyleri için referans popülasyonda genom düzeyinde yüksek ilişkili birey ya da bireyler bulunma ihtimalinin arttığını ortaya koymaktadır. Diğer yandan, $n_R = 5,000$ olduğu durumda dahi 2,000 TP bireyi arasından YI grubunda yer alanların sayısı $h^2 = 0.8$ ve $h^2 = 0.999$ durumları için sırası ile 69 ve 64'tur.

RP büyüklüğü arttıkça, tahmin isabeti hem DI hem de YI grupları için artış göstermiştir (Çizelge 4.4). $h^2 = 0.8$ durumunda $n_R = 500, 1,000, 2,000$ ve $5,000$ için DI grubuna ait R^2 değerleri, R_{DI}^2 , sırası ile 0.352, 0.477, 0.614 ve 0.699, YI grubuna ait R^2 değerleri, R_{YI}^2 , ise sırası ile 0.439, 0.469, 0.614 ve 0.733'tur. $h^2 = 0.999$

Çizelge 4.4. Farklı genomik ilişki grupları için tahmin isabeti ve tahmin edilebilirlik değerleri

h^2	n_R	n_T	n_{DI}	n_{YI}	R_{DI}^2	R_{YI}^2	UP_{DI}	UP_{YI}
0.8	500	2,000	1347	8	0.352(0.012)	0.439(0.083)	0.549(0.000)	0.569(0.003)
	1,000	2,000	972	17	0.477(0.006)	0.469(0.050)	0.777(0.000)	0.791(0.001)
	2,000	2,000	546	31	0.614(0.009)	0.614(0.043)	0.942(0.000)	0.947(0.000)
	5,000	2,000	163	69	0.699(0.011)	0.733(0.017)	1.000(0.000)	1.000(0.000)
0.999	500	2,000	1342	8	0.542(0.011)	0.533(0.080)	0.550(0.000)	0.575(0.002)
	1,000	2,000	963	15	0.756(0.005)	0.704(0.047)	0.777(0.000)	0.792(0.001)
	2,000	2,000	558	29	0.932(0.002)	0.939(0.008)	0.942(0.000)	0.948(0.000)
	5,000	2,000	166	64	0.995(0.000)	0.996(0.000)	1.000(0.000)	1.000(0.000)

n_R : referans popülasyondaki birey sayısı; n_T : test popülasyonundaki birey sayısı; n_{DI} ve n_{YI} sırası ile düşük ve yüksek ilişki gruplarındaki ortalama birey sayısı; R_{DI}^2 ve R_{YI}^2 sırası ile düşük ve yüksek ilişki gruplarındaki tahmin isabetleri; UP_{DI} ve UP_{YI} sırası ile düşük ve yüksek ilişki gruplarındaki ortalama UP değerleri

durumunda $n_R = 500, 1,000, 2,000$ ve $5,000$ için DI grubuna ait R^2 deęerleri, R^2_{DI} , sırası ile $0.542, 0.756, 0.932$ ve 0.995 , YI grubuna ait R^2 deęerleri, R^2_{YI} , ise sırası ile $0.533, 0.704, 0.939$ ve 0.996 'dir. Bu sonuçlar, TP ve RP bireyleri arasında genom düzeyindeki ikili ilişkilerin düşük olması durumunda dahi genetik deęer tahmininde yüksek isabet sağlanabileceğini ortaya koymaktadır.

5. TARTIŞMA

Tüm sonuçlar birlikte değerlendirildiğinde, R^2 için de los Campos vd (2013b) tarafından önerilen üst limitin, $[1 - (1 - b)^2]h^2$, geçerli olmadığı ve yeterli büyüklükte bir RP mevcut olduğu taktirde GBLUP yöntemi ile yakalanacak isabetin h_M^2 'yi yakalayabileceğini söylemek mümkündür.

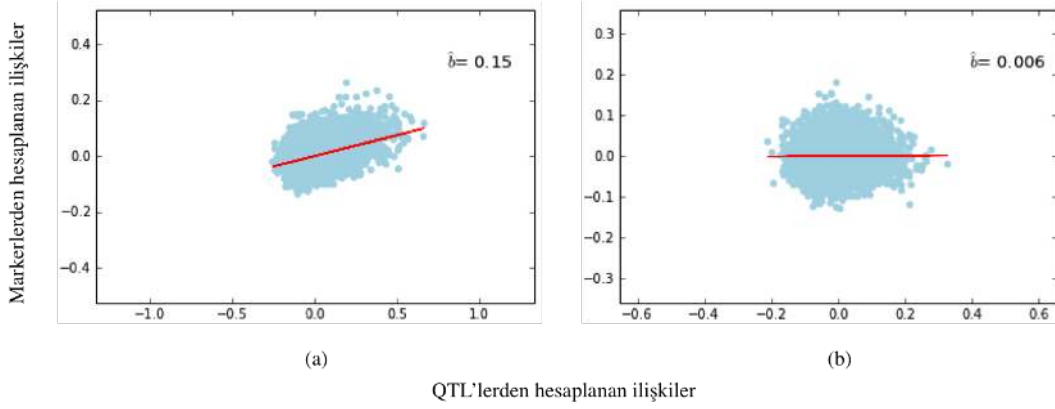
Formül (1.1) kullanılarak R^2 'nin asimptotik değerinin genomik kalıtım derecesi olduğu söylenebilir. Bu simülasyon çalışmasında ele alınan 3 senaryo için, $n_R = 8,000$, $M_e = \frac{2N_e L k}{\log(N_e L)} = 1448$, ve $h_M^2 = 0.698$, 0.724 ve 0.757 olmak üzere GBLUP yöntemi için elde edilen tahmini R^2 değerleri 0.554 , 0.579 ve 0.611 'tur. Bu değerler panelde yalnızca markerlerin bulunduğu durumda elde edilen R^2 değerlerinden (0.569 , 0.635 ve 0.673) ve panelde QTL'lerin de bulunduğu durumda elde edilen R^2 değerlerinden (0.702 , 0.739 ve 0.740) düşüktür.

Literatürde RP'da yer alan birey sayısı arttıkça eklemeli genetik değer tahmininde isabetin arttığını ortaya koyan çok sayıda çalışma mevcuttur (Goddard 2009; Daetwyler vd 2010b; Makowsky vd 2011). Benzer şekilde, çalışmamızda yalnızca marker genotiplerinin kullanıldığı durum için n_R 75'ten 8,000'e yükseltildiğinde R^2 'de yaklaşık 10-35 kat bir artış meydana gelmiştir. Meuwissen (2009), özellikle GBLUP yöntemi ile genetik değer tahmininde büyük referans popülasyonlar gerektiğini bildirmiştir. Buna karşılık, de los Campos vd (2013b), GBLUP yönteminin isabetini ele aldığı çalışmasında yalnızca 5,300 bireylik bir RP kullanarak 300,000 marker etkisini tahmin etmiştir. Daha önce de belirtildiği gibi genetik değer tahmininde isabeti etkileyen çok sayıda faktör olmasına karşın, söz konusu çalışmada elde edilen düşük R^2 değerlerinin muhtemel bir nedeni RP'nin yeteri kadar büyük olmamasıdır. Çalışmamızda, 4,200 marker etkisinin 75 bireylik bir RP kullanılarak tahmin edildiği durumda elde edilen R^2 değerleri S1 ve S2 için sırası ile 0.017 ve 0.064 olup, bu değerler de los Campos vd (2013b) tarafından GENOVA veri seti için bildirilenlere benzerdir. Diğer yandan, çalışmamızda artan RP büyüklükleri için genetik değer tahmininde isabetin, R^2 , arttığı ortaya konmuştur.

de los Campos vd (2013b), R^2 için bir üst limit ortaya koyduğu çalışmasında TP bireyleri için genetik değerlerin koşullu beklenen değerini

$$E(u_{T_i} | \mathbf{y}_R) = \mathbf{g}_{Q,i}[\mathbf{G}_Q \sigma_u^2 + \mathbf{I} \sigma_e^2]^{-1} \mathbf{y}_R, \quad (5.1)$$

şeklinde vermiştir. Burada $\mathbf{G}_Q$ RP bireylerinin QTL düzeyindeki ilişkilerini içeren matris ve $\mathbf{y}_R$ RP bireylerinin merkezileştirilmiş fenotipik değerleridir. Uygulamada bireylerin QTL genotipleri bilinmediğinden bireyler arasındaki genomik ilişkiler QTL'ler yerine markerler kullanılarak hesaplanmaktadır. Bu durumda, genetik değerlerin koşullu beklenen değeri aşağıdaki gibi yazılır:



Şekil 5.1. Senaryo 1'deki bir tekerrürde ($n_R = 8,000$) test popülasyonundaki bir bireyin referans popülasyondaki tüm bireylerle marker ve QTL düzeyindeki ilişkilerine ait serpilme grafiği
Kırmızı çizgi regresyon doğrusu olup, (a) ve (b) sırası ile en büyük ve en küçük $\hat{b}$ değerine sahip bireylere ait grafikleri göstermektedir.

$$E(u_{T_i} | \mathbf{y}_R) \approx \mathbf{g}_{M,i} [\mathbf{G}_M \sigma_u^2 + \mathbf{I} \sigma_e^2]^{-1} \mathbf{y}_R. \quad (5.2)$$

Burada, $\mathbf{G}_M$ RP bireylerinin marker düzeyindeki ilişkilerini içeren matristir. Ancak, de los Campos vd (2013b) tarafından R^2 için bir üst limit türetilirken $\mathbf{G}_Q$ 'nun bilindiği varsayılmış ve $\mathbf{g}_{M,i}$, $\mathbf{g}_{M,i} = b_i \mathbf{g}_{Q,i} + \epsilon_i$ şeklinde yazılmıştır:

$$E(u_{T_i} | \mathbf{y}_R) \approx b_i \mathbf{g}_{Q,i} [\mathbf{G}_Q \sigma_u^2 + \mathbf{I} \sigma_e^2]^{-1} \mathbf{y}_R. \quad (5.3)$$

Birbiri ile ilişkisiz bireylerden oluşan bir popülasyonda genom düzeyinde ilişkilerin beklenen değeri sıfırdır. Diğer yandan, genom düzeyindeki ilişkiler, $\mathbf{g}_{M,i}$ ve $\mathbf{g}_{Q,i}$, (de los Campos vd 2013b)'da olduğu gibi 300,000 marker ve 5,000 QTL kullanılarak hesaplandığında, $\mathbf{g}_{Q,i}$ 'nin varyansı $\mathbf{g}_{M,i}$ 'nin varyansından daha büyük olacaktır. Bu durumda da $\mathbf{g}_{M,i}$ ve $\mathbf{g}_{Q,i}$ için uydurulacak regresyon doğrusunun eğim katsayısı, b , küçük olacaktır (Şekil 5.1). Bundan dolayı da (5.3)'den elde edilecek R^2 değeri (5.1)'den çok daha küçük olacaktır. Buradan hareketle, uygun yaklaşım (5.2)'de hem $\mathbf{g}_{Q,i}$ hem de $\mathbf{G}_Q$ 'nun markerlerden kullanılarak hesaplanan karşılıklarının kullanılması olacaktır. Bu durumda, b katsayısı b_i 'lerin ortalama değeri olmak üzere, $\mathbf{g}_{M,i} = b \mathbf{g}_{Q,i}$ ve $\mathbf{G}_M = b \mathbf{G}_Q$, olacak ve (5.2)

$$E(u_{T_i} | \mathbf{y}_R) \approx \mathbf{g}_{Q,i} [\mathbf{G}_Q \sigma_u^2 + \mathbf{I} \frac{\sigma_e^2}{b}]^{-1} \mathbf{y}_R. \quad (5.4)$$

şeklinde elde edilecektir. Koşullu beklenen değere ilişkin bu yakınsamanın (5.4) eşitlik (5.1)'le hemen hemen aynı olduğu ve bu nedenle de (5.4)'den elde edilecek R^2 değerinin (5.1)'den elde edilecek R^2 değerine benzer olacağı görülmektedir.

van den Berg (2016), farklı sığır türleri için ırklar arası tahminde isabeti

ele aldıkları çalışmalarında, r^2 'deki azalmayı incelemişlerdir. Söz konusu çalışmada eşitlik 3.1'den tahmin edilen b katsayısı kullanılarak hesaplanan r^2 'deki minimum azalma ve gerçekte gözlenen durum incelenmiştir. Araştırmacılar, özellik görece az sayıda gen (50 ve 100) tarafından kontrol edildiğinde de los Campos vd (2013b) tarafından önerilen ve r^2 'deki minimum azalmaya işaret eden katsayının bazı durumlarda gerçekte gözlenenden daha büyük olduğunu bildirmişlerdir (van den Berg 2016). Bir başka ifade ile, gerçekte gözlenen r^2 değerleri, de los Campos vd (2013b) tarafından belirtilen $[1 - (1 - b)^2]$ kullanılarak elde edilecek değerlerden daha yüksek bulunmuştur.

Ticari kullanıma sunulan SNP panellerinin içerdiği marker sayısı, ya da bir başka ifadeyle SNP ciplerinin marker yoğunluğu, sürekli olarak artmaktadır. Diğer yandan, sekans verileri kullanıldığında model uydurmada kullanılan markerlerin QTL'leri içermesi de muhtemeldir. Ancak bu durumda kullanılacak SNP'lerin seçimi önemlidir (Druet vd 2014; van den Berg 2016). Buradan hareketle, mevcut çalışmada bireylerin QTL'lerinin de panelde olduğu durum dikkate alınmıştır. Bu durumda dahi GBLUP yöntemi ile sağlanan isabet çalışmada ele alınan RP büyüklüklerinde h_M^2 değerine ulaşamamıştır. Ancak, RP arttıkça R^2 'nin izlemiş olduğu seyre bakıldığında, yeterli büyüklükte bir RP sağlandığında GBLUP yönteminde elde edilecek isabetin genomik kalıtım derecesine erişebileceğini söylemek mümkündür.

RP yalnızca 75 birey içerdiği durumda her 3 yöntemin isabet derecesi birbirine yakın ve oldukça düşüktür. Bu durumda BayesB ve BayesC gibi değişken seçimine dayalı yöntemlerin GBLUP yöntemine karşı bir üstünlüğü olmadığı görülmektedir. Diğer yandan, n_R arttıkça BayesB ve BayesC yöntemleri GBLUP yöntemine göre daha yüksek isabet sağlamış, ancak $n_R = 8,000$ olduğu durumda söz konusu yöntemlerin üstünlükleri bir kez daha ortadan kalkmış ve her 3 yöntem de benzer R^2 değerleri ortaya koymuştur.

Daetwyler vd (2010b), GBLUP yönteminin isabetinin, etkili kromozom segmentlerinin sayısına, M_e , önemli ölçüde bağlı olduğunu, buna karşın BayesB yönteminin isabetinin M_e ile birlikte QTL sayısına da, n_{QTL} bağlı olduğunu ortaya koymuştur. Söz konusu çalışmada, Daetwyler vd (2008)'de verilen eşitlik yeniden düzenlenerek BayesB yöntemi ile genetik değer tahmininde elde edilecek isabet için bir yakınsama önerilmiştir: $n_R h^2 / [n_R h^2 + \min(n_{QTL}, M_e)]$. Daetwyler vd (2010b)'a göre $n_{QTL} < M_e$ olduğu durumda BayesB ve BayesC gibi değişken seçimine dayalı yöntemler, M_e tane parametreyi tahmin etmek yerine uydurulan markerler arasından özellik üzerinde etkili lokusları içeren bir alt kümeyi sectiklerinden bu yöntemlerden elde edilecek isabet de yüksek olacaktır. $n_R = 75$, $h^2 = 0.8$, $n_{QTL} = 70$ ve $M_e = 1,448$ değerleri Daetwyler vd (2010b) tarafından verilen eşitlikte yerine konulduğunda, BayesB ve GBLUP yöntemleri için tahmin edilen R^2 değerleri sırası ile 0.370 ve 0.032 olarak elde edilir. Ancak, çalışmamızda $n_R = 75$ olduğu durumda BayesB ve BayesC yöntemleri için elde edilen R^2 değerleri her üç senaryoda da, S1-S3, tahmin edilen bu R^2 değerlerinden oldukça düşüktür (0.027-0.077). Bu durumun kaynağını incelemek amacıyla Senaryo 1'deki bir BayesB analizi sonucunda

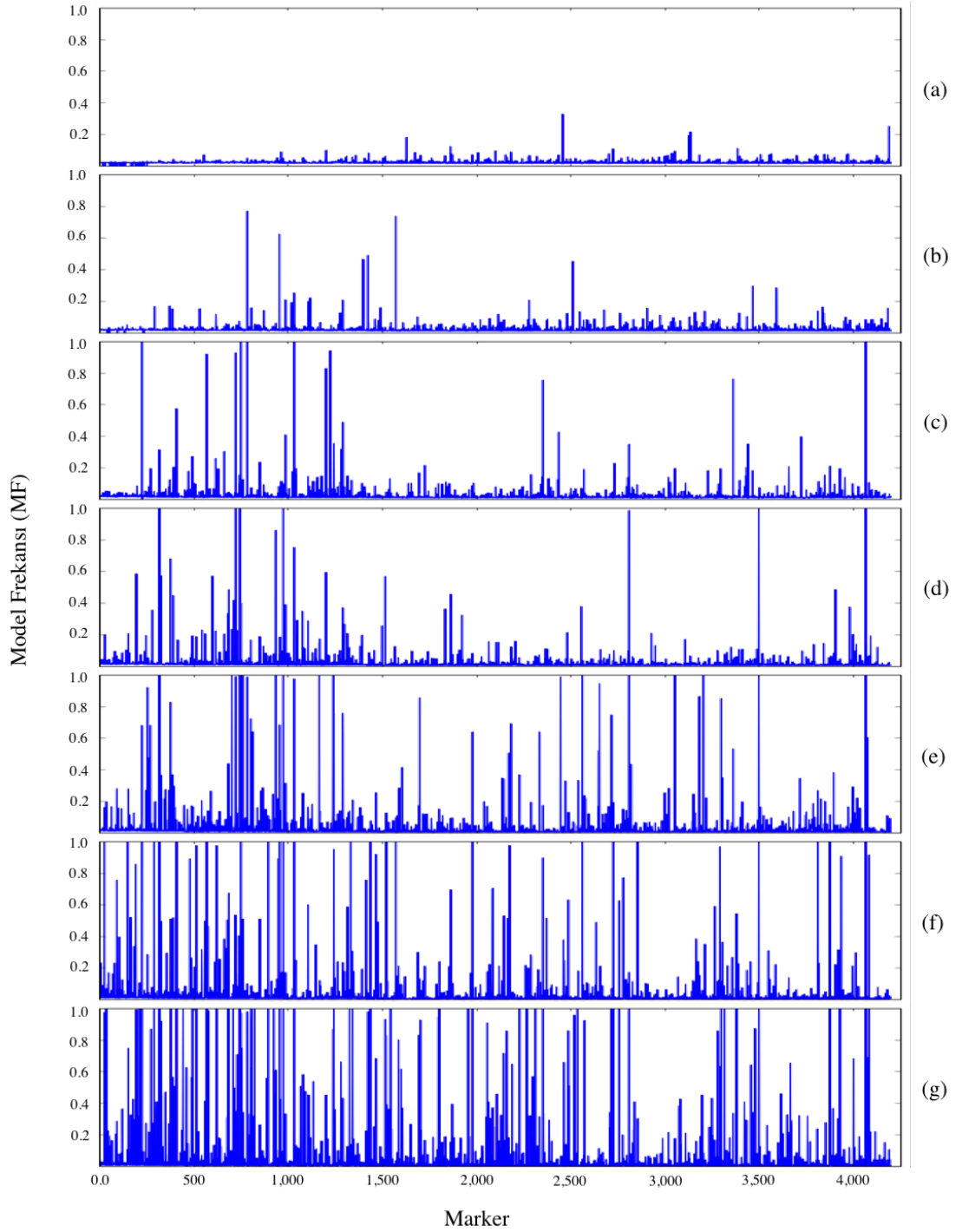
ortaya çıkan model frekansları, MF, incelenmiştir. Şekil 5.2 değişen n_R büyüklükleri için markerlere ilişkin MF'ni göstermektedir. $n_R = 75$ olduğu durumda MF'nin tekdüze bir dağılışı izlediği ve $MF > 0.4$ olan herhangi bir marker olmadığı görülmektedir. Buna karşın, $n_R = 8,000$ olduğunda çok sayıda marker için $MF > 0.4$ olmuştur. RP'daki birey sayısı 75'ten 8,000'e çıktığında markerlere ait MF değerlerinin varyansı yaklaşık 100 kat artmış ve 1.21×10^{-4} 'ten 214×10^{-4} 'de yükselmiştir. Bu sonuçlar ışığında, n_R küçük olduğunda veri indirgemeye dayalı yöntemlerin (örneğin BayesB) QTL'ler ile LD içerisinde olan markerleri doğru belirleyemediğini söylemek mümkündür. Dolayısı ile yeterli büyüklükte bir RP sağlanana kadar söz konusu yöntemleri kullanmanın bir avantajı olmayacaktır. RP büyüklüğünde 75'ten 8,000'e bir artış BayesB ve BayesC yöntemlerinden elde edilen R^2 değerlerinde 10-20 kat bir artış sağlamıştır (Çizelge 4.1-4.3).

RP büyüklüğü 8,000 olduğunda çalışmada ele alınan yüksek kalıtım dereceli özellik için her 3 yöntem de benzer sonuçlar vermiştir. BayesB ve BayesC gibi değişken seçimine dayalı yöntemler çok küçük etkiye sahip lokusların etkisini 0'a eşitlemekte ve bu lokuslar genetik değer tahmininde hesaba katılmamaktadır. Diğer yandan, GBLUP etkinin büyüklüğünü dikkate almaksızın tüm lokusların etkisini tahmin etmekte ve bu durum tahminlerde yalnızca bozucu bir etki yapmaktadır. Bu nedenle $n_{marker} \gg n_R$ olduğunda BayesB ve BayesC gibi değişken seçimine dayalı yöntemler GBLUP'tan daha iyi sonuçlar vermektedir. Yeterli büyüklükte bir referans popülasyon sağlandığında GBLUP etkisi küçük olan lokuslara ait etkileri de isabetli bir şekilde tahmin etmekte ve değişken seçimine dayalı yöntemler ile benzer sonuçlar vermektedir.

Panel QTL'leri de içerdiği durumda GBLUP ile BayesB ve BayesC yöntemlerinden elde edilen tahmin arasındaki farklar panelde yalnızca markerlerin bulunduğu durumdakinden daha büyüktür (Çizelge 4.1-4.3). Bu durumda BayesB ve BayesC, GBLUP yönteminin 8,000 bireylik bir RP ile ulaştığı R^2 değerlerine daha küçük bir RP ile ulaşmıştır (500-1,000).

Ober vd (2011) BayesB yönteminden elde edilen tahmin isabetinin GBLUP yönteminden elde edilen tahmin isabetine göre daha yüksek olmadığını bildirmiştir. Söz konusu çalışmada yaklaşık 2.5 milyon marker kullanılmış olmasına karşın $n_R = 155$ 'tir. Çalışmamızda ortaya koyduğumuz sonuçlar ışığında Ober vd (2011) tarafından bildirilen bu durumun RP büyüklüğünden kaynaklandığını söylemek olasıdır. Genetik değer tahmin yöntemlerinin isabetini karşılaştıran çok sayıda çalışma mevcuttur (Luan vd 2009; Meuwissen 2009; Daetwyler vd 2010b; Habier vd 2010b; Meuwissen ve Goddard 2010; Zhang vd 2010). Tüm çalışmalar birlikte ele alındığında, hiçbir yöntemin “en iyi” olmadığı ve bir yöntemin performansının RP büyüklüğü yanında üzerinde çalışılan özelliğin genetik yapısı ve kalıtım derecesine bağlı olduğunu söylemek mümkündür.

Bu çalışmada genetik değer tahmini sabit etkiler modelindeki tahmin edilebilirlik konusu şeklinde ele alınarak tahmin isabeti için bir üst sınır ortaya



Şekil 5.2. Senaryo 1’deki bir tekrürde BayesB yöntemi için markerlerin model frekansları

Senaryo 1 (S1): markerler ve QTL sırası ile yüksek ve düşük MAF bakımından seçilmiştir; (a) - (g): $n_R = 75 - n_R = 8,000$.

konmuştur. Çalışmanın sonuçları göstermiştir ki genetik değer tahmininde isabet TP bireylerinin genotiplerinin RP bireylerine ait genotipleri içeren matrisin satırlarının doğrusal bir kombinasyonu olarak ne ölçüde yazılabildiğine bağlıdır. RP büyüklüğü arttıkça $\mathbf{X}_R$ 'in satır uzayı da genişleme eğiliminde olacak ve bir TP bireyine ait genotip vektörünün, $\mathbf{x}_T$, $\mathbf{X}_R$ 'nin satır uzayında olma olasılığı da artacaktır. Bu sonuçlar ışığında, genetik değer tahmininde isabetin bir ölçüsü olarak kullanılan R^2 için üst sınır $h^2\bar{U}P$ şeklindedir.

Literatürde TP ve RP bireyleri arasındaki akrabalık derecesinin genetik değer tahmininde isabet üzerinde etkili olduğunu gösteren bazı çalışmalar mevcuttur. Habier vd (2007; 2010b), gerçek veriler ve simülasyonlar kullanılarak TP ve RP bireyleri arasındaki yüksek akrabalığın genetik değer tahmininde isabeti artırdığını bildirmiştir. Legarra vd (2008), farelerde yaptığı çalışmada aile içi tahminlerin aileler arası tahmine göre daha yüksek olduğunu bildirmiştir. Clark vd (2011), damızlık değer tahmininin TP ve RP bireyleri arasındaki ilişkinin derecesine bağlı olduğunu bildirmiştir. Pszczola vd (2011) RP bireyleri arasında ve TP-RP bireyleri arasında farklı akrabalık düzeylerini incelemiş, genetik değer tahmininde yüksek isabet için RP ve TP bireyleri arasındaki akrabalığın da yüksek olması gerektiğini bildirmiştir. Makowsky vd (2011), TP bireylerinin RP'nda yakın akrabalarının bulunması durumunda tahmin isabetinin arttığını bildirmiştir. Diğer yandan, Luan vd (2009) ve Perez-Cabal vd (2012), farklı sığır ırkları için genetik değer tahminleri gerçekleştirmiş ve daha önce bahsedilen çalışmaların aksine TP ve RP arasındaki ilişkinin tahmin isabeti üzerine etkisi için önemli bir bulguya ulaşmamışlardır. Bu çalışmada, TP bireylerinin genotipleri $\mathbf{X}_R$ 'nin satır uzayında olduğu müddetçe, özelliğin kalıtım derecesi ve RP büyüklüğüne bağlı olarak genetik değer tahmininde yüksek isabet sağlanabileceği ortaya konulmuştur. Bu durum TP ve RP bireyleri arasında genom düzeyindeki mevcut olan ikili ilişkilerin büyüklüğünden bağımsızdır. Genetik değer tahmininde isabet için bu çalışmada ortaya konan teorik altyapı ve simülasyon çalışması sonuçları, Calus (2010) tarafından yapılan “tahminlerin isabetli olabilmesi için mümkün tüm genotip ve fenotipleri barındıran bir RP veri seti kullanılmalıdır” önerisi ile uyum göstermektedir.

6. SONUÇ

Son yıllarda moleküler genetik biliminde ve bilgisayar teknolojilerinde meydana gelen gelişmelere bağlı olarak gerek insanlarda, gerekse bitki ve hayvan türlerinde pek çok önemli özellikle ilişkili kromozom bölgelerinin belirlenebilmesi amacıyla genom boyu ilişki analizi (Genome wide association study-GWAS) mümkün hale gelmiştir. Bu çalışmalar sonucunda, insanlarda ve pek çok çiftlik hayvanı türünde yüzlerce özellikle ilişkili olduğu bildirilen binlerce gen tespit edilmiştir. Buna karşın; insan, bitki ve hayvanlarda üzerinde durulan pek çok özellik karmaşık yapıdadır ve bu özelliklerin ortaya çıkmasında her biri küçük etkiye sahip çok sayıda gen rol oynamaktadır. Dolayısıyla da GWAS çalışmalarıyla belirlenen sınırlı sayıdaki kantitatif özellik lokusunun (quantitative trait loci-QTL) ortaya çıkan fenotipik varyasyonu açıklama oranı oldukça düşüktür.

Genomik değerlendirme, genetik değer tahmininde QTL'lerin tüm genoma yayılan ve sayıları on binleri bulan markerlerden en az biri ile bağlantı dengesizliği (linkage disequilibrium-LD) içerisinde olduğu varsayımına dayanmaktadır. GWAS çalışmalarının aksine, genomik değerlendirmede QTL'lerin belirlenmesi değil, insanlarda çeşitli özelliklere ilişkin, ya da bitki ve hayvan türlerinde seleksiyon adaylarına ait genetik değerlerin tüm genoma yayılan markerler yardımıyla tahmin edilmesi amaçlanmaktadır. Bu amaçla, genellikle tüm genom regresyon yöntemleri adı verilen Bayesci yöntemler yaygın olarak kullanılmaktadır.

Genomik değerlendirmede başarı genetik değer tahminlerinin isabetli bir şekilde tahmin edilmesine bağlıdır. Genomik değerlendirmede isabeti henüz veri toplanmadan belirlemek amacıyla önerilmiş çok sayıda yakınsama bulunmakta olup, bunlardan en yaygın kullanılanlarına tezin ilgili kısımlarında değinilmiştir. Bu çalışmaların öncülüğünü sabit etkiler modeli ve lokusların birbirinden bağımsızlığı altında Daetwyler vd (2008) yapmış olup, başka bir çalışmada ise (Goddard 2009) lokuslar bağımsız varsayıldığı durumda etkin kromozom bölgelerinin sayısının, M_e , kullanılabileceğini bildirmiştir. Söz konusu yakınsamalar; referans popülasyon (RP) büyüklüğü (n_R), özelliğin kalıtım derecesi (h^2) ve bağımsız lokusların ya da etkin kromozom bölgelerinin sayısına bağlıdır. Marker ve QTL'ler arasında tam/kusursuz bir LD varsayımının sağlanması en azından mevcut SNP panelleri dikkate alındığında mümkün olamamıştır. Örneğin yüksek yoğunluklu SNP panelleri ile elde edilen isabetler, beklentilerin aksine orta yoğunluklu paneller ile elde edilenlerden önemli ölçüde yüksek bulunmamıştır. Bunun muhtemel bir sebebi, pek çok QTL'in düşük minör allel frekansına (MAF) sahip olması, buna karşın panellerde yer alan SNP'lerin daha yüksek allel frekansları bakımından seçilmesidir.

Genomik değerlendirmede isabeti etkileyen faktörleri (i) araştırmacının kontrolünde olmayan ve popülasyona özgü faktörler ve (ii) araştırmacının kontrolünde olan faktörler olarak iki ayrı grupta değerlendirmek mümkündür. Birinci grupta; ortalama kromozom uzunluğu (L), her bir popülasyona özgü bir büyüklük olan ve popülasyonun evrilme sürecine bağlı olan efektif popülasyon

büyüklüğü (N_e), ilgilenilen özelliği etkileyen QTL sayısı (n_{QTL}) ve özelliğin kalıtım derecesi (h^2) yer almakta iken, ikinci grupta ise; RP'un büyüklüğü ve yapısı, marker yoğunluğu ve kullanılan tahmin yöntemi yer almaktadır. Genomik değerlendirme, ilk grupta yer alan faktörler dikkate alınarak uygun bir çalışma dizayn edilebildiği ölçüde başarılı olacaktır.

Çalışmamız sonucunda şu önemli sonuçlara ulaşılmıştır:

1) Yeterli büyüklükte bir RP mevcut olduğunda, R^2 genomik kalıtım derecesine (h_M^2) yaklaşacaktır.

2) Güvenilirlik, r^2 , için üst sınır:

$$UP_i = \frac{\mathbf{x}_{T_1}' \mathbf{x}_{T_1}}{\mathbf{x}_T' \mathbf{x}_T} \quad (6.1)$$

şeklinde. $UP_i = 0$ olduğunda tahminin isabet derecesi RP'un büyüklüğünden bağımsız olarak «0» olacaktır. Bununla birlikte, $UP_i = 1$ olduğunda RP'daki birey sayısı arttıkça r^2 de 1'e yaklaşacaktır. Diğer yandan, $UP_i < a$, olduğunda r^2 RP'nin büyüklüğünden bağımsız olarak a 'dan daha küçük olacaktır.

3) Genomik değerlendirmede isabetin bir ölçüsü olan R^2 'nin üst sınırı $h^2 UP_i$ 'dir.

4) Genomik değerlendirmede, RP'de yakın akrabası bulunmayan bireyler için de yüksek isabetli tahminler yapmak mümkündür.

5) TP bireylerinin genotipleri $\mathbf{X}_R$ 'nin satır uzayında olduğu müddetçe, özelliğin kalıtım derecesi ve RP büyüklüğüne bağlı olarak genetik değer tahmininde yüksek isabet sağlanması mümkündür.

6) QTL etkileri normal dağılış gösterdiği durumda, GBLUP yöntemi, RP yeterli sayıda birey barındırdığı taktirde en az Bayesci değişken seçim yöntemleri kadar isabetli tahminler sağlamaktadır. Asimptotik olarak tüm yöntemler benzer sonuçlar vermektedir.

7) QTL etkileri normal dağılış gösterdiği durumda, genomik değerlendirmede Bayesci değişken seçim yöntemlerinden elde edilecek isabetin GBLUP yönteminden elde edilecek isabetten yüksek olması, yeterli büyüklükte bir RP bulunmasına bağlıdır. Bu durum sağlanmadığı taktirde Bayesci değişken seçim yöntemlerinin GBLUP yöntemine karşı bir avantajı olmayacaktır.

8) QTL ve markerlere ait MAF dağılımları benzer olduğunda genomik değerlendirmede isabet, QTL'lerden farklı MAF dağılımına sahip markerler kullanılarak elde edilenden daha yüksek olacaktır.

7. KAYNAKLAR

- AFFYMETRICS, INC. Agrigenomic solutions. http://www.affymetrix.com/estore/browse/level_three_category_and_children.jsp?category=cat380018&categoryIdClicked=cat380018&expand=true&parent=35923 [Son erişim tarihi: 10.10.2016].
- AGUILAR, I., MISZTAL, I., JOHNSON, D.L., LEGARRA, A., TSURUTA, S., and LAWLOR, T.J. 2010. A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of holstein final score. *J Dairy Sci*, 93:743-752.
- ANONIM. DNA Genes, and SNPs <https://www.coriell.org/personalized-medicine/dna-genes-and-snps> [Son erişim tarihi: 10.10.2016].
- AZIZIAN, S., SHADPARVAR A.A., HOSSEIN-ZADEH, N.G., and JOEZY-SHEKALGORABI, S. 2016. Effect of increasing accuracy of genomic evaluations on economic efficiency of dairy cattle breeding programs. *Ital J Anim Sci*, 15(3):379-385.
- BASTIAANSEN, J.W.M., BINK, M.C.A.M., COSTER, A., MALIEPAARD, C., and CALUS, M.P.L. 2010. Comparison of analysis of the qtlmas xiii common dataset. i: Genomic selection. *BMC Proc* 4(Suppl. 1): S1.
- BERMINGHAM, M.L., PONG-WONG, R., SPILIOPOULOU, A., HAYWARD, C., RUDAN, I., CAMPBELL, H., et al. 2015. Application of high-dimensional feature selection: evaluation for genomic prediction in man. *Sci Rep*, 5(10312).
- BERNARDO R., and YU, J. 2007. Prospects for genome-wide selection for quantitative traits in maize. *Crop Sci*, 47:1082-1090.
- BRITO, F.V., NETO, J.B., SARGOLZAEI, M., COBUCI, J.A., and SCHENKEL, F.S. 2011. Accuracy of genomic selection in simulated populations mimicking the extent of linkage disequilibrium in beef cattle. *BMC Genet*, 12(80).
- CALUS, M.P.L. 2010. Genomic breeding value prediction: methods and procedures. *Animal*, 4(2):157-164.
- CALUS, M.P.L., and VEERKAMP, R.F. 2007. Accuracy of breeding values when using and ignoring the polygenic effect in genomic breeding value estimation with a marker density of one snp per cm. *J Anim Breed Genet*, 124:362-368.
- CHENG, H., GARRICK, D., and FERNANDO, R. 2015. Xsim: Simulation of descendants from ancestors with sequence data. *G3: Genes/Genomes/Genetics*, 5(7):1415-1417.

- CHRISTENSEN, O., and LUND, M. 2010. Genomic prediction when some animals are not genotyped. *Genet Sel Evol*, 42(2).
- CLARK, S.A., HICKEY, J.M., and VAN DER WERF, H.J. 2011. Different models of genetic variation and their effect on genomic evaluation. *Genet Sel Evol*, 43(18).
- CLARK, S.A., HICKEY, J.M., DAETWYLER, HD., and VAN DER WERF, J.H. 2012. The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. *Genet Sel Evol*, 44(4).
- CROSSA, J., DE LOS CAMPOS, G., PEREZ, P., GIANOLA, D., BURGUENO, J., et al. 2010. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics*, 186:713-724.
- DAETWYLER, H.D., VILLANUEVA, B., and WOOLLIAMS, J.A. 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS ONE*, 3(10):e3395.
- DAETWYLER, HD., HICKEY, JM., HENSHALL, JM., DOMINIK, S., GREDLER, B., et al. 2010a. Accuracy of estimated genomic breeding values for wool and meat traits in a multi-breed sheep population. *Anim Prod Sci*, 50:1004-1010.
- DAETWYLER, H.D., PONG-WONG, R., VILLANUEVA, B., and WOOLLIAMS, J.A. 2010b. The impact of genetic architecture on genome-wide evaluation methods. *Genetics*, 185(3):1021-1031.
- DE LOS CAMPOS, G., PONG-WONG, R., HICKEY, J.M., DAETWYLER, HD., and CALUS, M.P.L. 2013a. Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics*, 193(2):327-345.
- DE LOS CAMPOS, G., VAZQUEZ, A.I., FERNANDO, R.L., KLIMENTIDIS, Y.C., and SORENSEN, D. 2013b. Prediction of complex human traits using the genomic best linear unbiased predictor. *PLoS Genet*, 9(7):e1003608.
- DE LOS CAMPOS, G., SORENSEN, D., and GIANOLA, D. 2015. Genomic heritability: What is it? *PLoS Genet*, 11:e1005048.
- DE ROOS, A.P.W., HAYES, B.J., SPELMAN, R., and GODDARD, M.E. 2008. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. *Genetics* 179:1503-1512.
- DEKKERS, J.C.M. 2007. Prediction of response to marker assisted and genomic selection using selection index theory. *J Anim Breed Genet*, 124:591-611.
- DRUET, T., MACLEOD, I.M., and HAYES, R.J. 2014. Toward genomic prediction from whole-genome sequence data: impact of sequencing design on genotype imputation and accuracy of predictions. *Heredity*, 112: 39-47.

- FALCONER, D.S., and MACKAY, T.F.C. 1996. Introduction to quantitative genetics. Benjamin Cummings.
- FAN, B., ZHI-QIANG, D.U., GORBACH, D.M., and ROTHSCILD, M.F. 2010. Development and application of high-density snp arrays in genomic studies of domestic animals. *Asian-Aust J Anim Sci*, 23(7):833-847.
- FERNANDO, R.L. 1998. Genetic evaluation and selection using genotypic, phenotypic and pedigree information. Genetics Applied to Livestock Production (WCGALP), Armidale, Australia, 11-16 January 1998.
- FERNANDO, R.L., and GARRICK, D.J. 2013. Bayesian Methods Applied to GWAS. In Genome-Wide Association Studies and Genomic Prediction. Springer: Humana Press.
- FERNANDO, R.L., DEKKERS, J.C.M., and GARRICK, D.J. 2014. A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole genome analyses. *Genet Sel Evol*, 46(50).
- GARRICK, D. 2011. The nature, scope and impact of genomic prediction in beef cattle in the united states. *Genet Sel Evol*, 43(17).
- GARRICK, D.J., and FERNANDO, R.L. 2013. Implementing a QTL detection study (GWAS) using genomic prediction methodology. In Genome-Wide Association Studies and Genomic Prediction. Springer: Humana Press.
- GARRICK, D., DEKKERS, J., and FERNANDO, R. 2014. The evolution of methodologies for genomic prediction. *Livest Sci*, (166):10-18.
- GIANOLA, D., DE LOS CAMPOS, G., HILL, W.G., MANFREDI, E., and FERNANDO, R. 2009. Additive genetic variability and Bayesian alphabet. *Genetics*, 183(1):347-363.
- GODDARD, M. 2009. Genomic selection: prediction of accuracy and maximization of long term response. *Genetica*, 136:245-257.
- GODDARD, M., and HAYES, B.J. 2007. Genomic selection. *J Anim Breed Genet*, 124:323-330.
- GODDARD, M.E., and HAYES, B.J. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics*, 10:381-391.
- GODDARD, M.E., HAYES, B.J., and MEUWISSEN, T.H.E. 2011. Using the genomic relationship matrix to predict the accuracy of genomic selection. *J Anim Breed Genet*, 128:409-421.

- GONZALEZ-RECIO, O., GIANOLA, D., LONG, N., WEIGEL, K.A., ROSA, G.J., et al. 2008. Nonparametric methods for incorporating genomic information into genetic evaluations: an application to mortality in broilers. *Genetics*, 178:2305-2313.
- GRINBERG, N.F., LOVATT, A., HEGARTY, M., LOVATT, A., SKOT, P.K., KELLY, R., et al. 2016. Implementation of genomic prediction in *Lolium perenne* (L.) breeding populations. *Front Plant Sci*, 7:133.
- HABIER, D., FERNANDO, R.L., and DEKKERS, J.C. 2007. The impact of genetic relationship information on genome-assisted breeding values. *Genetics*, 177:2389-2397.
- HABIER, D., FERNANDO, R.L., KIZILKAYA, K., and GARRICK, D.J. 2010a. Extension of the Bayesian alphabet for genomic selection. *BMC bioinformatics*, 12.
- HABIER, D., TETENS, J., SEEFRIED, F.R., LICHTNER, P., and THALLER, G. 2010b. The impact of genetic relationship information on genomic breeding values in German holstein cattle. *Genet Sel Evol*, 42(5).
- HAYES, B.J., and GODDARD, M.E. 2001. The distribution of the effects of genes affecting quantitative traits in livestock. *Genet Sel Evol*, 33:209-229.
- HAYES, B.J., BOWMAN, P.J., CHAMBERLAIN, A.J., and GODDARD, M.E. 2009a. Invited review: Genomic selection in dairy cattle: progress and challenges. *J Dairy Sci*, 92:433-443.
- HAYES, B.J., BOWMAN, P.J., CHAMBERLAIN, A.C., VERBYLA, K., and GODDARD, M.E. 2009b. Accuracy of genomic breeding values in multi-breed dairy cattle populations. *Genet Sel Evol*, pages 41-51.
- HEFFNER, E., SORRELLS, M., and JANNINK, J. 2009. Genomic selection for crop improvement. *Crop Sci*, 49:1-12.
- HENDERSON, C.R. 1984. Applications of linear models in animal breeding. Guelph, Ontario, Canada: Univ. Guelph.
- HESLOT, N., SORRELLS, M.E., JANNINK, J.L., and YANG, H.P. 2012. Genomic selection in plant breeding: a comparison of models. *Crop Sci*, 52:146-160.
- ILLUMINA, INC. Products <http://www.illumina.com/products/all-products.html> [Son erişim tarihi: 10.10.2016]
- JANNINK, J.L., LORENZ, A.J., and IWATA, H. 2010. Genomic selection in plant breeding: From theory to practice. *Brief Funct Genomics*, 9:166-177.

- KAPELL, D.N.R.G., SORENSEN, D., SU, G., JANSS, L.L.G., ASHWORTH, C.J., and ROEHE, R. 2012. Efficiency of genomic selection using Bayesian multi-marker models for traits selected to reflect a wide range of heritabilities and frequencies of detected quantitative traits loci in mice. *BMC Genet*, 13(42).
- KARAMAN, E., CHENG, H., FIRAT, M.Z., GARRICK, D.J. and FERNANDO, R.L. 2016. An upper bound for accuracy of prediction using GBLUP. *PLoS One*, 11(8): e0161054.
- KIZILKAYA, K., FERNANDO, R.L., and GARRICK, D.J. 2010. Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. *J Anim Sci*, 88:544-551.
- KLUG, W.S., and CUMMINGS, M.R. 2003. Concepts of genetics. Upper Saddle River, N.J., Prentice Hall.
- LEGARRA, A. 2014. Bases for genomic prediction (v0.9). University of Georgia, course note, Athens, Georgia, USA.
- LEGARRA, A., ROBERT-GRANIE, C., MANFREDI, E., and ELSSEN, J.M. 2008. Performance of genomic selection in mice. *Genetics*, 180:611-618.
- LORENZANA, R.E., and BERNARDO, R. 2009. Accuracy fo genotypic value predictions for marker-based selection in biparental plant populations. *Theor Appl Genet*, 120:151-161.
- LUAN, T., WOOLLIAMS, J.A., LIEN, S., KENT, M., SVENDSEN, M., and MEUWISSEN, T.H.E. 2009. The accuracy of genomic selection in norwegian red cattle assessed by cross-validation. *Genetics*, 183:1119-1126.
- MAKOWSKY, R., PAJEWSKI, N.M., KLIMENTIDIS, Y.C., VAZQUEZ, A.I., DUARTE, C.W., and others. 2011. Beyond missing heritability: Prediction of complex traits. *PLoS Genet*, 7(4):e1002051.
- MEUWISSEN, T.H.E. 2009. Accuracy of breeding values of unrelated individuals predicted by dense snp genotyping. *Genet Sel Evol*, 41(35).
- MEUWISSEN, T.H.E., and GODDARD, M. 2010. Accurate prediction of genetic values for complex traits by whole-genome resequencing. *Genetics*, 185:623-631.
- MEUWISSEN, T.H.E., HAYES, B.J., and GODDARD, M.E. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157:1819-1829.
- MEUWISSEN, T.H.E., HAYES, B.J., and GODDARD, M.E. 2013. Accelerating improvement of livestock with genomic selection. *Annu Rev Anim Biosci*, 1:221-237.

- MISZTAL, I., AGUILAR, I., LEGARRA, A., TSURUTA, S., JOHNSON, D.L., and LAWLOR, T.J. 2010. A unified approach to utilize phenotypic, full pedigree and genomic information for genetic evaluation. In: Proceedings of the 9th World Congress Applied to Livestock Production, Leipzig, Germany.
- NEJATI-JAVAREMI, V., SMITH, C., and GIBSON, J.P. 1997. Effect of total allelic relationship on accuracy of evaluation and response to selection. *J Anim Sci*, 75:1738-1745.
- OBER, U., AYROLES, J.F., STONE, E.A., RICHARDS, S., ZHU, D., GIBBS, R.A., STRICKER, C., GIANOLA, D., SCHLATHER, M., MACKAY, T.F.C., and SIMIANER, H. 2011. Using whole-genome sequence data to predict quantitative trait phenotypes in drosophila melanogaster. *PLoS Genet*, 8(5).
- PEREZ-CABAL, M.A., VAZQUEZ, A.I., GIANOLA, D., ROSA, G.J.M., and WEIGEL, K.A. 2012. Accuracy of genome-enabled prediction in a dairy cattle population using different cross-validation layouts. *Frontiers in Genetics*, 3(27).
- PSZCZOLA, M., STRABEL, T., MULDER, H.A., and CALUS, M.P.L. 2011. Reliability of direct genomic values for animals with different relationships within and to the reference population. *J Dairy Sci*, 95:389-400.
- RESENDE, M.F.J.R., MUNOZ, P., RESENDE, M.D.V., GARRICK, D.J., FERNANDO, R.L., et al. 2012. Accuracy of genomic selection methods in a standard dataset of loblolly pine (pinus taeda l.). *Genetics*, 190:1503-1510.
- SAATCHI, M., MIRAEI-ASHTIANI, S.R., NEJATI-JAVAREMI, A., MORADI-SHAHREBABA, M., and MEHRABANI-YEGHANEH, H. 2010. The impact of information quantity and strength of relationship between training set and validation set on accuracy of genomic estimated breeding values. *Afr J Biotechnol*, 9(4):438-442.
- SAATCHI, M., MCCLURE, M.C., MCKAY, S.D., ROLF, M.M., KIM, J.W., et al. 2011. Accuracies of genomic breeding values in american angus beef cattle using k-means clustering for cross validation. *Genet Sel Evol*, 43(40).
- SCHEFFE, H. 1989. The analysis of variance, volume 72. John Wiley & Sons, NY, USA.
- STRANDEN, I., and GARRICK, D.J. 2009. Technical note: Derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. *J Dairy Sci*, 92(6):2971-2975.
- SU, G., BRONDUM, R., MA, P., GULDBRANDTSEN, B., AAMAND, G., and LUND, M. 2012. Comparison of genomic predictions using medium-density (54,000) and high-density (777,000) single nucleotide polymorphism marker panels in nordic holstein and red dairy cattle populations. *J Dairy Sci*, 95:4657-4665.

- SVED, J. 1971. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theoretical Population Biology*, 2(2):125-141.
- TAKAHATA, N. 1993. Allelic genealogy and human evolution. *Mol Biol Evol*, 10(1):2-22.
- The 1000 Genomes Project Consortium. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491:56-65.
- VAN DEN BERG, I., BOICHARD, D., GULDBRANTSSEN, B., and LUND, M.S. 2016. Using causative variants in linkage disequilibrium with causative mutations to improve across-breed prediction in dairy cattle: a simulation study. *G3: Genes Genomes Genetics*, 6:2553-2561.
- VAN RADEN, P.M. 2008. Efficient methods to compute genomic predictions. *J Dairy Sci*, 91(11):4414-4423.
- VAN RADEN, P.M., and SULLIVAN, P.G. 2010. International genomic evaluation methods for dairy cattle. *Genet Sel Evol*, 42(7).
- VAN RADEN, P.M., VAN TASSELL, C.P., WIGGANS, G.R., SONSTEGARD, T.S., SCHNABEL, R.D., TAYLOR, J.F., and SCHENKEL, F.S. 2009. Invited review: Reliability of genomic predictions for north american holstein bulls. *J Dairy Sci*, 92:16-24.
- VAZQUEZ, A.I., ROSA, G.J.M., WEIGEL, K.A., DE LOS CAMPOS, G., GIANOLA, D., et al. 2010. Predictive ability of subsets of single nucleotide polymorphisms with and without parent average in us holsteins. *J Dairy Sci*, 93:5942-5949.
- VERBYLA, K.L., HAYES, B.J., BOWMAN, P.J., and GODDARD, M.E. 2009. Accuracy of genomic selection using stochastic search variable selection in australian holstein friesian dairy cattle. *Genet Res (Camb)*, 91:307-311.
- WIENTJES, Y.C.J., CALUS, M.P.L., GODDARD, M.E., and HAYES, B.J. 2015. Impact of qtl properties on the accuracy of multi-breed genomic prediction. *Genet Sel Evol*, 47(42).
- WOLC, A., ARANGO, J., SETTAR, P., FULTON, J.E., O'SULLIVAN, N.P., et al. 2011a. Persistence of accuracy of genomic estimated breeding values over generations in layer chickens. *Genet Sel Evol*, 43(23).
- WOLC, A., STRICKER, C., ARANGO, J., SETTAR, P., FULTON, J.E., et al. 2011b. Breeding value prediction for production traits in layer chickens using pedigree or genomic relationships in a reduced animal model. *Genet Sel Evol*, 43(5).

- WOLC, A., ARANGO, J., SETTAR, P., FULTON, J.E., O'SULLIVAN, N.P., PREISINGER, R., FERNANDO, R., GARRICK, D.J., and DEKKERS, J.C. 2013. Analysis of egg production in layer chickens using a random regression model with genomic relationships. *Poult Sci*, 92(6):1486-1491.
- WU, X., LUND, M.S., SUN, D., ZHANG, Q., and SU, G. 2015. Impact of relationships between test and training animals and among training animals on reliability o genomic prediction. *J Anim Breed Genet*, 132(5): 366-375.
- YANG, J., BENYAMIN, B., MCEVOY, B.P., GORDON, S., HENDERS, A.K., et al. 2010. Common snps explain a large proportion of the heritability for human height. *Nature Genet*, 42:565-569.
- ZENG, J., PSZCZOLA, M., WOLC, A., STRABEL, T., FERNANDO, R.L., GARRICK, D.J., and DEKKERS, J.C.M. 2012. Genomic breeding value prediction and qtl mapping of QTLMAS2011 data using Bayesian and GBLUP methods. *BMC Proc* 6(Suppl. 2): S7, 2012.
- ZHANG, Z., LIU, J., DING, X., BIJMA, P., DE KONING, D.J., AND ZHANG, Q. 2010. Best linear unbiased prediction of genomic breeding values using a trait-specific marker-derived relationship matrix. *PLoS ONE*, 5(9):e12648.
- ZHANG Z., ZHANG Q., and DING X.D. 2011. Advances in genomic selection in domestic animals. *Chinese Sci Bull*, 56:2655-2663.

8. EKLER

Ek-1

Goddard vd (2011) genomik değerdendirmerinin isabetini tahmin etmek için aşğıdaki yakınsamayı önermiştir:

$$r^2 \approx q^2 \frac{\theta}{\theta + 1} \quad (8.1)$$

Burada, $\theta = n_R q^2 h^2 / M_e$ ve q^2 ise markerlar tarafından açıklanan genetik varyansın toplam genetik varyansa oranıdır, $q^2 = \sigma_q^2 / \sigma_g^2$. Yukarıdaki eşitlikte θ 'yi yerine koyarak,

$$r^2 \approx q^2 \frac{n_R q^2 h^2}{n_R q^2 h^2 + M_e}, \quad (8.2)$$

veya ana metinde (Giriş) verildiğı haliyle,

$$r^2 \approx q^2 \frac{n_R h^2}{n_R h^2 + \frac{M_e}{q^2}}. \quad (8.3)$$

elde ederiz. $h^2 = \sigma_g^2 / \sigma_y^2$ ve $q^2 = \sigma_q^2 / \sigma_g^2$ olduğundan, $h_M^2 = \sigma_q^2 / \sigma_y^2$ genomik kalıtım derecesi olmak üzere (de los Campos vd 2013b) eşitlik (8.2) yeniden düzenlenerek,

$$r^2 \approx q^2 \frac{n_R h_M^2}{n_R h_M^2 + M_e}, \quad (8.4)$$

elde edilir.

$R_{(\hat{u}, y)}^2 = \frac{Kov(\hat{u}, y)^2}{Var(\hat{u})Var(y)}$ ve $r_{(\hat{u}, u)}^2 = \frac{Kov(\hat{u}, u)^2}{Var(\hat{u})Var(u)}$ olduğundan $R_{(\hat{u}, y)}^2 = h^2 r_{(\hat{u}, u)}^2$ şeklindedir. Bu nedenle, isabetin ölçüsü olarak $\hat{u}$ and y arasındaki korelasyonun karesi kullanıldığında, yakınsama (8.4) kalıtım derecesi, h^2 , ile çarpılmalıdır. Bu yolla R^2 için aşğıdaki yakınsamaya ulaşırız (1.1):

$$R^2 \approx h_M^2 \frac{n_R h_M^2}{n_R h_M^2 + M_e} \quad (8.5)$$

ve

$$\lim_{n_R \rightarrow \infty} h_M^2 \left(\frac{n_R h_M^2}{n_R h_M^2 + M_e} \right) = h_M^2 \quad (8.6)$$

Ek-2

Kantitatif özellik y 'nin biallelik, bağımsız ve eklemeli olarak etki eden k adet lokus tarafından determine edildiğini varsayalım. Bu lokuslar yalnızca QTL'ler ile istatistiksel olarak ilişkili markerları barındırabileceği gibi, kantitatif özelliğe etki eden genleri de barındırabilir. Bu durumda aşağıdaki modeli kullanırız:

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + \epsilon_i \quad (8.7)$$

Burada, $\mathbf{x}_i$ referans popülasyonundaki (RP) i 'inci bireyin marker genotiplerini içeren vektör ve $\boldsymbol{\beta}$ da marker etkilerini içeren vektördür. Ayrıca, $\mathbf{V}_{XX}$, RP bireylerine ilişkin merkezileştirilmiş marker genotiplerini içeren $\mathbf{X}$ kullanılarak elde edilen varyans-kovaryans matrisi olmak üzere $\mathbf{X} \sim N(\mathbf{0}, \mathbf{V}_{XX})$ olduğunu varsayalım. $\mathbf{k}_i$ test popülasyonundaki (TP) i 'inci bireyin marker genotiplerini içeren vektör olmak üzere, söz konusu bireyin eklemeli genetik değeri $u_i = \mathbf{k}_i' \boldsymbol{\beta}$, ve bunun en küçük kareler tahmini de $\hat{u}_i = \mathbf{k}_i' \hat{\boldsymbol{\beta}}$ olacaktır.

Marker etkilerinin en küçük kareler tahmini,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \quad (8.8)$$

şeklindedir. Ayrıca, $\hat{\boldsymbol{\beta}}$ 'in koşullu ortalaması ve varyansı,

$$\begin{aligned} E(\hat{\boldsymbol{\beta}} \mid \mathbf{X}) &= \boldsymbol{\beta} \\ Var(\hat{\boldsymbol{\beta}} \mid \mathbf{X}) &= (\mathbf{X}'\mathbf{X})^{-1} \sigma_\epsilon^2 \end{aligned}$$

şeklindedir.

Eklemeli genetik değer tahmininde isabetin ölçüsü olarak,

$$r_{(u_i, \hat{u}_i)}^2 = \frac{Cov(u_i, \hat{u}_i \mid \mathbf{X})^2}{Var(u_i \mid \mathbf{X}) Var(\hat{u}_i \mid \mathbf{X})} \quad (8.9)$$

ile ilgilendiğimizi varsayalım. Bu durumda gerekli büyüklükler aşağıdaki gibi elde edilir:

$$\begin{aligned} Cov(u_i, \hat{u}_i \mid \mathbf{X}) &= Cov(\mathbf{k}_i' \boldsymbol{\beta}, \mathbf{k}_i' \hat{\boldsymbol{\beta}} \mid \mathbf{X}) \\ &= E[Cov(\mathbf{k}_i' \boldsymbol{\beta}, \mathbf{k}_i' \hat{\boldsymbol{\beta}} \mid \mathbf{k}_i, \mathbf{X})] \\ &\quad + Cov[E(\mathbf{k}_i' \boldsymbol{\beta} \mid \mathbf{k}_i, \mathbf{X}), E(\mathbf{k}_i' \hat{\boldsymbol{\beta}} \mid \mathbf{k}_i, \mathbf{X})] \end{aligned}$$

$E[Cov(\mathbf{k}_i' \boldsymbol{\beta}, \mathbf{k}_i' \hat{\boldsymbol{\beta}} \mid \mathbf{k}_i, \mathbf{X})] = 0$ olacaktır çünkü $\mathbf{k}_i' \boldsymbol{\beta}$, $\mathbf{k}_i$ üzerine koşullu dağılımı düşünüldüğünde şansa bağlı değildir.

$$\begin{aligned}
Cov[E(\mathbf{k}'_i \boldsymbol{\beta} \mid \mathbf{k}_i, \mathbf{X}), E(\mathbf{k}'_i \hat{\boldsymbol{\beta}} \mid \mathbf{k}_i, \mathbf{X})] &= Cov(\mathbf{k}'_i \boldsymbol{\beta}, \mathbf{k}'_i \boldsymbol{\beta}) \\
&= Var(\mathbf{k}'_i \boldsymbol{\beta}) \\
&= Var(u) \\
&= \sigma_u^2
\end{aligned}$$

$$\begin{aligned}
Var(u_i \mid \mathbf{X}) &= Var(\mathbf{k}'_i \boldsymbol{\beta} \mid \mathbf{X}) \\
&= \boldsymbol{\beta}' \mathbf{V}_{KK} \boldsymbol{\beta} \\
&= \sigma_u^2
\end{aligned}$$

$$\begin{aligned}
Var(\hat{u}_i \mid \mathbf{X}) &= Var(\mathbf{k}'_i \hat{\boldsymbol{\beta}} \mid \mathbf{k}_i, \mathbf{X}) \\
&= E[Var(\mathbf{k}'_i \hat{\boldsymbol{\beta}} \mid \mathbf{k}_i, \mathbf{X})] + Var[E(\mathbf{k}'_i \hat{\boldsymbol{\beta}} \mid \mathbf{k}_i, \mathbf{X})] \\
&= E[\mathbf{k}'_i Var(\hat{\boldsymbol{\beta}} \mid \mathbf{X}) \mathbf{k}_i] + Var(\mathbf{k}'_i \boldsymbol{\beta}) \\
&= E\{\mathbf{k}'_i [(\mathbf{X}'\mathbf{X})^{-1} \sigma_\epsilon^2] \mathbf{k}_i\} + \sigma_u^2
\end{aligned}$$

$$\begin{aligned}
E\{\mathbf{k}'_i [(\mathbf{X}'\mathbf{X})^{-1} \sigma_\epsilon^2] \mathbf{k}_i\} + \sigma_u^2 &= E\{tr(\mathbf{X}'\mathbf{X})^{-1} \mathbf{k}'_i \mathbf{k}_i\} \sigma_\epsilon^2 + \sigma_u^2 \\
&= tr\{E(\mathbf{X}'\mathbf{X})^{-1} \sigma_\epsilon^2\} \mathbf{V}_{KK} + \sigma_u^2
\end{aligned}$$

Lokusların ilişkisiz olduğunu varsaydığımızdan ve $\mathbf{X} \sim N(\mathbf{0}, \mathbf{V}_{XX})$ 'dan, ters Wishart dağılımının özelliğini kullanarak

$$E(\mathbf{X}'\mathbf{X})^{-1} = \frac{\mathbf{V}_{XX}^{-1}}{n_R - k - 1} \quad (8.10)$$

yazarız. Ayrıca, RP ve TP aynı popülasyondan geldiğinden, $\mathbf{V}_{XX} = \mathbf{V}_{KK}$ ve

$$\begin{aligned}
tr\{E(\mathbf{X}'\mathbf{X})^{-1} \mathbf{V}_{KK}\} \sigma_\epsilon^2 &= tr\left(\frac{\mathbf{V}_{XX}^{-1}}{n_R - k - 1} \mathbf{V}_{KK}\right) \sigma_\epsilon^2 \\
&= \frac{k}{n_R - k - 1} \sigma_\epsilon^2
\end{aligned}$$

Bu nedenle,

$$Var(\hat{u}_i \mid \mathbf{X}) = \frac{k}{n_R - k - 1} \sigma_\epsilon^2 + \sigma_u^2 \quad (8.11)$$

olacaktır. Elde ettiğimiz bu değerleri (8.9)'de yerine koyarsak,

$$\begin{aligned}
r_{(u_i, \hat{u}_i)}^2 &= \frac{(\sigma_u^2)^2}{\sigma_u^2 \left(\frac{k}{n_R - k - 1} \sigma_\epsilon^2 + \sigma_u^2 \right)} \\
&= \frac{\sigma_u^2}{\frac{k}{n_R - k - 1} \sigma_\epsilon^2 + \sigma_u^2}
\end{aligned}$$

sonucuna ulaşırız. Cebirsel bir takım işlemler ve hem payı hem de paydayı $\frac{1}{\sigma_y^2}$ ile çarparak

$$r_{(u_i, \hat{u}_i)}^2 = \frac{(n_R - k - 1)h^2}{k(1 - h^2) + (n_R - k - 1)h^2} \quad (8.12)$$

elde ederiz. Daetwyler vd (2008), çalışmalarında $\sigma_\epsilon^2 = \sigma_y^2 = 1$ olduğunu varsaymışlar ve, $\mathbf{X} \sim N(\mathbf{0}, \mathbf{V}_{XX})$ varsayımında bulunmamışlardır. Bu durumda,

$$r_{(u_i, \hat{u}_i)}^2 = \frac{n_R h^2}{k + n_R h^2} \quad (8.13)$$

olacaktır.

Ek-3

Bu kısımda, RP bireylerinin genotip matrisi ve TP bireylerinin genotip vektörlerini göstermek için Materyal ve Yöntem'in ilgili bölümünde kullanılan notasyon izlenecek ve $\tilde{\boldsymbol{\beta}} \in \Re(\mathbf{X}_R)$ olduğu gösterilecektir.

Aşağıdaki gibi verilen sabit etkiler modeli kullanılıyor olsun:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}. \quad (8.14)$$

Burada $\mathbf{y}$ merkezleştirilmiş gözlemleri içeren $n \times 1$ boyutlu vektör, $\mathbf{X}$ merkezleştirilmiş genotipleri içeren $n \times p$ boyutlu matris, $\boldsymbol{\beta}$ ortalaması sıfır ve varyans-kovaryans matrisi $\mathbf{I}\sigma_\beta^2$ olan $p \times 1$ boyutlu şansa bağlı marker etkilerini içeren vektör, ve $\mathbf{e}$ de $n \times 1$ boyutlu ortalaması sıfır ve varyans-kovaryans matrisi $\mathbf{I}\sigma_e^2$ olan şansa bağlı çevre etkilerini içeren vektördür. $\boldsymbol{\beta}$ 'nin en iyi doğrusal yansız tahmincisi aşağıdaki eşitliğin çözümünü sağlar:

$$(\mathbf{X}_R' \mathbf{X}_R + \mathbf{I}\lambda) \tilde{\boldsymbol{\beta}} = \mathbf{X}_R' \mathbf{y}_R. \quad (8.15)$$

Burada $\lambda = \frac{\sigma_e^2}{\sigma_\beta^2}$ 'dir. $\tilde{\boldsymbol{\beta}}$, $\mathbf{b}_1 \in \Re(\mathbf{X}_R)$ ve $\mathbf{b}_2 \perp \Re(\mathbf{X}_R)$ olmak üzere,

$$\tilde{\boldsymbol{\beta}} = \mathbf{b}_1 + \mathbf{b}_2, \quad (8.16)$$

gibi iki vektörün toplamı şeklinde yazılabilir (Scheffe 1989). Bu durumda, (8.16) kullanılarak, eşitlik (8.15)

$$(\mathbf{X}'_R \mathbf{X}_R \mathbf{b}_1 + \mathbf{I} \lambda \mathbf{b}_1) + (\mathbf{X}'_R \mathbf{X}_R \mathbf{b}_2 + \mathbf{I} \lambda \mathbf{b}_2) = \mathbf{X}'_R \mathbf{y}_R \quad (8.17)$$

şeklinde yazılabilir. $\mathbf{b}_1 \in \mathfrak{R}(\mathbf{X}_R)$ olduğundan $(\mathbf{X}'_R \mathbf{X}_R \mathbf{b}_1 + \mathbf{I} \lambda \mathbf{b}_1) \in \mathfrak{R}(\mathbf{X}_R)$ ve $\mathbf{b}_2 \perp \mathfrak{R}(\mathbf{X}_R)$ ve $\mathbf{X}'_R \mathbf{X}_R \in \mathfrak{R}(\mathbf{X}_R)$ olduğundan $\mathbf{X}'_R \mathbf{X}_R \mathbf{b}_2 = 0$ olacaktır. Bunu kullanarak, eşitlik (8.17):

$$\mathbf{I} \lambda \mathbf{b}_2 = \mathbf{X}'_R \mathbf{y}_R - \mathbf{X}'_R \mathbf{X}_R \mathbf{b}_1 - \mathbf{I} \lambda \mathbf{b}_1 \quad (8.18)$$

şeklinde yeniden yazılabilir. $\mathbf{X}'_R \mathbf{y}_R \in \mathfrak{R}(\mathbf{X}_R)$, $\mathbf{X}'_R \mathbf{X}_R \in \mathfrak{R}(\mathbf{X}_R)$ ve $\mathbf{b}_1 \in \mathfrak{R}(\mathbf{X}_R)$ olduğundan eşitlik (8.18)'un sağ tarafı da $\mathfrak{R}(\mathbf{X}_R)$ 'in bir elemanıdır. Yine bu eşitliğin (8.18) sol tarafı oluşturulma şekli gereği $\mathbf{X}_R$ 'in satır uzayına ortogonaldır. Bu nedenle, hem $\mathbf{b}_2 \in \mathfrak{R}(\mathbf{X}_R)$ hem de $\mathbf{b}_2 \perp \mathfrak{R}(\mathbf{X}_R)$ olacaktır. Bu durumda $\mathbf{b}_2$ kendisine ortogonal olacaktır, ki bu da yalnızca $\mathbf{b}_2 = \mathbf{0}$ olması durumunda olanaklıdır. Sonuç olarak $\tilde{\boldsymbol{\beta}} = \mathbf{b}_1$ 'dir ve $\mathfrak{R}(\mathbf{X}_R)$ 'in elemanıdır.

Ek-4

Bu kısımda $\hat{u}_i$ 'in isabeti,

$$r^2_{(u_i, \hat{u}_i)} = \frac{Var(\hat{u}_i)}{Var(u_i)}, \quad (8.19)$$

için bir üst sınır, (UP_i) ortaya koyacağız. Daha önce yazdığımız $\hat{u}_i = \mathbf{x}'_{T_1} \tilde{\boldsymbol{\beta}}$ eşitliğini kullanarak $\hat{u}$ için tahmin hatasının varyansı

$$\begin{aligned} Var(u_i - \hat{u}_i) &= Var(u_i) - Var(\hat{u}_i) \\ &= \mathbf{x}'_{T_1} (\mathbf{X}'_R \mathbf{X}_R + \mathbf{I} \lambda)^{-1} \mathbf{x}_{T_1} \sigma_e^2 \end{aligned}$$

şeklinde yazılabilir (Henderson 1984). Bu durumda,

$$\begin{aligned} Var(\hat{u}_i) &= Var(\mathbf{x}'_{T_1} \tilde{\boldsymbol{\beta}}) \\ &= Var(\mathbf{x}'_{T_1} \boldsymbol{\beta}) - \mathbf{x}'_{T_1} (\mathbf{X}'_R \mathbf{X}_R + \mathbf{I} \lambda)^{-1} \mathbf{x}_{T_1} \sigma_e^2 \end{aligned}$$

olacaktır. Buradan, $h^2 \rightarrow 1$, $\sigma_e^2 \rightarrow 0$ iken $Var(\mathbf{x}'_{T_1} \tilde{\boldsymbol{\beta}}) \rightarrow Var(\mathbf{x}'_{T_1} \boldsymbol{\beta})$ olduğu görülmektedir. $Var(\boldsymbol{\beta}) = \mathbf{I} \sigma_\beta^2$ olduğunu varsayımıyla, $Var(\mathbf{x}'_{T_1} \boldsymbol{\beta}) = \mathbf{x}'_{T_1} \mathbf{x}_{T_1} \sigma_\beta^2$, ve

$$UP_i = \frac{\mathbf{x}'_{T_1} \mathbf{x}_{T_1}}{\mathbf{x}'_T \mathbf{x}_T} \quad (8.20)$$

ile $r_{(u_i, \hat{u}_i)}^2$ için ve h^2UP_i ile de $R_{(y_i, \hat{u}_i)}^2$ için bir üst sınır elde edildiği görülecektir. $h^2 = 1$ olduğunda $\lambda = 0$ olacak ve $\tilde{\beta}$ da bu nedenle en küçük kareler tahmincisi β olacaktır. Bu durumda $\tilde{\beta} \notin \mathcal{R}(\mathbf{X}_R)$ 'in bir elemanı olmayacaktır.



ÖZGEÇMİŞ

1982 yılında Kocaeli’de doğdu. İlk öğrenimini Değirmendere Donanma İlköğretim okulunda, orta öğrenimini Gölcük İhsaniye Anadolu Lisesi’nde ve lise öğrenimini Antalya Adem Tolunay Anadolu Lisesi’nde tamamladı. Marmara Üniversitesi Ekonometri bölümünden 2008 yılında mezun oldu. Aynı yıl Akdeniz Üniversitesi Ziraat Fakültesi Zootekni Bölümü Biyometri ve Genetik Anabilim Dalı’nda yüksek lisansa başladı. Halen sözü edilen bölümde doktora eğitimine devam etmektedir.