



# UNIVERSIDAD AUTÓNOMA CHAPINGO

DEPARTAMENTO DE ENSEÑANZA, INVESTIGACIÓN  
Y SERVICIO EN ZOOTECNIA  
POSGRADO EN PRODUCCIÓN ANIMAL

## COMPARACIÓN DE VALORES GENÉTICOS UTILIZANDO MARCADORES SNP Y MODELOS UNI- Y BIVARIADOS TESIS

Que como requisito parcial  
para obtener el grado de:

### MAESTRO EN CIENCIAS EN INNOVACIÓN GANADERA

Presenta:

**JONATHAN EMANUEL VALERIO HERNÁNDEZ**

Bajo la supervisión de:  
Agustín Ruiz Flores, Ph. D.  
Paulino Pérez Rodríguez, Dr.



Enero, 2018  
Chapingo, México



DIRECCIÓN GENERAL ACADÉMICA  
DEPTO. DE SERVICIOS ESCOLARES  
OFICINA DE EXÁMENES PROFESIONALES

## COMPARACIÓN DE VALORES GENÉTICOS UTILIZANDO MARCADORES SNP Y MODELOS UNI- Y BIVARIADOS

Tesis realizada por **JONATHAN EMANUEL VALERIO HERNÁNDEZ** bajo la supervisión del comité asesor indicado, aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

### MAESTRO EN CIENCIAS EN INNOVACIÓN GANADERA

**DIRECTOR:**



Ph. D. AGUSTÍN RUÍZ FLORES

**CO-DIRECTOR:**



Dr. PAULINO PÉREZ RODRÍGUEZ

**ASESOR:**



Ph. D. RAFAEL NÚÑEZ DOMÍNGUEZ

**ASESOR:**



Ph. D. RODOLFO RAMÍREZ VALVERDE

## CONTENIDO

CONTENIDO.....	iii
LISTA DE CUADROS .....	vi
LISTA DE FIGURAS .....	vii
LISTA DE APÉNDICES.....	viii
ABREVIATURAS USADAS.....	ix
DEDICATORIAS .....	x
AGRADECIMIENTOS .....	xi
DATOS BIOGRÁFICOS .....	xii
1. INTRODUCCIÓN GENERAL.....	1
2. REVISIÓN DE LITERATURA .....	4
2.1. Marcadores genéticos .....	4
2.1.1. Marcadores morfológicos .....	4
2.1.2. Marcadores bioquímicos .....	5
2.1.3. Marcadores moleculares .....	6
2.1.4. Marcadores SNPs .....	7
2.2. Selección genómica.....	8
2.2.1. Implementación de un programa de mejoramiento genómico.....	10
2.2.2. Ventajas de la selección genómica .....	11
2.2.3. Desventajas de la selección genómica .....	12

2.3.	Algunas metodologías estadísticas para evaluación genómica.....	14
2.3.1.	Regresión Ridge .....	15
2.3.2.	Regresión LASSO .....	16
2.3.3.	Regresión Elastic Net.....	17
2.3.4.	Regresión Kernel (RKHS) .....	18
2.3.5.	GBLUP .....	19
2.4.	Complementación de metodologías cuantitativas y genómicas .....	20
2.5.	Análisis uni- vs bivariados .....	21
2.6.	Perspectivas de selección genómica.....	22
2.7.	Literatura Citada .....	23
3.	ANÁLISIS UNI- VS MULTIVARIADOS EN LA PREDICCIÓN DE VALORES GENÓMICOS DE CARACTERÍSTICAS DE CRECIMIENTO EN GANADO SUIZO EUROPEO.....	31
3.1.	Resumen .....	31
3.2.	Abstract .....	32
3.3.	Introducción .....	33
3.4.	Materiales y métodos.....	34
3.4.1.	Origen de la información .....	34

3.4.1.	Edición de información genotípica .....	35
3.4.2.	Edición de información fenotípica .....	35
3.4.3.	Análisis estadístico.....	36
3.4.4.	Criterios de comparación de los análisis .....	39
3.5.	Resultados y discusión .....	40
3.5.1.	Jerarquización de los valores genómicos predichos .....	41
3.5.2.	Habilidad de predicción de los modelos .....	44
3.5.3.	Estimadores de los coeficientes de correlación Pearson .....	46
3.6.	Conclusiones .....	48
3.7.	Literatura citada .....	48
3.8	Apéndices .....	52

## LISTA DE CUADROS

Cuadro 1 Estadísticos descriptivos y número de observaciones (n) utilizados en los análisis genómicos para los pesos al nacimiento (PN), al destete (PD) y al año (PA) de bovinos Suizo Europeo. ....	39
Cuadro 2. Estimadores de componentes de varianza genética aditiva ( $VG_A$ ), varianza genética aditiva materna ( $VG_M$ ), varianza fenotípica ( $VF$ ), heredabilidad directa ( $h^2_D$ ) y materna ( $h^2_M$ ) obtenidos mediante análisis univariados para los pesos al nacimiento (PN), al destete (PD) y al año (PA) en bovinos Suizo Europeo.....	41
Cuadro 3. Estimadores de los coeficientes de correlación Pearson y Spearman, en paréntesis nivel de significancia para la prueba de hipótesis del estimador igual a cero, entre los valores genómicos predichos con análisis uni- y bivariados para pesos al nacimiento, al destete y al año en bovinos Suizo Europeo. ....	43
Cuadro 4. Estimadores de los coeficientes de regresión ( $\pm$ error estándar), utilizando como variable independiente los valores genómicos predichos mediante análisis univariados y como variable dependiente los valores genómicos predichos con los análisis bivariados y viceversa; en paréntesis nivel de significancia para la prueba de hipótesis del estimador igual a cero, para pesos al nacimiento, al destete directo y al año en bovinos Suizo Europeo. ....	47

## LISTA DE FIGURAS

Figura 1. Promedios de los estimadores de correlación Pearson entre los valores genómicos predichos mediante análisis univariados (Uni) con los valores genómicos predichos promedio (PromBiv) de análisis bivariados (BPNPD, BPNPA y BPDPA) para los pesos al nacimiento (PN), al destete (PD) y al año (PA), de cinco repeticiones de validación cruzada. ....	44
--	----

## LISTA DE APÉNDICES

<b>Apéndice 1. Código en R para el ajuste de PN .....</b>	<b>52</b>
<b>Apéndice 2. Código en R para el ajuste de PD .....</b>	<b>53</b>
<b>Apéndice 3. Código en R para el ajuste de PA .....</b>	<b>54</b>
<b>Apéndice 4. Código en R para análisis univariado de PN y su validación cruzada.....</b>	<b>55</b>
<b>Apéndice 5. Código en R para análisis univariado de PD y su validación cruzada.....</b>	<b>57</b>
<b>Apéndice 6. Código en R para análisis univariado de PA y su validación cruzada.....</b>	<b>59</b>
<b>Apéndice 7. Código en R para análisis bivariado de PN y PD y su validación cruzada.....</b>	<b>61</b>
<b>Apéndice 8. Código en R para análisis bivariado de PN y PA y su validación cruzada.....</b>	<b>66</b>
<b>Apéndice 9. Código en R para análisis bivariado de PD y PA y su validación cruzada.....</b>	<b>71</b>



## ABREVIATURAS USADAS

ADN	Ácido desoxirribonucleico
AFLPs	Polimorfismos en la longitud de fragmentos amplificados (Amplified fragment length polymorphism)
BLUE	Mejor estimador lineal insesgado (Best Linear Unbiased Estimator)
BLUP	Mejor predictor lineal insesgado (Best Linear Unbiased Predictor)
EBVs	Valor genético predicho (Estimated Breeding Value)
GBLUP	Mejor predictor lineal insesgado genómico (Genomic Best Linear Unbiased Predictor)
GWAS	Estudio de asociación de genoma completo (Genome-Wide Association Study)
LARS	Regresión de ángulo mínimo (Least angle regression)
LASSO	Operador de contracción y selección mínimo absoluto (Least Absolute Shrinkage and Selection Operator)
LD	Desequilibrio de ligamiento (Linkage Disequilibrium)
PA (YW)	Peso al año (yearling weight)
PCR	Reacción en cadena de la polimerasa (Polymerase Chain Reaction)
PD (WW)	Peso al destete (weaning weight)
PN (BW)	Peso al nacimiento (birth weight)
QTL	Loci de características cuantitativas (Quantitative Trait Loci)
RAPDs	Polimorfismos amplificados aleatoriamente (Random Amplified Polymorphic)
REN	Regresión Elastic-Net
RFLPs	Polimorfismos de longitud de fragmentos de restricción (Restriction fragment length polymorphism)
RKHS	Regresión Kernel
RR	Regresión Ridge
SCE	Suma de cuadrados del error
SG	Selección genómica
SNPs	Polimorfismos de nucleótido simple (Single Nucleotide Polymorphism)
STS	Sitios marcados por secuencias (Sequence tagged sites)
VGn	Valores genómicos predichos
VNTR	Repeticiones en tándem de número variable (Variable Number Tandem Repeat)

## **DEDICATORIAS**

A Dios primeramente por permitirme durante estos dos años, adicionales a los primeros siete en Chapingo, entender cómo funciona un poco de su gran creación.

A mis padres por los consejos, motivaciones y apoyo durante mi estancia en Chapingo.

A la familia Tapia Barrera por abrirme la puerta de su casa, por los consejos, por el apoyo, atenciones para conmigo y considerarme parte de su familia.

A todas las personas que hicieron posible esto pero que no mencioné porque la lista sería muy grande.

**¡GRACIAS A TODOS!**

## **AGRADECIMIENTOS**

Al Consejo Nacional de Ciencia y Tecnología por el otorgamiento de la beca para realizar mis estudios de maestría.

A la Universidad Autónoma Chapingo por recibirme y permitirme formar parte de su alma mater.

Al Dr. Agustín Ruíz Flores por su supervisión, consejos y motivaciones durante la maestría.

Al Dr. Paulino Pérez Rodríguez por su supervisión, consejos y orientación de los análisis estadísticos del presente trabajo.

Al Dr. Rafael Núñez Domínguez y al Dr. Rodolfo Ramírez Valverde por las acertadas observaciones y recomendaciones.

Al IAZ. Susana Morales Pablo y al MC. Jorge Ángel Hidalgo Moreno por la amistad, apoyo y recomendaciones durante la realización del presente trabajo.

A todos mis profesores por ayudar en mi formación profesional, ética y moral.

## DATOS BIOGRÁFICOS



### Datos personales

Nombre Jonathan Emanuel Valerio Hernández.  
Fecha de nacimiento 24 de julio de 1993.  
Lugar de nacimiento Sontecomapan, Catemaco, Veracruz.  
CURP VAHJ930724HVZLRN07.  
Profesión Ingeniero Agrónomo Especialista en Zootecnia.  
Cédula profesional 099880.

### Desarrollo académico

2017 Maestro en Ciencias en Innovación Ganadera Universidad Autónoma Chapingo  
2015 Ingeniero Agrónomo Especialista en Zootecnia Universidad Autónoma Chapingo

## 1. INTRODUCCIÓN GENERAL

En programas de mejoramiento genético, la selección de animales con mérito genético sobresaliente para características de interés económico se espera impacte positivamente los sistemas de producción. Varias metodologías se han utilizado para predecir el valor genético de los animales, procurando lograr mayor precisión y respuesta a la selección. El avance en las tecnologías moleculares ha permitido la predicción de valores genómicos para el mejoramiento genético de plantas (Crossa et al., 2010; Jannink, Lorenz & Iwata, 2010) y animales (de los Campos, Hickey, Pong-Wong, Daetwyler & Calus, 2013; Hayes, Bowman, Chamberlain, Verbyla & Goddard, 2009; VanRaden et al., 2009). Paulatinamente, estas tecnologías moleculares han ampliado su utilización, particularmente por la reducción en el costo de la genotipificación en años recientes. Esto ha posibilitado obtener paneles densos de marcadores, que posteriormente se utilizan para predecir los valores genéticos en poblaciones de animales (Meuwissen, Hayes & Goddard, 2001).

Para predecir valores genéticos con exactitud, utilizando polimorfismos de un solo nucleótido (SNPs), se requieren estudios en los que un gran número de ellos cubran el genoma completo (de los Campos et al., 2013). Esto generalmente requiere disponer de individuos con información fenotípica confiable y un número considerable de SNPs; de no ser así, es complicado obtener ecuaciones apropiadas que relacionen SNPs con información fenotípica (Jannink et al., 2010; Meuwissen et al., 2001).

En su momento, cuando Meuwissen et al. (2001) propusieron la selección genómica, también decidieron que los efectos estimados de los marcadores fueran imparciales; es decir, evitaron la selección de marcadores en el desarrollo del modelo de predicción. Una consecuencia de esa decisión fue que se necesitan estimar más efectos predictores,  $p$ , que el número,  $n$ , de observaciones usualmente

disponibles. El problema de disponer de más SNPs que observaciones, se le refiere en la literatura como “la maldición de la dimensionalidad” (Bellman, 1969); o el problema llamado “grandes  $p$ , pequeños  $n$ ” (Jannink et al., 2010). La regresión lineal múltiple estándar no puede utilizarse sin selección de variables, lo que entra en conflicto con el objetivo original de evitar la selección de marcadores (Jannink et al., 2010).

Para superar el problema de la dimensionalidad se han propuesto una variedad de métodos; por ejemplo, la mejor predicción lineal no sesgada (BLUP), regresión ridge, regresión bayesiana, regresión kernel y métodos de aprendizaje automático (Jannink et al., 2010). Todos estos métodos fueron propuestos con el objetivo de desarrollar modelos de predicción en selección genómica. Sin embargo, aún hay necesidad de probar la predicción de valores genómicos utilizando algunos métodos desarrollados para genética cuantitativa, como modelos multivariados en los cuales se analizan simultáneamente más de una característica.

Ramírez-Valverde, Hernández-Alvarez, Núñez-Domínguez, Ruíz-Flores y García-Muñíz (2007) compararon análisis univariados contra multivariados de variables de crecimiento en evaluación genética de ganado Angus y Tropicarne, y concluyeron que para algunas características los análisis bivariados mejoraron los estimadores de la heredabilidad directa y materna, la exactitud de las predicciones de los valores genéticos y la habilidad predictiva de los modelos. Por su parte, Serão et al. (2013) realizaron un estudio considerando información genómica para el análisis uni- y bivariado de ganancia diaria de peso y consumo de materia seca, ambas componentes de eficiencia alimenticia. Los resultados de este estudio demuestran la complementariedad de los análisis uni- y bivariados en estudios GWAS. Estos autores argumentan que al considerar la covarianza entre los componentes se gana precisión estadística.

El análisis genómico multivariado es complejo de realizar y computacionalmente demandante debido a la gran cantidad de SNPs del genotipado, pero probablemente esto se justifique si se mejora la predicción de valores genómicos.

Por tal motivo el objetivo del presente estudio fue comparar análisis univariados y multivariados en la evaluación genómica de características de crecimiento en bovinos Suizo Europeo.

En el capítulo dos de esta Tesis, se revisa la literatura acerca de generalidades de selección genómica, análisis uni-y bivariados, marcadores moleculares, metodologías para evaluación genómica y las perspectivas de selección genómica.

Finalmente, en el capítulo tres, se presenta un estudio donde se comparan análisis uni- contra bi-variados utilizando marcadores SNPs para la predicción de valores genómicos de pesos al nacimiento, al destete y al año en bovinos Suizo Europeo.

## **2. REVISIÓN DE LITERATURA**

### **2.1. Marcadores genéticos**

Los marcadores genéticos son puntos de referencia o anclaje de fácil ubicación que ayudan en el estudio del material genético. Martínez, Helguera y Carrera (2010) mencionaron que para que un marcador sea un buen punto de señalamiento, es necesario que tenga una ubicación física en el genoma, ser polimórfico (dentro y entre especies), de herencia mendeliana (preferentemente no epistática), insensible a efectos ambientales, codominante (individuos heterocigotos capaces de diferenciarse de los homocigotos), de rápida identificación y simple análisis.

Existen gran variedad de marcadores, dependiendo del nivel biológico, que pueden ser desde cambios fenotípicos heredables fácilmente distinguibles, hasta la variación de un solo nucleótido de ADN. A continuación, se describen diferentes tipos de marcadores con más detalle.

#### **2.1.1. Marcadores morfológicos**

Los marcadores morfológicos son características fenotípicas de fácil identificación visual, de herencia monogénica y predecibles con las leyes de Mendel. Ejemplos de algunos marcadores morfológicos son color de fruto, forma de fruto, pubescencia, presencia de cuernos, color de pelaje, tamaño y altura, entre otros (Mártinez et al., 2010) .

Los marcadores morfológicos fueron de los primeros en utilizarse para explorar el material genético, debido a que durante mucho tiempo el trabajo genético dependió de encontrar variaciones de características fenotípicas. Muchos biólogos, incluyendo Charles Darwin, hicieron cruzamientos tratando de entender el comportamiento del material genético, pero a quien se atribuye el mayor mérito es a Gregor Mendel, quien descubrió como se segregaba la variación genética de una



cruza en casos de herencia de un solo carácter, determinado por un solo gen con dos formas alélicas (Eguiarte, Aguirre-Liguori, Jardón-Barbolla, Aguirre-Planter & Souza, 2013).

Los marcadores morfológicos tienen algunas limitantes; por ejemplo, algunas características que son influenciadas por el ambiente resultan en mediciones subjetivas que enmascaran el verdadero genotipo del individuo; pueden ser dominantes o recesivas por lo que no se podría distinguir un individuo portador de un gen recesivo; y existen características donde intervienen varios genes (epistasis), lo cual complica la observación del carácter fenotípico. También hay muchas características difíciles o caras de medir como las de la canal, resistencia a enfermedades, productividad de por vida y fertilidad, entre otras. Además, se requieren hacer cruza y observar la progenie, lo que puede implicar mucho tiempo cuando se trata de animales con intervalos generacionales largos. No obstante, los marcadores morfológicos son baratos, fáciles de usar y no requieren de equipo sofisticado, sólo se necesita determinar la característica fenotípica a observar; son útiles pues muchos de ellos se convierten en importantes descriptores que ayudan en la clasificación de nuevas variedades en plantas (Mártinez et al., 2010).

### **2.1.2. Marcadores bioquímicos**

Los marcadores bioquímicos son variaciones en ciertas proteínas codificadas por un gen, que pueden ser detectados por técnicas bioquímicas. Los marcadores bioquímicos más comunes son las isoenzimas y las proteínas de reserva. Las mutaciones en el ADN pueden ocasionar cambios en la secuencia de aminoácidos durante la síntesis de una proteína. Estos polimorfismos pueden detectarse mediante técnicas bioquímicas como la electroforesis, ya que los aminoácidos tienen diferentes pesos moleculares (Lewontin & Hubby, 1966). Los marcadores bioquímicos surgen en 1966 como una idea de Richard Lewontin, ya que los marcadores morfológicos, que eran los únicos con los que se contaba en aquel entonces, tenían muchas limitaciones. Lewontin sugirió aplicar una sencilla técnica

molecular para estudiar la variación en poblaciones naturales, llamada electroforesis de isoenzimas o alozimas, que básicamente consiste en separar proteínas según su carga y su tamaño usando un campo eléctrico. Estas proteínas son codificadas por genes de herencia mendeliana y pueden analizarse usando los fundamentos teóricos básicos de genética de poblaciones (Eguiarte et al., 2013).

Algunas desventajas de los marcadores bioquímicos se deben a limitantes de la técnica de electroforesis (Seeb et al., 2011). Eguiarte et al. (2013) explican estas limitantes: mediante electroforesis se pueden identificar variaciones del ADN que se reflejan como diferencias en la movilidad de las bandas en los geles, pero no todos los cambios en el ADN se reflejan como diferencias en dicha movilidad, ya que puede cambiar uno o varios aminoácidos sin que se modifique la movilidad de la proteína. Además, puede haber muchos cambios en el ADN que no necesariamente se reflejan como cambios en los aminoácidos, dado que existen diferentes codones que codifican para un mismo aminoácido. Weaver (2012) menciona que no más del 3% del ADN genómico es finalmente traducido a proteínas y que sólo unas pocas de estas pueden mutar sin alterar la viabilidad de los animales. Lo anterior es otra limitante porque no se podría estudiar el 97% del ADN restante.

### **2.1.3. Marcadores moleculares**

Los marcadores moleculares son polimorfismos a nivel de ADN, segmentos o nucleótidos simples, detectables bajo ciertas técnicas de biología molecular, que sirven como puntos de referencia para ilustrar la herencia de segmentos de cromosomas a través de las generaciones. Martínez et al. (2010) mencionan algunos marcadores moleculares y su clasificación dependiendo de las metodologías utilizadas para su detección: marcadores basados en hibridación de ADN; polimorfismos de longitud de los fragmentos de restricción (RFLP); repeticiones en tándem de número variable (VNTR); los marcadores basados en la amplificación arbitraria o semi arbitraria del ADN son polimorfismos amplificados aleatoriamente (RAPDs) y polimorfismos en la longitud de fragmentos amplificados

(AFLPs). Los marcadores moleculares basados en la amplificación sitio-específica del ADN son microsatélites y sitios marcados por secuencias (STS).

Con el desarrollo de las técnicas de biología molecular, surgieron diversos métodos de detección de polimorfismos genéticos directamente a nivel de ADN; inicialmente por el descubrimiento de las enzimas de restricción. Posteriormente, con el desarrollo de la reacción en cadena de la polimerasa (PCR), se pudo analizar polimorfismos de longitud de los fragmentos de restricción y desarrollar nuevos métodos de detección de marcadores moleculares (Grover & Sharma, 2014).

Entre las ventajas más destacadas de los marcadores moleculares se encuentran: alta heredabilidad, gran capacidad de detectar polimorfismos, posibilidad de acceder a todas las regiones del genoma, independencia del estado físico y de desarrollo del individuo, independencia de las condiciones ambientales y posibilidad de determinación en cualquier tipo de células que contenga núcleo. Sin embargo, algunos de ellos suelen ser caros, difíciles de implementar, de difícil lectura, o requieren de equipo sofisticado para su estudio. Por estos motivos, algunos tipos de marcadores se prefieren más que otros y se eligen dependiendo del tipo de estudio a realizar (Martínez et al., 2010).

#### **2.1.4. Marcadores SNPs**

Los polimorfismos de nucleótido simple (SNPs) pertenecen a la más reciente generación de marcadores moleculares. Estos se basan principalmente en la detección de polimorfismos provenientes del cambio de una sola base en una secuencia de ADN y representan solamente dos alelos simples (Martínez et al., 2010).

La mayoría de los SNPs se localizan en regiones no codificantes y no influyen directamente en el fenotipo de un individuo. No obstante, algunos introducen mutaciones en secuencias expresadas o en regiones que influyen en la expresión génica (promotores, potenciadores), y pueden inducir cambios en la estructura o

regulación de las proteínas (Martínez et al., 2010). Dichos SNPs tienen el potencial de detectar la variación genética funcional.

Los SNPs incluyen la técnica clásica de RFLPs, pero no exactamente detectando la aparición o abolición de un sitio de restricción. El que puedan detectarse en “arreglos” (chips) de fase sólida sin necesidad del uso de electroforesis en geles constituye una ventaja (Martínez et al., 2010). Polimorfismos específicos de diferencias en un solo nucleótido (sustitución, deleción o inserción) ocurren con frecuencia en el genoma.

## **2.2. Selección genómica**

La selección genómica (SG) es una variante de la selección asistida por marcadores moleculares, en la cual se utilizan como marcadores a los SNP distribuidos a lo largo de todo el genoma, por lo que se espera que todos los loci de características cuantitativas (QTL) queden en desequilibrio de ligamiento con al menos un marcador (Goddard & Hayes, 2007).

Hasta hace algunos años la SG consistía en rastrear genes a través de marcadores, dado que la mayoría de las características económicas de interés están influenciadas por muchos genes y muchas veces los marcadores estaban en equilibrio de ligamiento; la varianza genética explicada a través de estos métodos era pequeña (Goddard & Hayes, 2007; Jannink et al., 2010). Los primeros en proponer el término SG fueron Meuwissen et al. (2001), quienes sugirieron una variante de la selección asistida por marcadores en la que gran cantidad de SNPs se distribuyen por todo el genoma, asegurando de este modo que al menos un QTL quede en desequilibrio de ligamiento con algún marcador, para después usar el efecto de todos los marcadores y poder explicar la varianza genética.

La SG ha revolucionado el mejoramiento genético. Esto ha sido más notorio para características de baja heredabilidad, expresadas tarde en la vida de los animales

o costosas de medir. La SG asume que todos los marcadores pueden estar ligados a un gen que influye una característica y se concentra en la estimación de sus efectos más que en probar su significancia (Meuwissen, Hayes & Goddard, 2016). La SG provee predictores más precisos de los valores genéticos a edades tempranas en la vida de los animales. Dos de las ventajas consecuentes de esto son: a) incremento de la intensidad de selección, b) reducción del intervalo entre generaciones. Los incrementos en las tasas de mejoramiento genético varían de 10 a 100% dependiendo de la especie (Ibanez-Escriche & Simianer, 2015).

Los factores que permitieron que SG fuera factible (Meuwissen et al., 2001), fueron: a) un contexto metodológico adecuado, b) un mapa denso de marcadores SNP, y c) una tecnología de alto rendimiento para genotipar plataformas con un alto número de SNPs a costo razonable.

La SG en bovinos para carne no ha sido adoptada tan ampliamente como en bovinos lecheros, donde el impacto ha sido mayor (Ibanez-Escriche & Simianer, 2015). Este bajo nivel de adopción puede explicarse porque la exactitud es generalmente más baja y porque las ventajas económicas no son tan grandes. La SG es más ventajosa para características difícil de seleccionar de manera tradicional (Meuwissen et al., 2016).

De acuerdo con Goddard y Hayes (2007), la predicción del valor genético (EBV) de un individuo utilizando información genómica se realiza en tres pasos principalmente: i) utilizar los marcadores para deducir el genotipo de cada animal en cada QTL, ii) estimar el efecto del genotipo de cada QTL en la característica de interés, y iii) sumar el efecto de todos los genotipos de cada QTL en el animal para obtener su EBV genómico (VGn). Sin embargo, últimamente se ha optado por otros métodos para calcular los VGn; por ejemplo, se han utilizado los haplotipos en vez de marcadores únicos para aumentar la varianza genética explicada a través de SG (Hayes et al., 2007; Mulder, Calus & Veerkamp, 2010).

Existen metodologías alternativas para la predicción de valores genómicos; por ejemplo, Legarra, Aguilar y Misztal (2009); Aguilar, Misztal, Legarra y Tsuruta (2010) y Chen, Misztal, Aguilar, Legarra y Muir (2011) han trabajado metodologías que combinan pedigrí y relaciones genómicas. Este es un proceso que se realiza en una sola etapa y permite una evaluación que considera información fenotípica, de pedigrí y genómica disponible. Adicionalmente, este método no tiene limitaciones en los modelos para el análisis genómico.

### **2.2.1. Implementación de un programa de mejoramiento genómico**

La implementación de un programa de mejoramiento genético para bovinos productores de carne, basado en selección genómica sería menos ventajoso que en bovinos lecheros porque la prueba de progenie no es necesaria para características que pueden medirse en los candidatos a selección a edad temprana como la tasa de crecimiento (Meuwissen et al., 2016). Sin embargo, estos autores señalan que en esta especie existen características difíciles de seleccionar como eficiencia en la conversión alimenticia. Dado el costo de registrar estas características, resultaría oneroso establecer una población de entrenamiento grande. Estos autores señalan que se podría establecer una población de entrenamiento multirracial y realizarse análisis no lineales basados en paneles de SNPs de alta densidad o en información de secuenciación. A pesar de estas dificultades SG se ha implementado para Angus en Australia y Estados Unidos. Primeramente se obtienen los valores genómicos utilizando información de ADN. Hay dos maneras en que esto puede hacerse: primera, los genotipos pueden ser provistos a la instancia que habrá de obtener los valores genómicos para calcular las ecuaciones de predicción; segunda, una organización comercial puede proporcionar el servicio de genotipado y utilizar sus propias ecuaciones de predicción para obtener los valores genómicos, los que luego se transfieren a la instancia que habrá de incorporarlos a las evaluaciones genéticas (Meuwissen et al., 2016). El proceso, sin embargo, pudiera ser más complejo, lo importante es

distinguir sus diferentes etapas (Boichard, Ducrocq, Croiseau & Fritz, 2016; Goddard & Hayes, 2007).

### **2.2.2. Ventajas de la selección genómica**

El avance genético anual,  $R = r_{AF} i \sigma_A / L$ , está determinado por la precisión de la predicción del valor genético ( $r_{AF}$ ), la intensidad de selección aplicada ( $i$ ), la magnitud de las diferencias genéticas aditivas entre animales ( $\sigma_A$ ) y el intervalo entre generaciones ( $L$ ) (Boichard et al., 2016). Debido a la disponibilidad de tecnologías moleculares y la posibilidad de obtener el genotipo de un animal a edades tempranas, los intervalos generacionales tenderán a ser más cortos, por tanto el avance genético se espera sea mayor (Blasco & Toro, 2014; Goddard & Hayes, 2007; Yudin, Lukyanov, Voevoda & Kolchanov, 2016).

Las evaluaciones basadas en marcadores moleculares suelen ser más precisas que las evaluaciones cuantitativas tradicionales, lo cual representa una potencial tasa de avance genético mayor. Crossa et al. (2010) mencionan acerca de hallazgos basados en validaciones cruzadas, las cuales indican que los modelos que incluyen información genómica tienen una capacidad predictiva más alta que los modelos basados únicamente en el pedigrí. Para corroborar lo anterior, estos autores trabajaron con un modelo de pedigrí e información de plantas de trigo y maíz, las ganancias en cuanto a capacidad de predicción por la inclusión de marcadores variaron de 7.7 a 35.7%; mientras que la correlación de valores observados y predichos del conjunto de datos de maíz alcanzó valores de hasta 0.79.

La SG podría disminuir considerablemente el costo de operación en sistemas de mejoramiento genético; por ejemplo, en bovinos lecheros se pueden seleccionar animales jóvenes mediante marcadores y evitar las pruebas de progenie, lo que podría reducir los costos de operación hasta 92% (Boichard et al., 2016; Schaeffer, 2006).

En SG se separan las poblaciones de referencia y de prueba. Para la población de referencia es necesario tener los fenotipos y genotipos. Con metodologías genómicas se pueden evaluar características difíciles o caras de medir; por ejemplo, calidad de la canal, respuesta del animal a enfermedades, rendimiento bajo condiciones comerciales, vida productiva, entre otras. La población de entrenamiento puede ser ganado comercial; por su parte, para la población de evaluación basta con tener únicamente el genotipo. Lo anterior posibilita la reducción en los costos para producir animales, ya que no habría necesidad de registrar pedigríes y fenotipos, con excepción del costo del genotipado (Goddard & Hayes, 2007).

En general, la selección de animales ocasiona un aumento de consanguinidad. Sin embargo, la SG puede ser una herramienta útil para reducir la tasa de consanguinidad en programas de mejoramiento genético (Colleau et al., 2015; Yudin et al., 2016). Esto es porque con la genotipificación se puede saber con precisión el grado de parentesco entre individuos, permitiendo así una mejor toma de decisión al momento de aparear animales.

### **2.2.3. Desventajas de la selección genómica**

La selección genómica con la ayuda de marcadores moleculares requiere de poblaciones que cuenten con registros fenotípicos y conexión genética entre ellos; estas poblaciones son de ayuda para conocer el tamaño y tipo de chip adecuado para las características de interés, el método estadístico apropiado y la habilidad predictiva de estos entre otros. Sin embargo, esto resulta costoso y muchas veces estas poblaciones de referencia suelen ser pequeñas e insuficientes (Boichard et al., 2016; Jannink et al., 2010).

Una particularidad de SG es el uso de paneles de alta densidad de marcadores y la estimación del efecto de cada marcador, consecuentemente se necesitan estimar más efectos de predictores  $p$  que el número  $n$  de las observaciones disponibles



(Jannink et al., 2010). Por consiguiente se requieren metodologías de predicción más sofisticadas (de los Campos et al., 2013; Jannink et al., 2010).

En las evaluaciones genéticas tradicionales con metodologías cuantitativas, las principales fuentes de información son los registros fenotípicos y el pedigrí. En las evaluaciones genómicas, adicionalmente se requiere de información de los genotipos en términos de SNPs, la cual generalmente es costosa. Goddard y Hayes (2007) mencionan que ésta es la principal limitante para la implementación de SG. Además, esto puede originar una dependencia tecnológica, ya que la genotipificación se realiza con equipo especializado, generalmente costoso, técnicas especializadas y se tiene que realizar periódicamente.

Rajsic, Weersink, Navabi y Pauls (2016) estudiaron la viabilidad económica de la SG y encontraron que en la mayoría de las ocasiones es conveniente aplicar la SG en programas de mejoramiento genético. Sin embargo, en algunos casos el alto costo de algunas de sus fases dificulta su implementación. En otros casos, por ejemplo, animales adultos probados a través de su progenie, suelen tener exactitudes de predicción altas por lo que la ganancia adicional de exactitud a través de SG puede ser de baja magnitud (Goddard & Hayes, 2007).

Las ecuaciones de predicción usadas en SG requieren ser recalculadas cada cierto número de generaciones y no se pueden usar arbitrariamente en razas diferentes de donde fueron obtenidas (De Roos, Hayes & Goddard, 2009; Kachman et al., 2013). A este respecto, Kachman et al. (2013) utilizaron poblaciones de entrenamiento y evaluación de diferente raza. Los resultados que obtuvieron, aún en razas relativamente cercanas, indicaron una exactitud baja en la predicción de los valores genómicos cuando la raza evaluada no estuvo representada en la población de entrenamiento. Sin embargo, incluir información de otras razas a una población de entrenamiento de una raza, posibilita obtener valores genómicos más robustos a través de razas. Boichard et al. (2016) mencionan que la SG depende en gran medida de la relación entre los candidatos de selección y la población de referencia; conforme aumenta el número de generaciones que separan las

poblaciones de selección y entrenamiento se pierde la eficiencia de las ecuaciones de predicción debido a los eventos de recombinación que se acumulan entre QTLs y sus marcadores circundantes.

### **2.3. Algunas metodologías estadísticas para evaluación genómica**

En los modelos genéticos estándar, los resultados fenotípicos,  $y_i$  ( $i = 1, \dots, n$ ), se consideran como la suma de un valor genético,  $g_i$ , y un residual,  $\varepsilon_i$ ; es decir  $y_i = g_i + \varepsilon_i$ , en el caso de SG se requieren superar varios desafíos estadísticos, tales como modelos capaces de hacer frente al problema de dimensionalidad, la multicolinealidad entre marcadores o la complejidad de las características cuantitativas (Crossa et al., 2010).

En los modelos paramétricos para SG,  $g_i$  se describe como una regresión en covariables de marcadores  $x_{ij}$  ( $j = 1, \dots, p$ , marcadores moleculares), de forma tal que  $g_i = \sum_{j=1}^p x_{ij}\beta_j$ , tal que  $y_i = \sum_{j=1}^p x_{ij}\beta_j + \varepsilon_i$  (o en notación matricial,  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ ), donde  $\beta_j$  es la regresión de  $y_i$  en la covariable correspondiente al  $j$ -ésimo marcador  $x_{ij}$  (Crossa et al., 2010; Meuwissen et al., 2001).

La estimación de los coeficientes de regresión asociados con los marcadores,  $\beta$ , mediante regresión múltiple por métodos convencionales, como máxima verosimilitud y mínimos cuadrados no son viables. Esto es porque el número de observaciones fenotípicas,  $n$ , es mucho menor que el número de marcadores,  $p$ , ( $n \ll p$ ) y la inversa de  $X^tX$  puede no existir, ya que son muy probables las combinaciones lineales entre las columnas de  $X$ , es decir que exista multicolinealidad (Crossa et al., 2010; Nicolás, 2013).

Una opción para calcular los efectos de los marcadores es usar métodos penalizados tales como regresión Ridge (RR), LASSO, la reproducción de los

espacios de Hilbert (reproducing kernel Hilbert spaces, RKHS), la mejor predicción lineal insesgada genómica (GBLUP), la regresión elastic-net (REN) entre otros (Schaefer, Roi & Wolfe, 1984; López, 2012; de los Campos et al., 2013; Jacquin, Cao & Ahmadi, 2016). Estos métodos penalizados se basan en la minimización de la suma de cuadrados del error (SCE), a través de restricciones de los posibles valores de los estimadores para reducir la varianza, obteniendo así estimaciones más precisas (Kyung, Gilly, Ghoshz & Casellax, 2010; Nicolás, 2013).

### 2.3.1. Regresión Ridge

Uno de los métodos más antiguos de estimación penalizada es RR (Hoerl & Kennard, 1970). El estimador tipo Ridge de los coeficientes de regresión se obtiene mediante la ecuación:

$$\hat{\beta}_{RR} = (X^t X + \lambda_{RR} I)^{-1} X^t y$$

donde  $\lambda_{RR} \geq 0$  es una constante no negativa que controla la bondad de ajuste y la complejidad del modelo, medida por la suma de cuadrados de los coeficientes de regresión o la norma  $L_2$  de  $\beta$  (López, 2012). El parámetro de Ridge  $\lambda_{RR}$ , se debe estimar usando los datos (Le Cessie & Van Houwelingen, 1992; Lee & Silvapulle, 1988; Schaefer et al., 1984):

1.  $\lambda_0$  = el valor de  $\beta_R$  donde se estabiliza el gráfico, conocido como la traza de Ridge.
2.  $\lambda_1 = \frac{1}{\beta^t \beta}$
3.  $\lambda_2 = \frac{p+1}{\beta^t \beta}$
4.  $\lambda_3 = \frac{(\lambda_1 - 100\lambda_p)}{99}$ .

donde  $\lambda_1$  y  $\lambda_p$  son el menor y mayor valor propio de  $X^t X$ .

López (2012) menciona que la penalización en RR se hace para la suma de cuadrados de los residuales,  $RSS(y, \beta) = (y - X\beta)^t(y - X\beta)$ , sujeto a la norma L<sub>2</sub>:  $SS(\beta) = \sum_{j=1}^p \beta_j^2 \leq t$ ; o equivalentemente:

$$\hat{\beta}_{RR} = \text{mín}_{\beta} \left\{ (y - X\beta)^t(y - X\beta) + \lambda_{RR} \sum_{j=1}^p \beta_j^2 \right\}$$

La penalidad cuadrática induce la contracción de los coeficientes de regresión estimados hacia cero, lo que introduce un sesgo pero al mismo tiempo reduce la varianza de los estimadores (Hoerl & Kennard, 1970; Zou & Hastie, 2005).

Zou y Hastie (2005) mencionan que RR no proporciona un modelo parsimonioso, ya que mantiene todas las variables en el modelo, es decir, el grado de contracción es homogéneo entre los marcadores. Por esta razón, este método puede no ser apropiado ya que algunos marcadores están localizados en regiones no asociadas con la varianza genética, mientras que los marcadores en otras regiones pueden estar vinculados a QTL (Goddard & Hayes, 2007). Otra desventaja de este método es que  $\lambda_{RR}$  depende de  $\beta$ , que es desconocido y que penaliza a todos los  $\beta_j$ , estén o no en la relación de colinealidad (Nicolás, 2013).

### 2.3.2. Regresión LASSO

La regresión LASSO propuesta por Tibshirani (1996) es otro método de estimación penalizada, el cual, a diferencia de RR, estima los coeficientes de regresión lineal a través de mínimos cuadrados con la restricción L<sub>1</sub> (Park & Casella, 2008), esto es:

$$\hat{\beta}_L = \text{mín}_{\beta} \left\{ (y - X\beta)^t(y - X\beta) + \lambda_L \sum_{j=1}^p |\beta_j| \right\} \quad (1)$$

En los resultados de la Regresión LASSO, algunos coeficientes se hacen cero, lo que hace al modelo parsimonioso, es decir realiza selección de variable y contracción simultáneamente (López, 2012; Nicolás, 2013).

Para resolver el problema de la Regresión Lasso, ecuación (1), Efron, Hastie, Johnstone y Tibshirani (2004) propusieron el algoritmo LARS (Least Angle Regression). Este procedimiento contiene una fórmula matemática simple para acelerar los cálculos, que produce modelos estimados de forma semejante a los procedimientos de selección de variables hacia adelante y hacia atrás, pero que con una pequeña modificación también puede obtener estimaciones LASSO (Nicolás, 2013).

De manera similar que la RR, la regresión LASSO presenta algunas desventajas (Zou & Hastie, 2005). Así, cuando  $p > n$ , la regresión LASSO selecciona a lo más  $n$  variables; en presencia de altas correlaciones por pares en subconjuntos de variables, elige una de cada subconjunto sin importar cuál de ellas es seleccionada, y si todos los predictores están correlacionados, la regresión LASSO es superada por la RR (Nicolás, 2013; Tibshirani, 1996; Zou & Hastie, 2005).

### **2.3.3. Regresión Elastic Net**

Debido a las limitaciones de la regresión LASSO:

1. En el caso de que  $p > n$ , la regresión LASSO selecciona a lo más  $n$  variables debido a que es un problema de optimización convexa (López, 2012).
2. Si existe algún grupo de covariables entre las cuales las correlaciones por pares son muy altas, entonces la regresión LASSO tiende a seleccionar solamente una variable del grupo sin importar cuál de ellas es seleccionada (Nicolás, 2013).
3. En situaciones usuales  $p > n$ , la regresión LASSO se desempeña insatisfactoriamente cuando los predictores están correlacionados y en esta situación es superada por la RR (Tibshirani, 1996).

Zou y Hastie (2005) propusieron el método REN para mejorar el desempeño de la Regresión LASSO; cuya función de penalidad es la combinación de las normas de RR y la Regresión LASSO ( $L_1$  y  $L_2$ ), es decir:

$$\hat{\beta}_{EN} = \min_{\beta} \left\{ (y - X\beta)^t (y - X\beta) + \lambda_1 \sum_{j=1}^p |\beta_j| + \lambda_2 \sum_{j=1}^p \beta_j^2 \right\}$$

donde  $\lambda_1$  y  $\lambda_2$  son constantes no negativas, las cuales controlan el peso asignado a las penalidades  $L_1$  y  $L_2$  respectivamente. Este problema de optimización es equivalente a:

$$\min_{\beta} \{(y - X\beta)^t (y - X\beta)\}, \text{ sujeto a las restricciones: } (1 - \alpha) \sum_{j=1}^p |\beta_j| + \alpha \sum_{j=1}^p \beta_j^2 \leq t$$

donde  $\alpha = \lambda_2 / (\lambda_1 + \lambda_2)$  y  $t$  es una constante arbitraria positiva (López, 2012; Nicolás, 2013).

Para resolver el problema de la REN eficientemente, Zou y Hastie (2005) propusieron el algoritmo LARS-EN, basado en el algoritmo LARS propuesto por Efron et al. (2004) para resolver el problema de la regresión LASSO.

### 2.3.4. Regresión Kernel (RKHS)

La RKHS es un método de regresión semiparamétrica auxiliar para modelar la epistasis en la predicción de valores genómicos, ya que se considera que puede captar los efectos no aditivos debido a que no asume linealidad (Gianola & Van Kaam, 2008; Heslot, Yang, Sorrells & Jannink, 2012; Jiang & Reif, 2015). En este método se utiliza una función llamada kernel, mediante la cual se transforma el conjunto de datos (marcadores) en un conjunto de medidas de similitud; por ejemplo, distancias entre pares de observaciones dando como resultado una matriz cuadrada, que puede ser utilizada como matriz diseño en un modelo lineal normal (Nicolás, 2013). Las metodologías kernel se utilizan como métodos de regresión de elección para la predicción genómica (Jacquin et al., 2016). Las características cuantitativas están influenciadas por muchas regiones genómicas trabajando junto

con otras, generando de esta manera interacciones. Un número creciente de enfoques estadísticos basados en kernels intentan capturar los efectos genéticos no aditivos, de manera paramétrica o no paramétrica (Morota & Gianola, 2014).

El modelo para Regresión RKHS puede ser formulado de la siguiente manera (Heslot et al., 2012):

$$y = W\mu + K_h\alpha + \varepsilon$$

donde  $\mu$  es un vector de efectos fijos y  $\varepsilon$  es un vector de residuos aleatorios; se asume que los parámetros  $\alpha$  y  $\varepsilon$  tienen distribuciones previas independientes  $\alpha \sim N(0, K_h\sigma_\alpha^2)$  y  $\varepsilon \sim N(0, I\sigma_\varepsilon^2)$ , respectivamente. La matriz  $K_h$  depende de una función reproductora de kernel con un parámetro de suavizado  $h$ , que mide la "distancia genómica" entre genotipos y puede interpretarse como una matriz de correlación (Heslot et al., 2012).

### 2.3.5. GBLUP

Henderson (1984) desarrolló las ecuaciones de modelos mixtos, las cuales producen simultáneamente el mejor estimador lineal insesgado (BLUE) y el mejor predictor lineal insesgado (BLUE). El método GBLUP (Genomic Best Linear Unbiased Prediction) es una modificación de las ecuaciones de modelos mixtos en las cuales se usa una matriz de relaciones genómicas en lugar de la matriz de relaciones genéticas aditivas (Clark & van der Werf, 2013).

El modelo lineal mixto es:

$$y = X\beta + Zu + \varepsilon \tag{2}$$

donde  $X$  y  $Z$  son matrices de incidencia conocidas,  $\beta$  es el vector de efectos fijos y  $u$  es el vector de efectos aleatorios, además:

$$E \begin{bmatrix} u \\ \varepsilon \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \text{ y } V \begin{bmatrix} u \\ \varepsilon \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix}$$

Para estimar  $\beta$  y predecir  $u$ , dadas las matrices de (co)varianzas  $G$  y  $R$  en la ecuación (2) se reescriben de la manera siguiente:

$$y = X\beta + \epsilon$$

donde  $\epsilon = Zu + \varepsilon$ , se tiene que  $Cov(\epsilon) \equiv V = ZVZ^t = R$ , es decir, están correlacionados (Nicolás, 2013).

## **2.4. Complementación de metodologías cuantitativas y genómicas**

En un principio cuando se diseñó la selección a través de marcadores se pensó en encontrar cada uno de los genes que influyen en las características. Sin embargo, después se descubrió que muchas características están influenciadas por muchos genes, por lo que encontrar cada uno de los genes es un problema muy complejo. Posteriormente, Meuwissen et al. (2001) propusieron la SG con el uso de paneles de alta densidad de marcadores distribuidos a lo largo del genoma, para después determinar el efecto individual de cada uno. Las altas densidades de marcadores aseguran que por lo menos un marcador quede en desequilibrio de ligamiento con los genes de características cuantitativas (QTL).

Los factores que rigen las ganancias adicionales a la selección asistida por marcadores (Boichard et al., 2016; Goddard & Hayes, 2007) son:

- La mayor precisión que los valores genéticos predichos (EBV) pre-existentes. Si la precisión ya es alta, puede haber baja ganancia adicional. En consecuencia, las ganancias son mayores cuando la selección tradicional es más difícil; por ejemplo, características que se expresan solo en hembras, de baja heredabilidad o de medición costosa.
- La proporción de la varianza genética explicada por los marcadores de ADN.



- La precisión con la que se estima el efecto de los alelos marcados de los locus de características cuantitativas (QTL).
- El potencial de reducir el intervalo entre generaciones, utilizando SG al obtener los genotipos a una edad más temprana.

Considerando lo anterior, puede deducirse que SG no puede reemplazar las evaluaciones genéticas cuantitativas; antes bien, se pueden usar ambas para complementarse.

Crossa et al. (2013) compararon modelos de análisis de información fenotípica y genómica. Estos autores encontraron ganancias consistentes en la exactitud de la predicción al combinar en los análisis información fenotípica y de genotipado por secuenciación.

## **2.5. Análisis uni- vs bivariados**

Estudios sobre el uso de modelos multivariados reportan una mejor capacidad predictiva de los valores genéticos (Henderson & Quaas, 1976; Ramírez-Valverde et al., 2007). Sin embargo, en SG los reportes del uso de modelos multivariados en selección genómica son muy pocos ( Jia & Jannink, 2012; Malosetti, Ribaut, Vargas, Crossa & Van Eeuwijk, 2008).

El análisis por separado de una característica puede subutilizar la covarianza que entre dos características o sus componentes y consecuentemente perder precisión estadística para detectar los efectos de SNPs (Serão et al., 2013). Estos autores realizaron un estudio utilizando modelos uni- y multi-SNP, así como análisis GWAS uni- y bivariados. Los resultados del estudio demostraron el valor de la complementariedad de los análisis multivariados.

En ocasiones la SG directa (análisis univariados) resulta ineficaz, dependiendo de la heredabilidad y la complejidad de la arquitectura genética (de los Campos et al.,

2013; Dos Santos, De Castro Vasconcellos, Pires, Balestre & Von Pinho, 2016). En consecuencia, la SG indirecta (análisis multivariados) se puede usar para mejorar la exactitud de la predicción de características de heredabilidad baja con alta correlación con la característica de interés.

## **2.6. Perspectivas de selección genómica**

Se espera que en el futuro la SG se extenderá a diferentes especies. Actualmente en la mayoría de las especies, el costo de la SG limita su uso rutinario generalizado, ya que muchos esquemas de mejoramiento genético no pueden pagar la inversión para el genotipado en una población de referencia. Sin embargo, recientemente el desarrollo tecnológico ha reducido el costo de genotipificación, además se han diseñado estrategias para abaratar los costos globales de la genotipificación, por ejemplo, en especies prolíficas el genotipado con un chip de baja densidad es suficiente para rastrear todos los segmentos cromosómicos en los candidatos a selección, siempre y cuando los padres estén genotipados con chips de alta densidad (Boichard et al., 2016).

Con la integración de nueva información biológica, la SG podría ser más precisa y de uso más amplio. Se ha observado que la eficacia de la selección genómica depende en gran medida de la relación entre los candidatos de selección y la población de referencia, por tal motivo la predicción genómica es ineficiente entre razas (Boichard et al., 2016). Meuwissen et al. (2016) consideran que las futuras aplicaciones de SG serán dentro de raza, donde la exactitud se obtendrá manteniendo poblaciones de referencia grandes o a través de razas donde la exactitud se obtendrá de poblaciones de referencia, a través de razas y métodos de SG basados en paneles de alta densidad que se enfoquen en regiones causales.

**Nuevos fenotipos y objetivos de mejora.** Debido a las bondades de la SG de poder seleccionar para una característica tan pronto como se tenga una población de entrenamiento, se están generando poblaciones de referencia para rasgos que

durante mucho tiempo se creyeron imposibles de seleccionar, gracias al desarrollo de técnicas de registro innovadoras (como la espectrometría de infrarrojo medio para composición y propiedades de la leche), el uso de datos de ganado (salud, reproducción, comportamiento), el uso de datos comerciales (tarjetas sanitarias, características de la canal en rastro), o la colaboración internacional para características costosas de medir; por ejemplo, la eficiencia de la alimentación o la emisión de metano (Boichard et al., 2016).

## 2.7. Literatura Citada

Aguilar, I., Misztal, I., Legarra, A., & Tsuruta, S. (2010). Efficient computation of the genomic relationship matrix and other matrices used in single-step evaluation. *Journal of Animal Breeding and Genetics*, 128, 422-428.

Bellman, R. (1969). A new type of approximation leading to reduction of dimensionality in control processes. *Journal of Mathematical Analysis and Applications*, 27, 454-459, [https://doi.org/10.1016/0022-247X\(69\)90061-4](https://doi.org/10.1016/0022-247X(69)90061-4).

Blasco, A., & Toro, M. A. (2014). A short critical history of the application of genomics to animal breeding. *Livestock Science*, 166(1), 4–9, <https://doi.org/10.1016/j.livsci.2014.03.015>.

Boichard, D., Ducrocq, V., Croiseau, P., & Fritz, S. (2016). Genomic selection in domestic animals: Principles, applications and perspectives. *Comptes Rendus Biologies*, 339(7-8), 274–277, <https://doi.org/10.1016/j.crv.2016.04.007>.

Chen, C. Y., Misztal, I., Aguilar, I., Legarra, A., & Muir, W. M. (2011). Effect of different genomic relationship matrices on accuracy and scale. *Journal of Animal Science*, 89, 2673-2679.

Clark S.A., van der Werf J. (2013). Genomic Best Linear Unbiased Prediction

(gBLUP) for the Estimation of Genomic Breeding Values. In: Gondro C., van der Werf J., Hayes B. (eds) *Genome-Wide Association Studies and Genomic Prediction. Methods in Molecular Biology (Methods and Protocols)*, Vol 1019. Humana Press, Totowa, NJ.

Colleau, J. J., Fritz, S., Guillaume, F., Baur, A., Dupassieux, D., Boscher, M. Y., Journaux, L., Eggen, A., & Boichard, D. (2015). Simulation des potentialités de la sélection génomique chez les bovins laitiers. *Productions Animales*, 28(3), 251–258. Available at: [http://www.journees3r.fr/IMG/pdf/2009\\_12\\_05\\_Colleau.pdf](http://www.journees3r.fr/IMG/pdf/2009_12_05_Colleau.pdf). Accessed on October 20, 2017.

Crossa, J., Beyene, Y., Kassa, S., Pérez, P., Hickey, J. M., Chen, C., de los Campos, G., Burgueño, J., Windhausen, V. S., Buckler, E., Jannink, J-L., Lopez C., M. A., & Babu, R. (2013). Genomic prediction in maize breeding populations with genotyping-by-sequencing. *Genes, Genomes, Genetics*, 3, 1903-1926, doi: 10.1534/g3.113.008227.

Crossa, J., de Los Campos, G., Pérez, P., Gianola, D., Burgueño, J., Araus, J. L., Makumbi, D., Singh, R. P., Dreisigacker, S., Yan, J., Arief, V., Banzinger, M., & Braun, H.-J. (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics*, 186(2), 713–724, <https://doi.org/10.1534/genetics.110.118521>.

de los Campos, G., Hickey, J. M., Pong-Wong, R., Daetwyler, H. D., & Calus, M. P. L. (2013). Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics*, 193(2), 327–345, <https://doi.org/10.1534/genetics.112.143313>

De Roos, A. P. W., Hayes, B. J., & Goddard, M. E. (2009). Reliability of genomic predictions across multiple populations. *Genetics*, 183(4), 1545–1553. Available at: <https://doi.org/10.1534/genetics.109.104935>. Accessed on October 20, 2017.

- Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, 32(2), 407-499. Available at: <https://doi.org/doi:10.1214/009053604000000067>. Accessed on October 20, 2017.
- Eguiarte, L. E., Aguirre-Liguori, J. A., Jardón-Barbolla, L., Aguirre-Planter, E., & Souza, V. (2013). Genómica de poblaciones: nada en evolución va a tener sentido si no es a la luz de la genómica, y nada en genómica tendrá sentido si no es a la luz de la evolución. *TIP Rev. Esp. Cienc. Quím. Biol.*, 16(1), 42-56, doi:10.1016/s1405-888x(13)72077-1.
- Gianola, D., & Van Kaam, J. B. C. H. M. (2008). Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. *Genetics*, 178(4), 2289-2303. Available at: <https://doi.org/10.1534/genetics.107.084285>. Accessed on October 21, 2017.
- Goddard, M. E., & Hayes, B. J. (2007). Genomic selection. *Journal of Animal Breeding and Genetics*, 124(6), 323–330. Available at: <https://doi.org/10.1111/j.1439-0388.2007.00702.x>. Accessed on October 20, 2017.
- Grover, A., & Sharma, P. C. (2014). Development and use of molecular markers: past and present. *Critical Reviews In Biotechnology*, 36(2), 290-302, doi:10.3109 / 07388551.2014.959891.
- Hayes, B. J., Chamberlain, A. J., McPartlan, H., Macleod, I., Sethuraman, L., & Goddard, M. E. (2007). Accuracy of marker-assisted selection with single markers and marker haplotypes in cattle. *Genetical Research*, 89(4), <https://doi.org/10.1017/S0016672307008865>.
- Hayes, B. J., Bowman, P. J., Chamberlain, A. C., Verbyla, K., & Goddard, M. E. (2009). Accuracy of genomic breeding values in multi-breed dairy cattle populations. *Genetics Selection Evolution*, 41(1), 51,

<https://doi.org/10.1186/1297-9686-41-51>

Henderson, C. R. (1984). Applications of linear models in animal breeding. Univ. Guelph, Guelph, Canada.

Henderson, C. R., & Quaas, R. L. (1976). Multiple trait evaluation using relatives' records. *Journal of Animal Science*, 43(6), 1188-1197.

Heslot, N., Yang, H.-P., Sorrells, M. E., & Jannink, J.-L. (2012). Genomic Selection in Plant Breeding: A Comparison of Models. *Crop Science*, 52(1), 146. Available at: <https://doi.org/10.2135/cropsci2011.09.0297>. Accessed on October 21, 2017.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55–67, <https://doi.org/10.1080/00401706.1970.10488634>.

Ibanez-Escriche, N., & Simianer, H. (2015). Animal breeding in the genomics era. *Animal Frontiers*, 5:4-5. Doi:10.2527/af.2016-0001.

Jannink, J.-L., Lorenz, A. J., & Iwata, H. (2010). Genomic selection in plant breeding: from theory to practice. *Briefings in Functional Genomics*, 9(2), 166–177, <https://doi.org/10.1093/bfgp/elq001>.

Jacquín, L., Cao, T. V., & Ahmadi, N. (2016). A unified and comprehensible view of parametric and kernel methods for genomic prediction with application to rice. *Frontiers in Genetics*, 7(AUG), 1–16. Available at: <https://doi.org/10.3389/fgene.2016.00145>. Accessed on October 21, 2017.

Jiang, Y., & Reif, J. C. (2015). Modeling epistasis in genomic selection. *Genetics*, 201(2), 759-768. Available at: <https://doi.org/10.1534/genetics.115.177907>. Accessed on October 21, 2017.

- Kachman, S. D., Spangler, M. L., Bennett, G. L., Hanford, K. J., Kuehn, L. A., Snelling, W. M., Thallman, R. M., Saatchi, M., Garrick, D. J., Schnabel, R. D., Taylor, J. F. & Pollak, E. J. (2013). Comparison of molecular breeding values based on within- and across-breed training in beef cattle. *Genetics Selection Evolution*. 45, 30-38.
- Kyung, M., Gilly, J., Ghoshz, M., & Casellax, G. (2010). Penalized regression, standard errors, and Bayesian lassos. *Bayesian Analysis*, 5(2), 369–412. Available at: <https://doi.org/10.1214/10-BA607>. Accessed on October 2017.
- Le Cessie, S., & Van Houwelingen, J. C. (1992). Ridge Estimators in Logistic Regression. *Appl. Statist*, 41(1), 191–201. Available at: [http://gs3sr3zm5k.search.serialssolutions.com.ezproxy.conricyt.org/?ctx\\_ver=Z39.88-2004&ctx\\_enc=info%3Aofi%2Fenc%3AUTF-8&rft\\_id=info%3Aid%2Fsummon.serialssolutions.com&rft\\_val\\_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=article&rft.atitle=RIDGE+](http://gs3sr3zm5k.search.serialssolutions.com.ezproxy.conricyt.org/?ctx_ver=Z39.88-2004&ctx_enc=info%3Aofi%2Fenc%3AUTF-8&rft_id=info%3Aid%2Fsummon.serialssolutions.com&rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=article&rft.atitle=RIDGE+). Accessed on October 20, 2017.
- Lee, A. H., & Silvapulle, M. J. (1988). Ridge estimation in logistic regression. *Communications in Statistics -Simulation and Computation*, 17(4), 1231-1257. Available at: <https://doi.org/10.1080/03610918808812723>. Accessed on October 20, 2017.
- Legarra, A., Aguilar, I., & Misztal, I. (2009). A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science*, 92, 4656-4663.
- Lewontin, R. C., & Hubby, J. L. (1966). A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics*, 54, 595-609. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1211186/pdf/595.pdf>. Consultado el 29 de Mayo de 2017.

López C., M. A. (2012). Aplicación del Elastic Net LASSO y modelos relacionados en Selección Genómica basados en Marcadores Moleculares. Colegio de Postgraduados (COLPOS). Disponible en: [http://www.biblio.colpos.mx:8080/jspui/bitstream/handle/10521/256/Sanchez\\_Borja\\_M\\_DC\\_Fitosanidad\\_2010.pdf?sequence=1](http://www.biblio.colpos.mx:8080/jspui/bitstream/handle/10521/256/Sanchez_Borja_M_DC_Fitosanidad_2010.pdf?sequence=1). Consultado el 20 de octubre de 2017.

Martínez, M. C., Helguera, M., & Carrera, A. (2010). Marcadores Moleculares. En Biotecnología y Mejoramiento Vegetal II. Levitus, G., Echenique, V., Rubinstein, C., Hopp, E., & Mroginski, L. (Eds.). 2a ed. Pp. 70-85. Instituto Nacional de Tecnología Agropecuaria. Buenos Aires, Argentina. INTA. Disponible en: [http://intainforma.inta.gov.ar/wp-content/uploads/2010/09/bio\\_WEB.pdf](http://intainforma.inta.gov.ar/wp-content/uploads/2010/09/bio_WEB.pdf). Consultado el 19 de octubre de 2017.

Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense markers maps. *Genetics*. falta volumen y páginas <https://doi.org/11290733>.

Meuwissen, T., Hayes, B., & Goddard, M. (2016). Genomic selection: A paradigm shift in animal breeding. *Animal Frontiers*, 5, 7-14, doi:10.2527/af.2016-0002.

Morota, G., & Gianola, D. (2014). Kernel-based whole-genome prediction of complex traits: a review. *Frontiers in Genetics*, 5, 1-13. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4199321/pdf/fgene-05-00363.pdf>. Accessed on October 20, 2017.

Mulder, H. A., Calus, M. P. L., & Veerkamp, R. F. (2010). Prediction of haplotypes for ungenotyped animals and its effect on marker-assisted breeding value estimation. *Genetics Selection Evolution*, 42(1), 10-24. Available at: <https://doi.org/10.1186/1297-9686-42-10>. Accessed on October 19, 2017.

Nicolás, E. (2013). Método De Regresión Bayesiana Para Selección Genómica. Colegio de Postgraduados (COLPOS). Disponible en:



[http://colposdigital.colpos.mx:8080/jspui/bitstream/handle/10521/2089/Nicolas\\_Popoca\\_E\\_MC\\_Estadistica\\_2013.pdf?sequence=1](http://colposdigital.colpos.mx:8080/jspui/bitstream/handle/10521/2089/Nicolas_Popoca_E_MC_Estadistica_2013.pdf?sequence=1). Consultado el 20 de octubre de 2017.

Park, T., & Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association*, 103(482), 681-686. Available at: <https://doi.org/10.1198/016214508000000337>. Accessed on October 20, 2017.

Rajic, P., Weersink, A., Navabi, A., & Peter Pauls, K. (2016). Economics of genomic selection: The role of prediction accuracy and relative genotyping costs. *Euphytica*, 210(2), 259–276. Available at: <https://doi.org/10.1007/s10681-016-1716-0>. Accessed on October 20, 2017.

Ramírez-Valverde, R., Hernández-Alvarez, C. O., Núñez-Domínguez, R., Ruíz-Flores, A., & García-Muñíz, J. G. (2007). Univariate vs multivariate analyses in genetic evaluation of growth traits in two cattle breeds. *Agrociencia*, 41, 271–282.

Schaefer, R. L., Roi, L. D., & Wolfe, R. A. (1984). A ridge logistic estimator. *Commun. Statist.-Theor. Meth.*, 13(1), 99–113, <https://doi.org/10.1080/03610928408828664>.

Schaeffer, L. R. (2006). Strategy for applying genome wide selection in dairy cattle. *Journal of Animal Breeding and Genetics*, 123(4), 218–223.

Seeb, J. E., Carvalho, G., Hauser, L., Naish, K., Roberts, S., & Seeb, L. W. (2011). Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. *Molecular Ecology Resources*, 11(S1), 1-8, doi:10.1111/j.1755-0998.2010.02979.x.

Serão, N. V. L., González-Peña, D., Beever, J. E., Bollero, G. A., Southey, B. R., Faulkner, D. B., & Rodriguez-Zas, S. L. (2013). Bivariate genome-wide

association analysis of the growth and intake components of feed efficiency. *PLoS ONE*, 8(10), e78530, doi:10.1371/journal.pone.0078530.

Tibshirani, R. (1996). Regression selection and shrinkage via the Lasso. *Journal of the Royal Statistical Society B*, 58(1), 267–288. Available at: <https://doi.org/10.2307/2346178>. Accessed on October 20, 2017.

Weaver, R. F. (2012). *Molecular Biology*. 5<sup>th</sup> Edition. Lange, M. (Ed.). McGraw-Hill. New York, NY. Available at: [https://aliazamani.files.wordpress.com/2015/09/molecular\\_biology\\_r-\\_f-\\_weaver\\_5th\\_ed.pdf](https://aliazamani.files.wordpress.com/2015/09/molecular_biology_r-_f-_weaver_5th_ed.pdf). Accessed on May 20th, 2017.

VanRaden, P. M., Van Tassell, C. P., Wiggans, G. R., Sonstegard, T. S., Schnabel, R. D., Taylor, J. F., & Schenkel, F. S. (2009). Invited review: Reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science*, 92(1), 16–24, <https://doi.org/10.3168/jds.2008-1514>.

Yudin, N. S., Lukyanov, K. I., Voevoda, M. I., & Kolchanov, N. A. (2016). Application of reproductive technologies to improve dairy cattle genomic selection. *Russian Journal of Genetics: Applied Research*, 6(3), 321–329, <https://doi.org/10.1134/S207905971603014X>.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 67(2), 301–320, <https://doi.org/10.1111/j.1467-9868.2005.00503.x>. Accessed on October 20, 2017.

### **3. ANÁLISIS UNI- VS MULTIVARIADOS EN LA PREDICCIÓN DE VALORES GENÓMICOS DE CARACTERÍSTICAS DE CRECIMIENTO EN GANADO SUIZO EUROPEO**

#### **3.1. Resumen**

El objetivo fue comparar análisis genómicos univariados vs bivariados para pesos al nacer (PN), al destete (PD) y al año (PA) en bovinos Suizo Europeo. Los criterios de comparación fueron la jerarquización de los valores genómicos predichos (VGn), la habilidad de predicción de los modelos y los coeficientes de regresión de los VGn predichos con una alternativa de análisis sobre los VGn obtenidos con la otra, y viceversa. Los estimadores de heredabilidad genómica directa de PN, PD y PA fueron 0.72, 0.65 y 0.64 para PN, PD y PA, respectivamente; el correspondiente a heredabilidad genómica materna para PD fue 0.21. Los estimadores de los coeficientes de correlación de Pearson y Spearman entre los VGn predichos con análisis univariados y bivariados variaron de 0.845 a 0.994 ( $p \leq 0.0001$ ) y de 0.843 a 0.989 ( $p \leq 0.0001$ ), respectivamente. Los análisis bivariados tuvieron mejor habilidad predictiva que los univariados; excepto los análisis bivariados de PN con PD y PA. Los estimadores de los coeficientes de correlación Pearson entre VGn provenientes de análisis univariados (VGnU) y los VGn provenientes de análisis bivariados (VGnB), cuando los VGnU fueron las variables independientes, variaron de  $0.96 \pm 0.06$  a  $1.37 \pm 0.06$  ( $p \leq 0.0001$ ); mientras que cuando los VGnB lo fueron, variaron de  $0.70 \pm 0.02$  a  $0.82 \pm 0.01$  ( $p \leq 0.0001$ ). En conclusión, la jerarquización de los valores genómicos predichos con modelos univariados es similar a la obtenida con modelos bivariados, la habilidad predictiva de los modelos bivariados es mayor que la de los univariados, especialmente para características registradas en el tiempo más cerca la una de la otra, y los VGnU pueden utilizarse para predecir los VGnB y viceversa.

Palabras clave: Evaluación genómica, análisis univariados, análisis bivariados.

## UNI- VS MULTIVARIATE ANALYSES IN PREDICTION OF GENOMIC VALUES FOR GROWTH TRAITS IN BRAUNVIEH CATTLE

### 3.2. Abstract

The study aimed to compare the univariate vs bivariate genomic analysis of birth (BW), weaning (WW), and yearling weight (YW) of Braunvieh cattle. The criteria for the comparison were the ranking of the predicted genomic values (GVn), the ability of prediction of the models, and the regression coefficient estimates of GVn predicted with one method of analysis on the VGn predicted with the other alternative, and vice-versa. The direct genomic heritability estimates for BW, WW, and YW were 0.72, 0.65, and 0.64. The maternal genomic heritability estimate for WW was 0.21. The estimates for the Pearson and Spearman correlation coefficients between GVn from uni- and bivariate analyses varied from 0.845 to 0.994 ( $p \leq 0.0001$ ), and from 0.843 to 0.989 ( $p \leq 0.0001$ ), respectively. The bivariate analysis had better predictive ability than the univariate alternative; exception was the bivariate analyses for BW with WW and YW. Regarding the estimates for the Pearson regression coefficients of GVn from univariate analysis (GVnU) on GVn from bivariate analysis (GVnB); when the independent variables were GVnU, the estimates varied from  $0.96 \pm 0.06$  to  $1.37 \pm 0.06$  ( $p \leq 0.0001$ ); meanwhile, when they were GVnB, the estimates varied from  $0.70 \pm 0.02$  to  $0.82 \pm 0.01$  ( $p \leq 0.0001$ ). In conclusion, the ranking of GVnU is similar to that obtained with the GVnB, the predictive ability of bivariate models is higher than that of the univariate alternative. This is especially true for traits recorded closer one to each other. The GVnU could be used to predict the GVnB, and vice versa.

Keywords: Genomic evaluation, univariate analysis, bivariate analysis.

### 3.3. Introducción

De acuerdo con Van der Sluis et al. (2015), los análisis del genoma completo (genome-wide association studies, GWAS) usualmente prueban la asociación entre un fenotipo y un número grande polimorfismos de nucleótido simple (single nucleotide polymorphisms, SNPs). Estos autores afirman que esto constituye una limitante en dos maneras: primera, las características frecuentemente son multivariadas, perdiéndose potencia estadística si se analizan por separado; y segunda, los análisis basados en genes deberán preferirse para disminuir los problemas de múltiples pruebas. Otra alternativa de análisis es considerar la información de un segmento del genoma para realizar el GWAS (gene-based GWAS; Basu et al., 2013). De aquí el interés de analizar características productivas de manera simultánea.

De acuerdo con Henderson y Quaas (1976), los modelos multivariados utilizando registros fenotípicos de parientes, son de gran utilidad especialmente cuando las características tienen baja heredabilidad o no se expresan en los animales candidatos a selección. Ramírez-Valverde et al. (2007) compararon análisis uni- y multivariados de pesos al nacimiento, al destete y al año en bovinos Angus y Tropicarne. Los criterios de comparación incluyeron la magnitud de los estimadores de los parámetros genéticos, la habilidad de predicción de los modelos y la exactitud de los valores genéticos predichos. Para algunas variables los análisis bivariados dieron mejores resultados en los tres criterios de comparación; sin embargo, las ventajas al usar análisis trivariados no se mantuvieron. Además, el cambio en la alternativa de análisis se reflejó en cambios en la jerarquización de los valores genéticos predichos de los animales para casi todas las variables. Por su parte, Serão et al. (2013) realizaron un estudio considerando información genómica para el análisis uni- y bivariado de ganancia diaria de peso y consumo de materia seca. Los resultados de este estudio demuestran la complementariedad de los análisis uni- y bivariados en estudios GWAS. Estos autores argumentan que al considerar la covarianza entre los componentes se gana precisión estadística. Sin embargo,

debido al incremento en la demanda computacional de los análisis genómicos, frecuentemente no se les suele utilizar en evaluaciones genéticas, dada la gran cantidad de información que se maneja. Con base en lo anterior, el objetivo del presente estudio fue comparar análisis uni- vs bivariados en la predicción de valores genómicos para pesos al nacimiento, al destete y al año en bovinos Suizo Europeo, utilizando varios de criterios para esta comparación.

### **3.4. Materiales y métodos**

#### **3.4.1. Origen de la información**

La información fenotípica del ganado Suizo Europeo provino de la base de datos de la Asociación Mexicana de Criadores de Ganado Suizo de Registro. Esta asociación registra los antecedentes genealógicos de los animales, la identificación del animal de acuerdo con el propietario (número privado), el número de registro del animal (número único de identificación) y las características de crecimiento, con base en el peso vivo del animal al nacimiento (PN), al destete ajustado a 218 días (PD) y al año (PA).

Para la información genómica se recolectaron muestras de pelo de animales de diferentes ranchos ubicados en los estados de Colima, Jalisco y Veracruz. Posteriormente las muestras se enviaron a la empresa GeneSeek (Lincoln, NE, USA) para obtener el genotipado de las muestras. El chip utilizado para genotipar los animales fue el Genomic Profile Bovine LD, con 30,000 marcadores SNPs. Este chip se eligió porque sus SNP se seleccionaron para obtener la máxima información en las razas bovinas Angus, Red Angus, Charolais, Gelbvieh, Hereford, Limousin, Simmental, Holstein, Jersey, Brown Swiss, Ayrshire, Guernsey, Nellore, Brahman, Gir, Girolando, Brangus, Beefmaster y Braford.

### **3.4.1. Edición de información genotípica**

La edición de los genotipos Cuadro 1 se realizó recodificando los genotipos, esto es, se cambió AA, AB y BB por 0, 1 y 2, respectivamente. Se eliminaron los SNPs monomórficos y aquellos cuya frecuencia de alelo fue mayor que 0.95 o menor que 0.05. Los genotipos faltantes de marcadores se imputaron utilizando las muestras de distribución marginal de los genotipos de los marcadores. Después de la edición, quedaron 26,207 marcadores.

### **3.4.2. Edición de información fenotípica**

Para PN se utilizaron 32,405 registros, se eliminaron valores que estuvieron tres desviaciones estándar por arriba o por debajo de la media. Registros sin información de hato y edad de la madre también fueron descartados. Los grupos contemporáneos se definieron combinando los efectos de hato, año y época de nacimiento. Las épocas de nacimiento se definieron considerando la cantidad y distribución de la precipitación y el promedio de temperatura media mensual históricos en las estaciones climatológicas más cercanas a los ranchos estudiados (SMN, 2017). Los grupos contemporáneos con varianza cero fueron descartados. Al final se retuvieron 29,718 registros de PN.

Para PD se utilizaron inicialmente 29,059 registros. Los registros eliminados fueron aquellos con edades al destete erróneas, los que estuvieron fuera de la media más, y menos, tres desviaciones estándar y los que no tuvieron información de hato y edad de la madre. Los grupos contemporáneos se formaron de manera similar que para PN. El manejo para PD se definió de tres maneras: primera, alimentación de la cría con leche de la madre; segunda, leche de la madre más alimento balanceado; y tercera, leche de la madre y de una nodriza más alimento balanceado. Los registros sin información de manejo fueron eliminados. Después de la edición quedaron 25,505 registros de PD.

Para PA inicialmente se tuvieron 19,911 registros. Los datos descartados fueron edades de registros de PA erróneos, los que estuvieran fuera del rango de la media más, y menos, tres desviaciones estándar y los que no tuvieron información de hato. Los grupos contemporáneos se formaron de manera similar que para PN y PD. El manejo para PA se definió de tres maneras: primera, el animal se mantuvo en pastoreo; segunda, el animal estuvo en pastoreo y recibió alimento balanceado; y tercera, el animal estuvo estabulado y recibió una ración balanceada. Al final quedaron 17,985 registros de PA. La edición de la información fenotípica se realizó con el software R (R Core Team, 2017) y Microsoft Excel (2016).

### **3.4.3. Análisis estadístico**

Las evaluaciones genómicas, tanto para análisis uni- como bivariados, se realizaron en dos etapas. En la primera, se pre-ajustaron los registros fenotípicos de la base de datos completa, editada como se describió previamente. En los apéndices del 1 al 3 se presentan los códigos en R utilizados para realizar estos ajustes. Posteriormente, de ella se extrajeron los registros fenotípicos de los animales genotipados y se utilizaron como variable de respuesta en los análisis finales. Esta manera fue elegida con la finalidad de utilizar la mayor información posible.

También en la primera etapa se probaron diferentes modelos estadísticos que incluyeron diversos efectos para cada característica, con el fin de determinar los mejores modelos de ajuste para los fenotipos y determinar los efectos ambientales para los cuales habría que pre-ajustar. Los pre-ajustes para PN se hicieron para el efecto aleatorio de grupo contemporáneo, así como para los efectos fijos de sexo del animal, y los efectos lineal y cuadrático de la covariable edad de la madre. Los pre-ajustes de PD se realizaron para el efecto aleatorio de grupo contemporáneo y para los efectos fijos de sexo, manejo, y los efectos lineal y cuadrático de la covariable edad de la madre. Los pre-ajustes para PA se realizaron para grupo contemporáneo, considerándolo como efecto aleatorio, y para sexo del animal, manejo y los efectos lineal y cuadrático de la covariable edad al registro de PA,



considerándolos como efectos fijos. En el Cuadro1 se presentan los estadísticos descriptivos para PN, PD y PA, una vez realizados los ajustes.

En la segunda etapa, los valores genómicos y los componentes de varianza de los modelos univariados se obtuvieron con el paquete BGLR (Pérez & de los Campos, 2014; de los Campos & Pérez-Rodríguez, 2016), utilizando el método de mejor predicción lineal insesgada genómica (GBLUP) propuesto por VanRaden et al. (2009), desarrollado a partir del método BLUP (Henderson, 1984).

El modelo utilizado para la predicción de los VGn para los análisis univariados fue.

$$y = X\beta + Zu + e$$

donde  $y$  es el vector de fenotipos;  $X$  y  $Z$  son matrices de incidencia que relacionan las mediciones fenotípicas con los efectos fijos  $\beta$  y los efectos aleatorios  $u$ ;  $y$   $e$  es el vector de efectos aleatorios de error residual. Las varianzas aditivas genética y residual se supusieron:  $Var(u) = G\sigma_a^2$  y  $Var(e) = I\sigma_e^2$ , respectivamente.

Las ecuaciones de modelos mixtos para GBLUP se describen como:

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

En donde:  $\alpha$  es la proporción entre la varianza residual y la varianza genética aditiva  $\alpha = \sigma_e^2/\sigma_a^2$  y  $G$  es una matriz de relaciones genómicas aditivas (VanRaden et al., 2009).

Para el análisis univariado de PD, el modelo incluyó los efectos genéticos aditivos maternos. La heredabilidad se obtuvo como la proporción de los componentes de varianza genética aditiva directa sobre la varianza fenotípica para las tres características analizadas. Adicionalmente, para PD se obtuvo la heredabilidad materna como la proporción de la varianza genética aditiva materna sobre la varianza fenotípica.

El modelo para predecir los VGn de los análisis bivariados fue:

$$y_i = X_i\beta_i + Z_iu_i + e_i,$$

que en notación matricial expandida corresponde a:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

donde  $y_i$  es el vector de registros fenotípicos de la  $i$ -ésima característica de crecimiento;  $X_i$  y  $Z_i$  son matrices de incidencia que relacionan las mediciones fenotípicas con los efectos sistemáticos fijos,  $\beta_i$ , y con los efectos aleatorios genéticos directos,  $u_i$ , en la  $i$ -ésima característica.

Los vectores de efectos aleatorios, se asume que siguen una densidad normal multivariada, centrada en cero, y con estructura de (co)varianzas  $Cov(u, u') = G_0 \otimes A$ ,  $Cov(e, e') = I_u \otimes R_0$  y  $Cov(u, e') = 0$ , donde  $I$  es una matriz identidad de tamaño  $u$ ,  $G_0$  es una matriz de covarianzas  $i \times i$  del efecto genético aditivo directo de las diferentes características,  $\otimes$  denota el producto Kronecker de matrices, y  $A = \{a(i, i')\}$  es una matriz de relación de numerador  $u \times u$ .

Los análisis bivariados se realizaron utilizando simultáneamente los paquetes computacionales BGLR y MTM (de los Campos & Grüneberg, 2015), los cuales están basados en el método GBLUP. Debido a restricciones en el paquete MTM, cuando los análisis bivariados incluyeron PD, sólo se consideraron los efectos genéticos aditivos directos para esta característica. En los apéndices del 4 al 6 se presentan los códigos en R para los análisis univariados y su validación cruzada. En los apéndices del 7 al 9 se presentan los códigos en R para los análisis bivariados, así como su validación cruzada.

Cuadro 1 Estadísticos descriptivos y número de observaciones (n) utilizados en los análisis genómicos para los pesos al nacimiento (PN), al destete (PD) y al año (PA) de bovinos Suizo Europeo.

Característica	Mínimo	Máximo	$\bar{X} \pm \sigma$	CV <sup>z</sup> (%)	n
PN <sup>1</sup> (kg)	31	50	38.15±4.76	12.48	95
PD <sup>2</sup> (kg)	135	308	207.62±28.42	13.69	95
PA <sup>3</sup> (kg)	177	420	314.74±47.36	14.91	95

<sup>1</sup>Los pre-ajustes fueron por sexo y GC como efectos fijos y edad de la vaca como covariable lineal y cuadrática. <sup>2</sup>Los pre-ajustes fueron por sexo, GC y régimen alimenticio al destete como efectos fijos y edad de la vaca y edad al destete como covariables lineal y cuadrática. <sup>3</sup>Los pre-ajustes fueron por sexo, GC y régimen alimenticio al año como efectos fijos y edad a la pesada al año como covariables lineal y cuadrática.

#### 3.4.4. Criterios de comparación de los análisis

Las alternativas de análisis uni- y bivariados fueron comparadas siguiendo los criterios de comparación:

**Jerarquización de los valores genómicos predichos.** Para determinar el grado de similitud entre las clasificaciones o jerarquización de animales por sus valores genómicos, se obtuvieron los estimadores de correlación Spearman entre los valores genómicos (VGn) obtenidos con los análisis univariados con los correspondientes VGn en las mismas características provenientes de las dos alternativas de análisis bivariados. Adicionalmente, se estimaron los coeficientes de correlación Pearson entre los valores genómicos de los análisis univariados y sus correspondientes provenientes de análisis bivariados.

**Habilidad de predicción de los modelos.** Para evaluar la habilidad predictiva de los modelos uni y bivariados se realizó validación cruzada. Para ello se dividió la información en conjuntos de entrenamiento y validación, utilizando el conjunto de validación para evaluar la capacidad predictiva del modelo entrenado. En los

análisis univariados se eliminó aleatoriamente 20% de los valores fenotípicos de las características de crecimiento PD, PD y PA. Posteriormente se obtuvieron las predicciones de los VGn para dichas características y se obtuvo el promedio de cinco repeticiones de los coeficientes de correlación de los fenotipos eliminados con los predichos. En los análisis bivariados se eliminó 20% de los registros fenotípicos de una característica y se obtuvo el promedio de cinco repeticiones de los coeficientes de correlación entre los VGn predichos con la información fenotípica eliminada con los predichos con la información completa.

**Coefficiente de regresión lineal simple.** Los estimadores de los coeficientes de regresión lineal simple fueron utilizados para determinar los cambios en magnitud de los valores genómicos, primero se utilizaron como variables independientes los valores genómicos predichos mediante análisis univariados y como variables dependientes los valores genómicos predichos con los análisis bivariados. Posteriormente se invirtieron las variables dependientes e independientes y se estimaron nuevamente los estimadores de los coeficientes de regresión. La hipótesis nula probada en ambos casos fue que el estimador del coeficiente de regresión fue igual que cero.

### 3.5. Resultados y discusión

De los Campos, Sorensen y Gianola (2015) definen heredabilidad genómica como la proporción de la varianza fenotípica de una característica que puede explicarse en la población bajo estudio mediante regresión lineal sobre un conjunto de marcadores moleculares. Aunque los parámetros genéticos generalmente se definen basándose en los supuestos estadísticos del modelo instrumental para el análisis de la información (de los Campos et al., 2015). En el Cuadro 2 se presentan los estimadores de los componentes de varianza y heredabilidad genómica obtenidos con resultados de análisis univariados. Los estimadores de heredabilidad directa de PN, PD y PA fueron mayores que los reportados en literatura para estas características obtenidos con análisis convencionales (fenotipos y pedigrí), 0.37, 0.27 y  $0.36 \pm 0.02$ , respectivamente, (Ríos-Utrera, 2008; Torres-Vázquez & Spangler,

2016). Ryu y Lee (2014) reportaron estimadores para heredabilidad genómica de pesos a 6, 9, 12, 15, 18, 21 y 24 meses de ganado bovino Coreano, los cuales variaron de 0.58 a 0.76, similares a los obtenidos en el presente estudio. Estos autores observaron una tendencia a mayores estimadores para este parámetro para los pesos a menor edad, sugiriendo que una evaluación genómica para características de crecimiento en bovinos a edad temprana incrementaría el mérito genético de los animales. Por su parte, el estimador de heredabilidad materna para PD fue moderado, similar a lo reportado en literatura, 0.17, con modelos lineales de efectos mixtos (Ríos-Utrera, 2008). En la literatura no se encontraron estudios en los que se reporten estimadores genómicos para heredabilidad materna.

Cuadro 2. Estimadores de componentes de varianza genética aditiva ( $VG_A$ ), varianza genética aditiva materna ( $VG_M$ ), varianza fenotípica ( $VF$ ), heredabilidad directa ( $h^2_D$ ) y materna ( $h^2_M$ ) obtenidos mediante análisis univariados para los pesos al nacimiento (PN), al destete (PD) y al año (PA) en bovinos Suizo Europeo.

	PN	PD	PA
Varianza genética aditiva	8.82	228.35	669.99
Varianza genética materna		73.67	
Varianza fenotípica	12.34	357.63	1067.52
Heredabilidad directa	0.72	0.65	0.64
Heredabilidad materna		0.21	

### 3.5.1. Jerarquización de los valores genómicos predichos

Los estimadores de los coeficientes de correlación de Spearman y Pearson entre los VGn predichos con análisis univariados y bivariados se muestran en el Cuadro 3. Estos estimadores variaron de 0.845 a 0.994 ( $p \leq 0.0001$ ) y de 0.843 a 0.989 ( $p \leq 0.0001$ ), para las correlaciones Pearson y Spearman, respectivamente. Estos estimadores indican un alto grado de asociación entre los VGn predichos con análisis uni- y bivariados; así como pocos cambios importantes en la jerarquización

de los VGn de los animales predichos con las alternativas de análisis comparadas. Estos resultados sugieren que con fines de seleccionar animales como progenitores se puede utilizar una u otra alternativa de análisis.

Los resultados del presente estudio son similares a los observados por Verbyla et al. (2010), en un estudio de simulación en el que compararon cuatro métodos bayesianos: Bayes BLUP, Bayes A, Bayes A/B y Bayes C, para predecir VGn de individuos sin información fenotípica, asumiendo diferentes distribuciones *a priori*. Los estimadores de los coeficientes de correlación Spearman para 100 animales, variaron de 0.69 a 0.73 para los cuatro métodos evaluados. Estos resultados sugieren que la jerarquización de los VGn para los animales evaluados cambia poco, independientemente del método utilizado para predecir los VGn. Los autores concluyeron que los modelos no parecen sensibles a la elección de las distribuciones *a priori*, consistente con los resultados de los métodos con información real. En otro estudio, Zambrano, Rincón, López y Echeverri (2015) estimaron los coeficientes de la correlación Spearman entre los VGn obtenidos mediante el método Bayes C y los EBVs predichos convencionalmente mediante el mejor predictor lineal insesgado para producción de leche, porcentajes de proteína y grasa en leche, y conteo de células somáticas en bovinos Holstein. Los estimadores obtenidos por estos autores variaron de  $0.932 \pm 0.02$  a  $0.991 \pm 0.0005$  ( $p \leq 0.01$ ). En un estudio realizado con bovinos Holstein Friesian en Australia, Moser, Tier, Crump, Khatkar y Raadsma (2009) encontraron valores de 0.26 a 0.68 para los estimadores de los coeficientes de correlación entre valores genéticos predichos y valores genómicos, para porcentaje de proteína y un índice de selección australiano.

Cuadro 3. Estimadores de los coeficientes de correlación Pearson y Spearman, en paréntesis nivel de significancia para la prueba de hipótesis del estimador igual a cero, entre los valores genómicos predichos con análisis uni- y bivariados para pesos al nacimiento, al destete y al año en bovinos Suizo Europeo.

	PNBPDD <sup>1</sup>	PNBPA	PDDBPBPN	PDDBPBPA	PABPN	PABPDD
Estimadores de correlación Pearson						
PNU	0.994 (<.0001)	0.994 (<.0001)				
PDDU			0.989 (<.0001)	0.988 (<.0001)		
PAU					0.978 (<.0001)	0.845 (<.0001)
Estimadores de correlación Spearman						
PNU	0.987 (<.0001)	0.989 (<.0001)				
PDDU			0.985 (<.0001)	0.983 (<.0001)		
PAU					0.977 (<.0001)	0.843 (<.0001)

<sup>1</sup> PNU = peso al nacimiento univariado; PDDU = peso al destete directo univariado, PAU = peso al año univariado; PNBPDD = peso al nacimiento bivariado con peso al destete directo; PNBPA = peso al nacimiento bivariado con peso al año; PDDBPBPN = peso al destete directo bivariado con peso al nacimiento; PDDBPBPA = peso al destete directo bivariado con peso al año; PABPN = peso al año bivariado con peso al nacimiento; PABPDD = peso al año bivariado con peso al destete directo.

### 3.5.2. Habilidad de predicción de los modelos

En la Figura 1 se muestran los promedios de las correlaciones de las validaciones cruzadas de los VGn de modelos univariados y bivariados. En general puede observarse que los análisis bivariados tuvieron mejor habilidad predictiva que los univariados; con excepción de los análisis bivariados de PN con PD y PA. Lo anterior corresponde al mayor tiempo que transcurre entre los registros de las dos características. Por el contrario, los mayores estimadores de correlación corresponden a VGn de características registradas más cercana la una de la otra en la vida del animal.

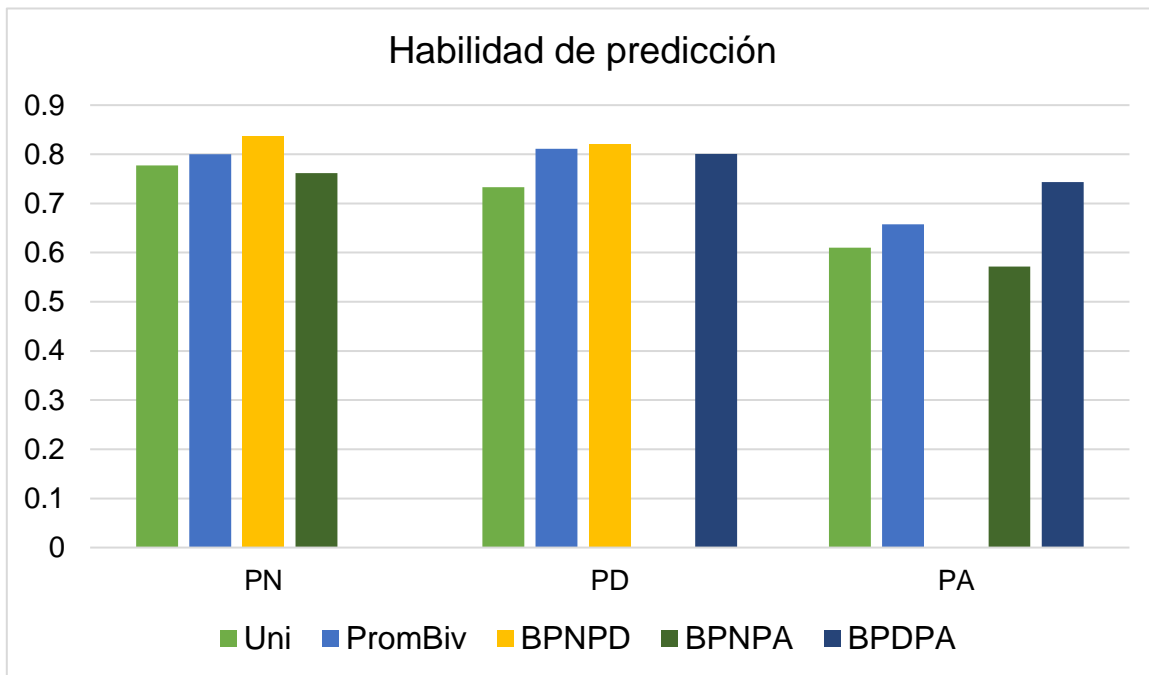


Figura 1. Promedios de los estimadores de correlación Pearson entre los valores genómicos predichos mediante análisis univariados (Uni) con los valores genómicos predichos promedio (PromBiv) de análisis bivariados (BPNPD, BPNPA y BPDPA) para los pesos al nacimiento (PN), al destete (PD) y al año (PA), de cinco repeticiones de validación cruzada.

En un estudio realizado en bovinos Angus y Tropicarne, Ramírez-Valverde et al. (2007) estimaron coeficientes de correlación entre valores genéticos predichos obtenidos de pares de muestras usando validación cruzada en bovinos Angus y Tropicarne para leche, PN, PD y PA. Los estimadores variaron de 0.12 a 0.24 y de



0.11 a 1.0 en Angus, de 0.23 a 0.51 y de 0.53 a 0.91 en Tropicarne, en los análisis univariados y bivariados, respectivamente. Estos autores también reportaron promedios de 0.00 a 0.12 y de 0.02 a 0.19 en Angus; de 0.42 a 0.48 y de 0.39 a 0.54 en Tropicarne para exactitud de valores genéticos predichos para leche, PN, PD y PA con modelos uni- y bivariados, respectivamente. Los resultados sugieren ventajas de los análisis bivariados. Estos autores realizaron ambos tipos de análisis mediante el modelo animal.

Estudios uni- y multivariados también se han realizado recientemente en plantas. Por ejemplo, Wang et al. (2017) predijeron VGn para ocho características de rendimiento en híbridos de arroz utilizando GBLUP uni- y multivariados. Estos autores utilizaron validación cruzada para investigar la capacidad predictiva de los modelos. La habilidad predictiva de los modelos multivariados fue mayor que la de los univariados. Sin embargo, estos autores señalan que los análisis univariados son suficientes cuando se analizan características con alta heredabilidad. La ventaja de los análisis multivariados está en utilizarlos para características de baja heredabilidad. Por el contrario, Dos Santos, Vasconcellos, Pires, Balestre y Von Pinho (2016) no encontraron diferencias significativas en capacidad predictiva entre modelos uni- y bivariados en varios escenarios de evaluación para cinco características en maíz. La principal explicación que estos autores esgrimen para estas pequeñas diferencias, es la baja a moderada magnitud de las correlaciones entre las características estudiadas y las moderadas heredabilidades. Estas condiciones no favorecen el uso de las ventajas de los análisis multivariados. Cuando estos autores incluyeron los efectos de dominancia en los modelos, la capacidad predictiva de los modelos mejoró, así como la calidad en la estimación de los componentes de varianza.

### 3.5.3. Estimadores de los coeficientes de regresión Pearson

Los estimadores de los coeficientes de regresión Pearson entre VGn provenientes de análisis univariados y VGn provenientes de análisis bivariados para las características analizadas se presentan en el Cuadro 4. Los estimadores fueron obtenidos utilizando como variable independiente los VGn obtenidos con los análisis univariados y como variable dependiente los VGn predichos con los análisis bivariados, y viceversa. Cuando las variables independientes fueron los VGn predichos con análisis univariados, los estimadores de las regresiones variaron de  $0.96 \pm 0.06$  a  $1.37 \pm 0.06$  ( $p \leq 0.0001$ ); mientras que cuando lo fueron los VGn predichos con los análisis bivariados, los estimadores variaron de  $0.70 \pm 0.02$  a  $0.82 \pm 0.01$  ( $p \leq 0.0001$ ). Estos resultados indican que los VGn predichos con una alternativa de análisis se pueden utilizar para predecir los VGn con la otra alternativa de análisis. Estos resultados fueron similares a los observados por Zambrano et al. (2016), quienes al estimar los coeficientes de regresión entre EBV predichos con el método BLUP sobre VGn predichos mediante el método Bayes C, obtuvieron de 0.954 a 1.051 para cuatro características de producción en bovinos Holstein. En la medida que los estimadores de los coeficientes de regresión sean diferentes de la unidad, el sesgo en la predicción de los VGn con las alternativas de análisis en comparación es mayor (Mäntysaari et al., 2010).

Por su parte Ryu y Lee (2014) estimaron coeficientes de regresión para probar la asociación de SNPs con pesos corporales a diferentes edades. Estos autores obtuvieron estimadores para estos coeficientes de 0.66 a 0.98. Estos estimadores decrecieron cuando las diferencias en edad fueron mayores

Cuadro 4. Estimadores de los coeficientes de regresión ( $\pm$ error estándar), utilizando como variable independiente los valores genómicos predichos mediante análisis univariados y como variable dependiente los valores genómicos predichos con los análisis bivariados y viceversa; en paréntesis nivel de significancia para la prueba de hipótesis del estimador igual a cero, para pesos al nacimiento, al destete directo y al año en bovinos Suizo Europeo.

	PNBPDD	PNBPA	PNU	PDDDBPN	PDDDBPA	PDU	PABPN	PABPDD	PAU
PNU <sup>1</sup>	0.80 $\pm$ 0.01 ( $<0.0001$ )	0.82 $\pm$ 0.01 ( $<0.0001$ )							
PNBPDD			1.21 $\pm$ 0.01 ( $<0.0001$ )						
PNBPA			1.21 $\pm$ 0.01 ( $<0.0001$ )						
PDDU				0.78 $\pm$ 0.01 ( $<0.0001$ )	0.78 $\pm$ 0.01 ( $<0.0001$ )				
PDDDBPN						1.26 $\pm$ 0.02 ( $<0.0001$ )			
PDDDBPA						1.25 $\pm$ 0.02 ( $<0.0001$ )			
PAU							0.70 $\pm$ 0.02 ( $<0.0001$ )	0.74 $\pm$ 0.05 ( $<0.0001$ )	
PABPN									1.37 $\pm$ 0.03 ( $<0.0001$ )
PABPDD									0.96 $\pm$ 0.06 ( $<0.0001$ )

<sup>1</sup> PNU = valor genómico para peso al nacimiento (PN) predicho con el análisis univariado, PNBPDD = valor genómico predicho con análisis bivariado de PN y peso al destete directo (PDD), PNBPA = valor genómico predicho con análisis bivariado de PN y peso al año (PA), PDDU = valor genómico predicho con análisis univariado de PD (efectos directos, el modelo consideró efectos directos y maternos), PDDDBPN = valores genómicos para PD (efectos directos) predichos con análisis bivariado de PDD y PN, PAU = valor genómico para PA predicho con análisis univariado, PABPN = valor genómico para PA predicho con análisis bivariado con PN, PABPDD = valor genómico para PA predicho con análisis bivariado con PDD.

A través de estudios de simulación, Basu et al. (2013) compararon análisis uni- vs multivariados a través de regresión lineal múltiple multivariada, encontrando que con el enfoque multivariado se mantuvo un nivel de error tipo I correcto, aun cuando se tuvo conjuntos de SNPs en fuerte desequilibrio de ligamiento (LD). Estos autores también observaron una ganancia sustancial en potencia para detectar un gene cuando está asociado con un subconjunto de características.

De acuerdo con las recomendaciones de varios autores, entre ellos Calus (2010), dado el reducido número de animales que conformaron la población de referencia, los resultados obtenidos en el presente trabajo deberán tomarse con precaución. En el futuro, deberán realizarse estudios considerando un número mayor de animales genotipados y por varias generaciones. Esto con la finalidad de estimar los efectos de los marcadores dado el rompimiento del LD entre ellos y los QTL a través de generaciones para mantener niveles adecuados de exactitud de los VGn (Calus, 2010).

### **3.6. Conclusiones**

La jerarquización de los valores genómicos predichos con modelos univariados es similar a la obtenida con modelos bivariados.

La habilidad predictiva de los modelos bivariados es mayor que la de los modelos univariados, especialmente para las características registradas más cerca la una de la otra.

Los valores genómicos predichos con análisis univariados pueden utilizarse para predecir los valores genómicos de análisis bivariados y viceversa.

### **3.7. Literatura citada**

Basu, S., Zhang, Y., Ray, D., Miller, M. B., Iacono, W.G., & McGueM, M. (2013). A rapid genome-wide gene-based association test with multivariate trait. *Human*

*Heredity*, 76(2), 53-63, doi:10.1159/000356016.

Calus, M. P. L. (2010). Genomic breeding value prediction: Methods and procedures. *Animal*, 4(2), 157-164, doi:10.1017/S1751731109991352.

de los campos, G., & Grüneberg. (2015). MTM: Multivariate gaussian models for genomic and pedigree data. R package. Michigan State University. Available at: <https://github.com/QuantGen/MTM/wiki>. Consulted on December 15th, 2017.

de los Campos, G., Sorensen, D., & Gianola, D. (2015). Genomic heritability: What is it? *PLoS Genet*, 11(5), e1005048, doi:10.1371/journal.pgen.100504.

de los Campos, G., & Pérez-Rodríguez, P. (2016). BGLR: Bayesian generalized linear regression. R package version 1.0.5. R Foundation for Statistical Computing. <https://cran.r-project.org/web/packages/BGLR/index>. Accessed Dec 15th, 2017.

Dos Santos, J. P. R., Vasconcellos, R. C. de C., Pires, L. P. M., Balestre, M., Von Pinho, R. G. (2016). Inclusion of dominance effects in the multivariate GBLUP Model. *PLoS ONE*, 11(4), e0152045, doi:10.1371/journal.pone.0152045.

Henderson, C. R. (1984) Applications of linear models in animal breeding. U. Guelph, ON. Available at: <http://cgil.uoguelph.ca/pub/Henderson.html>. Consulted on December 17<sup>th</sup>, 2017.

Henderson, C. R., & Quaas, R. L. (1976). Multiple trait evaluation using relatives' records. *Journal of Animal Science*, 43(6), 1188-1197.

Mäntysaari, E., Zengting, L., & Van Raden, P. (2010). Interbull validation test for genomics evaluations. *Interbull Bulletin*, 41, 17-22.

Moser, G., Tier, B., Crump, R. E., Khatkar M. S., & Raadsma H. W. (2009). A comparison of five methods to predict genomic breeding values of dairy bulls

- from genome-wide SNP markers. *Genetics Selection Evolution*, 41, 56-61, <https://doi.org/10.1186/1297-9686-41-56>
- Pérez, P., & de los Campos, G. (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genetics*, 198, 483-495.
- R Core Team. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Available at: <https://www.R-project.org/> Accessed on November 12<sup>th</sup>, 2017.
- Ramírez-Valverde, R., Hernández-Alvarez, C. O., Núñez-Domínguez, R., Ruíz-Flores, A., & García-Muñiz, J. G. (2007). Univariate vs multivariate analyses in genetic evaluation of growth traits in two cattle breeds. *Agrociencia*, 41, 271–282.
- Ríos-Utrera., A. (2008). Estimadores de parámetros genéticos para características de crecimiento predestete de bovinos. Revisión. *Técnica Pecuaria en México*, 46(1), 37-67.
- Ryu, J., & Lee, L. (2014). Genomic heritability of bovine growth using a mixed model. *Asian-Australasian Journal of Animal Science*, 27(11), 1521–1525, doi:10.5713/ajas.2014.14287.
- Serão, N. V. L., González-Peña, D., Beever, J. E., Bollero, G. A., Southey, B. R., Faulkner, D. B., & Rodriguez-Zas, S. L. (2013). Bivariate genome-wide association analysis of the growth and intake components of feed efficiency. *PLoS ONE*, 8(10), e78530, doi:10.1371/journal.pone.0078530.
- SMN (Servicio Meteorológico Nacional). (2017). Información climatológica. Disponible en: <http://smn.cna.gob.mx/es/climatologia/informacion-climatologica>. Consultada el 30 de agosto de 2017.
- Torres-Vázquez, J. A., & Spangler, M. L. (2016). Genetic parameters for docility, weaning weight, yearling weight, and intramuscular fat percentage in

- Hereford cattle. *Journal of Animal Science*, 94, 21–27, doi:10.2527/jas2015-9566.
- Van der Sluis, S., Dolan, C. V., Li, J., Song, Y., Sham, P., Posthuma, D., & Li, M.-L. (2015). MGAS: a powerful tool for multivariate gene-based genome-wide association analysis. *Bioinformatics*, 31(7), 1007–1015, doi: 10.1093/bioinformatics/btu783.
- VanRaden, P. M., Van Tassell, C. P., Wiggans, G. R., Sonstegard, T. S., Schnabel, R. D., Taylor, J. F., & Schenkel, F. S. (2009). Invited review: Reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science*, 92, 16-24.
- Verbyla, K. L., Bowman, P. J., Hayes, B. J., Raadsma, H., & Goddard, M. E. (2010). Sensitivity of genomic selection to using different prior distributions. *BMC Proceedings*, 4(Suppl 1), S5.
- Wang, X., Li, L., Yang, Z., Zheng, X., Yu, S, Xu, C., & Hu, Z. (2017). Predicting rice hybrid performance using univariate and multivariate GBLUP models based on North Carolina mating design II. *Heredity*, 118, 302–310, doi:10.1038/hdy.2016.87.
- Zambrano A., J., Rincón F., J., López H., A., y Echeverri Z., J. (2015). Estimation and comparison of conventional and genomic breeding values in Holstein cattle of Antioquia, Colombia. *Rev. MVZ Córdoba*, 20(3), 4739-475.

## 3.8 Apéndices

### Apéndice 1. Código en R para el ajuste de PN

```
library(Matrix)
library(lme4)

#Código para llamar el archivo donde se encuentran los datos
setwd("C:/Users/jev_h_/Documents/Maestría/Genética/Análisis de
datos/Univariados/UniPN/Ajuste PN")
SuizoE=read.csv(file="Ajuste de PN.csv", header=TRUE)
fix(SuizoE)

#Estadísticos descriptivos para variables de interés
length(SuizoE$BW)
min(SuizoE$BW)
max(SuizoE$BW)
mean(SuizoE$BW)
sd(SuizoE$BW)

#Distribución de frecuencias para BW
hist(SuizoE$BW,main="Distribución de frecuencias para BW",
     ylab="Frecuencia", xlab="BW(Bird Weigth (kg))", col="green")

#Modelo YW=Sex+GC+EM+EM^2 Fijos(Sex,EM,EM2) Aleatorios(GC)
AJUSTEBW<-lmer(BW~as.factor(Sex)+(1|GC)+EM+EM2, data=SuizoE)
summary(AJUSTEBW)
YHAT<-predict(AJUSTEBW)

#Exportar Resultados
library(XLConnectJars)
library(XLConnect)
PAAj<-loadWorkbook("PAAj.xlsx", create=TRUE)
createSheet(PAAj, name="Ajustado")
createName(PAAj, name="Ajustado", formula="Ajustado!$A$1")
writeNamedRegion(PAAj, YHAT, name="Ajustado")
saveWorkbook(PAAj)
```



## Apéndice 2. Código en R para el ajuste de PD

```
library(Matrix)
library(lme4)

#Código para llamar el archivo donde se encuentran los datos
setwd("C:/Users/jev_h_/Documents/Maestría/Genética/Análisis de
datos/Univariados/UniPD/Ajuste PD")
SuizoE=read.csv(file="Ajuste de PD.csv", header=TRUE)
fix(SuizoE)

#Estadísticos descriptivos para variables de interés
length(SuizoE$WW)
min(SuizoE$WW)
max(SuizoE$WW)
mean(SuizoE$WW)
sd(SuizoE$WW)

#Distribución de frecuencias para BW
hist(SuizoE$WW,main="Distribución de frecuencias para WW",
      ylab="Frecuencia", xlab="WW(Weaning Weigh (kg))", col="green")

#Modelo YW=Sex+GC+Man+EM+EM^2+ED+ED2 Fijos(Sex,Man,EA,EA2)
Aleatorios(GC)
AJUSTEWW<-lmer(WW~as.factor(Sex)+(1|GC)+as.factor(Man)+EM+EM2+DED+DED2,
data=SuizoE)
summary(AJUSTEWW)
YHAT<-predict(AJUSTEWW)

#Exportar Resultados
library(XLConnectJars)
library(XLConnect)
PDAj<-loadWorkbook("PDAj.xlsx", create=TRUE)
createSheet(PDAj, name="Ajustado")
createName(PDAj, name="Ajustado", formula="Ajustado!$A$1")
writeNamedRegion(PDAj, YHAT, name="Ajustado")
saveWorkbook(PDAj)
```

### Apéndice 3. Código en R para el ajuste de PA

```
rm(list=ls())
library(Matrix)
library(lme4)

#Código para llamar el archivo donde se encuentran los datos
setwd("C:/Users/jev_h_/Documents/Maestría/Genética/Análisis de
datos/Univariados/UniPA/Ajuste PA")
SuizoE=read.csv(file="Ajuste de PA.csv", header=TRUE)
fix(SuizoE)

#Estadísticos descriptivos para variables de interés
length(SuizoE$YW)
min(SuizoE$YW)
max(SuizoE$YW)
mean(SuizoE$YW)
sd(SuizoE$YW)

#Distribución de frecuencias para BW
hist(SuizoE$YW,main="Distribución de frecuencias para YW",
     ylab="Frecuencia", xlab="YW(Year Weigth (kg))", col="green")

#Modelo  $YW = \text{Sex} + \text{GC} + \text{Man} + \text{EA} + \text{EA}^2$  Fijos(Sex,Man,EA,EA2) Aleatorios(GC)
AJUSTEYW<-lmer(YW~as.factor(Sex)+(1|GC)+as.factor(Man)+DEA+DEA2, data=SuizoE)
summary(AJUSTEYW)
YHAT<-predict(AJUSTEYW)

#Exportar Resultados
library(XLConnectJars)
library(XLConnect)
PAAj<-loadWorkbook("PAAj.xlsx", create=TRUE)
createSheet(PAAj, name="Ajustado")
createName(PAAj, name="Ajustado", formula="Ajustado!$A$1")
writeNamedRegion(PAAj, YHAT, name="Ajustado")
saveWorkbook(PAAj)
```

#### Apéndice 4. Código en R para análisis univariado de PN y su validación cruzada

```
rm(list=ls())
library(BGLR)
setwd("C:/Users/jev_h_/Documents/Maestría/Genética/Análisis de
datos/Univariados/UniPN/GBLUP PN")

MatrizX<-load("C:/Users/jev_h_/Documents/Maestría/Genética/Análisis de
datos/Univariados/UniPN/GBLUP PN/Matriz X.RData")

FenPN=read.csv(file="FenPN.csv",header=TRUE)
dim(FenPN)
UNY<-as.matrix(FenPN)
class(Y)

#Genomic Relationship matrix
Z=scale(X, center=TRUE,scale=TRUE)
G=tcrossprod(Z)/ncol(Z)
dim(G)

ids <- as.character(FenPN$ID)
idsall<-colnames(G)
idsall<-rownames(G)
indice=idsall%in%ids
G1<-G[indice,]
G2<-G1[,indice]
dim(G2)
rownames(G2)

#Linear predictor
ETA=list(list(K=G2, model="RKHS"))
fUN<-BGLR(y=as.numeric(UNY[,4]), ETA=ETA, nIter=10000, burnIn=5000, thin=10)

cor(fUN$yHat,as.numeric(UNY[,4]))

##### Validación cruzada #####

### Creación de poblaciones de entrenamiento y prueba ###
Ind.sets=sample(rep(seq(1,5,by=1),19),95,replace=FALSE)
Ind.sets

A=which(Ind.sets==1)
B=which(Ind.sets==2)
C=which(Ind.sets==3)
D=which(Ind.sets==4)
E=which(Ind.sets==5)

sets<-Ind.sets
```

```
##### Grupo A
UNy<-UNY[,4]
UNyNa=UNy
UNyNa[A]=NA
#Predictor Lineal
ETA=list(list(K=G2, model="RKHS"))
fUNA<-BGLR(y=as.numeric(UNyNa), ETA=ETA, nIter=10000, burnIn=5000, thin=10)
```

```
##### Grupo B
UNy<-UNY[,4]
UNyNa=UNy
UNyNa[B]=NA
#Predictor Lineal
ETA=list(list(K=G2, model="RKHS"))
fUNB<-BGLR(y=as.numeric(UNyNa), ETA=ETA, nIter=10000, burnIn=5000, thin=10)
```

```
##### Grupo C
UNy<-UNY[,4]
UNyNa=UNy
UNyNa[C]=NA
#Predictor Lineal
ETA=list(list(K=G2, model="RKHS"))
fUNC<-BGLR(y=as.numeric(UNyNa), ETA=ETA, nIter=10000, burnIn=5000, thin=10)
```

```
##### Grupo D
UNy<-UNY[,4]
UNyNa=UNy
UNyNa[D]=NA
#Predictor Lineal
ETA=list(list(K=G2, model="RKHS"))
fUND<-BGLR(y=as.numeric(UNyNa), ETA=ETA, nIter=10000, burnIn=5000, thin=10)
```

```
##### Grupo E
UNy<-UNY[,4]
UNyNa=UNy
UNyNa[E]=NA
#Predictor Lineal
ETA=list(list(K=G2, model="RKHS"))
fUNE<-BGLR(y=as.numeric(UNyNa), ETA=ETA, nIter=10000, burnIn=5000, thin=10)
```

```
CorVC<-c(cor(fUNA$yHat[A],as.numeric(UNy[A])),
          cor(fUNB$yHat[B],as.numeric(UNy[B])),
          cor(fUNC$yHat[C],as.numeric(UNy[C])),
          cor(fUND$yHat[D],as.numeric(UNy[D])),
          cor(fUNE$yHat[E],as.numeric(UNy[E])))
```

```
CorVC
```

## Apéndice 5. Código en R para análisis univariado de PD y su validación cruzada

```
rm(list=ls())
library(BGLR)
setwd("C:/Users/jev_h_/Documents/Maestría/Genética/Análisis de
datos/Univariados/UniPD/GBLUP PD")

MatrizX<-load("C:/Users/jev_h_/Documents/Maestría/Genética/Análisis de
datos/Univariados/UniPD/GBLUP PD/Matriz X.RData")

FenPD=read.csv(file="FenPD.csv",header=TRUE)
dim(FenPD)
UDY<-as.matrix(FenPD)
class(Y)

#Genomic Relationship matrix
Z=scale (X, center=TRUE,scale=TRUE)
G=tcrossprod(Z)/ncol(Z)
dim(G)

ids <- as.character(FenPD$ID)
idsall<-colnames(G)
idsall<-rownames(G)
indice=idsall%in%ids
G1<-G[indice,]
G2<-G1[,indice]
dim(G2)
rownames(G2)

#Linear predictor
ETA=list(list(K=G2, model="RKHS"))
fUD<-BGLR(y=as.numeric(UDY[,4]), ETA=ETA, nlter=10000, burnIn=5000, thin=10)

cor(fUD$yHat,as.numeric(UDY[,4]))

##### Validación cruzada #####

### Creación de poblaciones de entrenamiento y prueba ###
Ind.sets=sample(rep(seq(1,5,by=1),19),95,replace=FALSE)
Ind.sets

A=which(Ind.sets==1)
B=which(Ind.sets==2)
C=which(Ind.sets==3)
D=which(Ind.sets==4)
E=which(Ind.sets==5)
```

```

sets<-Ind.sets

##### Grupo A
UDy<-UDY[,4]
UDyNa=UDy
UDyNa[A]=NA
#Predictor Lineal
ETA=list(list(K=G2, model="RKHS"))
fUDA<-BGLR(y=as.numeric(UDyNa), ETA=ETA, nIter=10000, burnIn=5000, thin=10)

##### Grupo B
UDy<-UDY[,4]
UDyNa=UDy
UDyNa[B]=NA
#Predictor Lineal
ETA=list(list(K=G2, model="RKHS"))
fUDB<-BGLR(y=as.numeric(UDyNa), ETA=ETA, nIter=10000, burnIn=5000, thin=10)

##### Grupo C
UDy<-UDY[,4]
UDyNa=UDy
UDyNa[C]=NA
#Predictor Lineal
ETA=list(list(K=G2, model="RKHS"))
fUDC<-BGLR(y=as.numeric(UDyNa), ETA=ETA, nIter=10000, burnIn=5000, thin=10)

##### Grupo D
UDy<-UDY[,4]
UDyNa=UDy
UDyNa[D]=NA
#Predictor Lineal
ETA=list(list(K=G2, model="RKHS"))
fUDD<-BGLR(y=as.numeric(UDyNa), ETA=ETA, nIter=10000, burnIn=5000, thin=10)

##### Grupo E
UDy<-UDY[,4]
UDyNa=UDy
UDyNa[E]=NA
#Predictor Lineal
ETA=list(list(K=G2, model="RKHS"))
fUDE<-BGLR(y=as.numeric(UDyNa), ETA=ETA, nIter=10000, burnIn=5000, thin=10)

CorVC<-c(cor(fUDA$yHat[A],as.numeric(UDy[A])),
          cor(fUDB$yHat[B],as.numeric(UDy[B])),
          cor(fUDC$yHat[C],as.numeric(UDy[C])),
          cor(fUDD$yHat[D],as.numeric(UDy[D])),
          cor(fUDE$yHat[E],as.numeric(UDy[E])))

CorVC

```

## Apéndice 6. Código en R para análisis univariado de PA y su validación cruzada

```
rm(list=ls())
library(BGLR)
setwd("C:/Users/jev_h_/Documents/Maestría/Genética/Análisis de
datos/Univariados/UniPA/GBLUP PA")

MatrizX<-load("C:/Users/jev_h_/Documents/Maestría/Genética/Análisis de
datos/Univariados/UniPA/GBLUP PA/Matriz X.RData")

FenPA=read.csv(file="FenPA.csv",header=TRUE)
dim(FenPA)
UAY<-as.matrix(FenPA)
class(Y)

#Genomic Relationship matrix
Z=scale(X, center=TRUE,scale=TRUE)
G=tcrossprod(Z)/ncol(Z)
dim(G)

ids <- as.character(FenPA$ID)
idsall<-colnames(G)
idsall<-rownames(G)
indice=idsall%in%ids
G1<-G[indice,]
G2<-G1[,indice]
dim(G2)
rownames(G2)

#Linear predictor
ETA=list(list(K=G2, model="RKHS"))
fUA<-BGLR(y=as.numeric(UAY[,4]), ETA=ETA, nIter=10000, burnIn=5000, thin=10)

cor(fUA$yHat,as.numeric(UAY[,4]))

##### Validación cruzada #####

### Creación de poblaciones de entrenamiento y prueba ###
Ind.sets=sample(rep(seq(1,5,by=1),19),95,replace=FALSE)
Ind.sets

A=which(Ind.sets==1)
B=which(Ind.sets==2)
C=which(Ind.sets==3)
D=which(Ind.sets==4)
E=which(Ind.sets==5)

sets<-Ind.sets
```

```
##### Grupo A
UAY<-UAY[,4]
UAYNa=UAY
UAYNa[A]=NA
#Predictor Lineal
ETA=list(list(K=G2, model="RKHS"))
fUAA<-BGLR(y=as.numeric(UAYNa), ETA=ETA, nIter=10000, burnIn=5000, thin=10)
```

```
##### Grupo B
UAY<-UAY[,4]
UAYNa=UAY
UAYNa[B]=NA
#Predictor Lineal
ETA=list(list(K=G2, model="RKHS"))
fUAB<-BGLR(y=as.numeric(UAYNa), ETA=ETA, nIter=10000, burnIn=5000, thin=10)
```

```
##### Grupo C
UAY<-UAY[,4]
UAYNa=UAY
UAYNa[C]=NA
#Predictor Lineal
ETA=list(list(K=G2, model="RKHS"))
fUAC<-BGLR(y=as.numeric(UAYNa), ETA=ETA, nIter=10000, burnIn=5000, thin=10)
```

```
##### Grupo D
UAY<-UAY[,4]
UAYNa=UAY
UAYNa[D]=NA
#Predictor Lineal
ETA=list(list(K=G2, model="RKHS"))
fUAD<-BGLR(y=as.numeric(UAYNa), ETA=ETA, nIter=10000, burnIn=5000, thin=10)
```

```
##### Grupo E
UAY<-UAY[,4]
UAYNa=UAY
UAYNa[E]=NA
#Predictor Lineal
ETA=list(list(K=G2, model="RKHS"))
fUAE<-BGLR(y=as.numeric(UAYNa), ETA=ETA, nIter=10000, burnIn=5000, thin=10)
```

```
CorVC<-c(cor(fUAA$yHat[A],as.numeric(UAY[A])),
          cor(fUAB$yHat[B],as.numeric(UAY[B])),
          cor(fUAC$yHat[C],as.numeric(UAY[C])),
          cor(fUAD$yHat[D],as.numeric(UAY[D])),
          cor(fUAE$yHat[E],as.numeric(UAY[E])))
```

```
CorVC
```



## Apéndice 7. Código en R para análisis bivariado de PN y PD y su validación cruzada

```
rm(list=ls())
library(MTM)
library(BGLR)

#Cargar carpeta de trabajo
setwd("C:/Users/jev_h_/Documents/Maestría/Genética/Análisis de
datos/Bivariados/BivPN_PD")
dir()

BNDY=read.csv(file="FenPN_PD.csv",header=TRUE)
fix(BNDY)
PNPD<-cbind(BNDY$PNAj,BNDY$PDAj)

MatrizX<-load("C:/Users/jev_h_/Documents/Maestría/Genética/Análisis de
datos/Bivariados/BivPN_PD/Matriz X.RData")

#Genomic Relationship matrix
Z=scale(X, center=TRUE,scale=TRUE)
G=tcrossprod(Z)/ncol(Z)
dim(G)

ids <- as.character(BNDY$ID)
idsall<-colnames(G)
idsall<-rownames(G)
indice=idsall%in%ids
G1<-G[indice,]
G2<-G1[,indice]
dim(G2)
rownames(G2)

##### Sin validación
fBND <- MTM(Y = PNPD,K = list(list(K = G2,COV = list(type = 'UN',df0 = 4,S0 = diag(2)))),
          resCov = list(type = 'UN',S0 = diag(2),df0 = 4),nlter = 10000,burnIn = 500,thin =
5,saveAt = 'ex1_')

cor(fBND$YHat,PNPD)

# Recuperando estimaciones
fBND$YHat # predicitons
fBND$resCov$R # Matriz de covarianzas residual
fBND$K[[1]]$G # Matriz de covarianzas geneticas
fBND$K[[1]]$U # Efectos aleatorios

cor(fBND$YHat[,1],fBND$YHat[,2])
```

```
##### Validación cruzada PN CON PD #####
```

```
### Creación de poblaciones de entrenamiento y prueba ###
```

```
Ind.sets=sample(rep(seq(1,5,by=1),19),95,replace=FALSE)
```

```
Ind.sets
```

```
A=which(Ind.sets==1)
```

```
B=which(Ind.sets==2)
```

```
C=which(Ind.sets==3)
```

```
D=which(Ind.sets==4)
```

```
E=which(Ind.sets==5)
```

```
sets<-Ind.sets
```

```
## A ##
```

```
BNDy1<-BNDY$PNAj
```

```
BNDyNa1=BNDy1
```

```
BNDyNa1[A]=NA
```

```
BNDy2<-BNDY$PDAj
```

```
BNDyNa2=BNDy2
```

```
#BNDyNa2[A]=NA
```

```
BNDy<-cbind(BNDyNa1,BNDyNa2)
```

```
fBNDa <- MTM(Y = BNDy,K = list(list(K = G2,COV = list(type = 'UN',df0 = 4,S0 =  
diag(2))))),
```

```
resCov = list(type = 'UN',S0 = diag(2),df0 = 4),nlter = 10000,burnIn = 500,thin =  
5,saveAt = 'ex1_')
```

```
## B ##
```

```
BNDy1<-BNDY$PNAj
```

```
BNDyNa1=BNDy1
```

```
BNDyNa1[B]=NA
```

```
BNDy2<-BNDY$PDAj
```

```
BNDyNa2=BNDy2
```

```
#BNDyNa2[B]=NA
```

```
BNDy<-cbind(BNDyNa1,BNDyNa2)
```

```
fBNDB <- MTM(Y = BNDy,K = list(list(K = G2,COV = list(type = 'UN',df0 = 4,S0 =  
diag(2))))),
```

```
resCov = list(type = 'UN',S0 = diag(2),df0 = 4),nlter = 10000,burnIn = 500,thin =  
5,saveAt = 'ex1_')
```

```
## C ##
```

```
BNDy1<-BNDY$PNAj
```

```
BNDyNa1=BNDy1
```

```
BNDyNa1[C]=NA
```

```
BNDy2<-BNDY$PDAj
```

```

BNDyNa2=BNDy2
#BNDyNa2[C]=NA

BNDy<-cbind(BNDyNa1,BNDyNa2)

fBNDc <- MTM(Y = BNDy,K = list(list(K = G2,COV = list(type = 'UN',df0 = 4,S0 =
diag(2)))),
      resCov = list(type = 'UN',S0 = diag(2),df0 = 4),nlter = 10000,burnIn = 500,thin =
5,saveAt = 'ex1_')

## D ##
BNDy1<-BNDY$PNAj
BNDyNa1=BNDy1
BNDyNa1[D]=NA
BNDy2<-BNDY$PDAj
BNDyNa2=BNDy2
#BNDyNa2[D]=NA

BNDy<-cbind(BNDyNa1,BNDyNa2)

fBNDd <- MTM(Y = BNDy,K = list(list(K = G2,COV = list(type = 'UN',df0 = 4,S0 =
diag(2)))),
      resCov = list(type = 'UN',S0 = diag(2),df0 = 4),nlter = 10000,burnIn = 500,thin =
5,saveAt = 'ex1_')

## E ##
BNDy1<-BNDY$PNAj
BNDyNa1=BNDy1
BNDyNa1[E]=NA
BNDy2<-BNDY$PDAj
BNDyNa2=BNDy2
#BNDyNa2[E]=NA

BNDy<-cbind(BNDyNa1,BNDyNa2)

fBNDe <- MTM(Y = BNDy,K = list(list(K = G2,COV = list(type = 'UN',df0 = 4,S0 =
diag(2)))),
      resCov = list(type = 'UN',S0 = diag(2),df0 = 4),nlter = 10000,burnIn = 500,thin =
5,saveAt = 'ex1_')

CorVCPNPD<-c(cor(fBNDa$YHat[A],[, 1],PNPD[A],[, 1]),
      cor(fBNDb$YHat[B],[, 1],PNPD[B],[, 1]),
      cor(fBNDc$YHat[C],[, 1],PNPD[C],[, 1]),
      cor(fBNDd$YHat[D],[, 1],PNPD[D],[, 1]),
      cor(fBNDe$YHat[E],[, 1],PNPD[E],[, 1]))

CorVCPNPD

```

```
##### Validación cruzada PD CON PN #####
```

```
### Creación de poblaciones de entrenamiento y prueba ###
```

```
Ind.sets=sample(rep(seq(1,5,by=1),19),95,replace=FALSE)
```

```
Ind.sets
```

```
A=which(Ind.sets==1)
```

```
B=which(Ind.sets==2)
```

```
C=which(Ind.sets==3)
```

```
D=which(Ind.sets==4)
```

```
E=which(Ind.sets==5)
```

```
sets<-Ind.sets
```

```
## A ##
```

```
BNDy1<-BNDY$PNAj
```

```
BNDyNa1=BNDy1
```

```
#BNDyNa1[A]=NA
```

```
BNDy2<-BNDY$PDAj
```

```
BNDyNa2=BNDy2
```

```
BNDyNa2[A]=NA
```

```
BNDy<-cbind(BNDyNa1,BNDyNa2)
```

```
fBDNA <- MTM(Y = BNDy,K = list(list(K = G2,COV = list(type = 'UN',df0 = 4,S0 =  
diag(2))))),
```

```
resCov = list(type = 'UN',S0 = diag(2),df0 = 4),nlter = 10000,burnIn = 500,thin =  
5,saveAt = 'ex1_')
```

```
## B ##
```

```
BNDy1<-BNDY$PNAj
```

```
BNDyNa1=BNDy1
```

```
#BNDyNa1[B]=NA
```

```
BNDy2<-BNDY$PDAj
```

```
BNDyNa2=BNDy2
```

```
BNDyNa2[B]=NA
```

```
BNDy<-cbind(BNDyNa1,BNDyNa2)
```

```
fBDNB <- MTM(Y = BNDy,K = list(list(K = G2,COV = list(type = 'UN',df0 = 4,S0 =  
diag(2))))),
```

```
resCov = list(type = 'UN',S0 = diag(2),df0 = 4),nlter = 10000,burnIn = 500,thin =  
5,saveAt = 'ex1_')
```

```
## C ##
```

```
BNDy1<-BNDY$PNAj
```

```
BNDyNa1=BNDy1
```

```
#BNDyNa1[C]=NA
```

```
BNDy2<-BNDY$PDAj
```

```
BNDyNa2=BNDy2
BNDyNa2[C]=NA
```

```
BNDy<-cbind(BNDyNa1,BNDyNa2)
```

```
fBDNC <- MTM(Y = BNDy,K = list(list(K = G2,COV = list(type = 'UN',df0 = 4,S0 =
diag(2))))),
  resCov = list(type = 'UN',S0 = diag(2),df0 = 4),nlter = 10000,burnIn = 500,thin =
5,saveAt = 'ex1_')
```

```
## D ##
```

```
BNDy1<-BNDY$PNAj
BNDyNa1=BNDy1
#BNDyNa1[D]=NA
BNDy2<-BNDY$PDAj
BNDyNa2=BNDy2
BNDyNa2[D]=NA
```

```
BNDy<-cbind(BNDyNa1,BNDyNa2)
```

```
fBDND <- MTM(Y = BNDy,K = list(list(K = G2,COV = list(type = 'UN',df0 = 4,S0 =
diag(2))))),
  resCov = list(type = 'UN',S0 = diag(2),df0 = 4),nlter = 10000,burnIn = 500,thin =
5,saveAt = 'ex1_')
```

```
cor(fBDND$YHat[D,],PNPD[D,])
```

```
## E ##
```

```
BNDy1<-BNDY$PNAj
BNDyNa1=BNDy1
#BNDyNa1[E]=NA
BNDy2<-BNDY$PDAj
BNDyNa2=BNDy2
BNDyNa2[E]=NA
```

```
BNDy<-cbind(BNDyNa1,BNDyNa2)
```

```
fBDNE <- MTM(Y = BNDy,K = list(list(K = G2,COV = list(type = 'UN',df0 = 4,S0 =
diag(2))))),
  resCov = list(type = 'UN',S0 = diag(2),df0 = 4),nlter = 10000,burnIn = 500,thin =
5,saveAt = 'ex1_')
```

```
CorVCPDPN<-c(cor(fBDNA$YHat[A,][,2],PNPD[A,][,2]),
  cor(fBDNB$YHat[B,][,2],PNPD[B,][,2]),
  cor(fBDNC$YHat[C,][,2],PNPD[C,][,2]),
  cor(fBDND$YHat[D,][,2],PNPD[D,][,2]),
  cor(fBDNE$YHat[E,][,2],PNPD[E,][,2]))
```

```
CorVCPDPN
```

## Apéndice 8. Código en R para análisis bivariado de PN y PA y su validación cruzada

```
rm(list=ls())
library(MTM)
library(BGLR)

#Cargar carpeta de trabajo
setwd("C:/Users/jev_h_/Documents/Maestría/Genética/Análisis de
datos/Bivariados/BivPN_PA")
dir()

BNAY=read.csv(file="FenPN_PA.csv",header=TRUE)
PNPA<-cbind(BNAY$PNAj,BNAY$PAAj)

MatrizX<-load("C:/Users/jev_h_/Documents/Maestría/Genética/Análisis de
datos/Bivariados/BivPN_PA/Matriz X.RData")

#Genomic Relationship matrix
Z=scale(X, center=TRUE,scale=TRUE)
G=tcrossprod(Z)/ncol(Z)
dim(G)

ids <- as.character(BNAY$ID)
idsall<-colnames(G)
idsall<-rownames(G)
indice=idsall%in%ids
G1<-G[indice,]
G2<-G1[,indice]
dim(G2)
rownames(G2)

##### Sin validación
fBNA <- MTM(Y = PNPA,K = list(list(K = G2,COV = list(type = 'UN',df0 = 4,S0 = diag(2)))),
      resCov = list(type = 'UN',S0 = diag(2),df0 = 4),nlter = 10000,burnIn = 500,thin =
5,saveAt = 'ex1_')

cor(fBNA$YHat,PNPA)

# Recuperando estimaciones
fBNA$YHat # predicitons
fBNA$resCov$R # Matriz de covarianzas residual
fBNA$K[[1]]$G # matriz de covarianzas genéticas
fBNA$K[[1]]$U # Efectos aleatorios

cor(fBNA$YHat[,1],fBNA$YHat[,2])
```

```
##### Validación cruzada PN CON PA #####
```

```
### Creación de poblaciones de entrenamiento y prueba ###
```

```
Ind.sets=sample(rep(seq(1,5,by=1),19),95,replace=FALSE)
```

```
Ind.sets
```

```
A=which(Ind.sets==1)
```

```
B=which(Ind.sets==2)
```

```
C=which(Ind.sets==3)
```

```
D=which(Ind.sets==4)
```

```
E=which(Ind.sets==5)
```

```
sets<-Ind.sets
```

```
## A ##
```

```
BNay1<-BNAY$PNAj
```

```
BNayNa1=BNay1
```

```
BNayNa1[A]=NA
```

```
BNay2<-BNAY$PAAj
```

```
BNayNa2=BNay2
```

```
#BNayNa2[A]=NA
```

```
BNay<-cbind(BNayNa1,BNayNa2)
```

```
fBNAA <- MTM(Y = BNay,K = list(list(K = G2,COV = list(type = 'UN',df0 = 4,S0 = diag(2)))),  
            resCov = list(type = 'UN',S0 = diag(2),df0 = 4),nlter = 10000,burnIn = 500,thin =  
5,saveAt = 'ex1_')
```

```
## B ##
```

```
BNay1<-BNAY$PNAj
```

```
BNayNa1=BNay1
```

```
BNayNa1[B]=NA
```

```
BNay2<-BNAY$PAAj
```

```
BNayNa2=BNay2
```

```
#BNayNa2[B]=NA
```

```
BNay<-cbind(BNayNa1,BNayNa2)
```

```
fBNAB <- MTM(Y = BNay,K = list(list(K = G2,COV = list(type = 'UN',df0 = 4,S0 = diag(2)))),  
            resCov = list(type = 'UN',S0 = diag(2),df0 = 4),nlter = 10000,burnIn = 500,thin =  
5,saveAt = 'ex1_')
```

```
## C ##
```

```
BNay1<-BNAY$PNAj
```

```
BNayNa1=BNay1
```

```
BNayNa1[C]=NA
```

```
BNay2<-BNAY$PAAj
```

```
BNayNa2=BNay2
```

```
#BNayNa2[C]=NA
```

```
BNAY<-cbind(BNAYNa1,BNAYNa2)
```

```
fBNAC <- MTM(Y = BNAY,K = list(list(K = G2,COV = list(type = 'UN',df0 = 4,S0 =  
diag(2)))),  
resCov = list(type = 'UN',S0 = diag(2),df0 = 4),nlter = 10000,burnIn = 500,thin =  
5,saveAt = 'ex1_')
```

```
## D ##
```

```
BNAY1<-BNAY$PNAj  
BNAYNa1=BNAY1  
BNAYNa1[D]=NA  
BNAY2<-BNAY$PAAj  
BNAYNa2=BNAY2  
#BNAYNa2[D]=NA
```

```
BNAY<-cbind(BNAYNa1,BNAYNa2)
```

```
fBNAD <- MTM(Y = BNAY,K = list(list(K = G2,COV = list(type = 'UN',df0 = 4,S0 =  
diag(2)))),  
resCov = list(type = 'UN',S0 = diag(2),df0 = 4),nlter = 10000,burnIn = 500,thin =  
5,saveAt = 'ex1_')
```

```
## E ##
```

```
BNAY1<-BNAY$PNAj  
BNAYNa1=BNAY1  
BNAYNa1[E]=NA  
BNAY2<-BNAY$PAAj  
BNAYNa2=BNAY2  
#BNAYNa2[E]=NA
```

```
BNAY<-cbind(BNAYNa1,BNAYNa2)
```

```
fBNAE <- MTM(Y = BNAY,K = list(list(K = G2,COV = list(type = 'UN',df0 = 4,S0 = diag(2)))),  
resCov = list(type = 'UN',S0 = diag(2),df0 = 4),nlter = 10000,burnIn = 500,thin =  
5,saveAt = 'ex1_')
```

```
CorVCPNPA<-c(cor(fBNAE$YHat[A],[,1],PNPA[A],[,1]),  
cor(fBNAB$YHat[B],[,1],PNPA[B],[,1]),  
cor(fBNAC$YHat[C],[,1],PNPA[C],[,1]),  
cor(fBNAD$YHat[D],[,1],PNPA[D],[,1]),  
cor(fBNAE$YHat[E],[,1],PNPA[E],[,1]))
```

```
CorVCPNPA
```



```
##### Validación cruzada PA CON PN #####
```

```
### Creación de poblaciones de entrenamiento y prueba ###
```

```
Ind.sets=sample(rep(seq(1,5,by=1),19),95,replace=FALSE)
```

```
Ind.sets
```

```
A=which(Ind.sets==1)
```

```
B=which(Ind.sets==2)
```

```
C=which(Ind.sets==3)
```

```
D=which(Ind.sets==4)
```

```
E=which(Ind.sets==5)
```

```
sets<-Ind.sets
```

```
## A ##
```

```
BNAY1<-BNAY$PNAj
```

```
BNAYNa1=BNAY1
```

```
#BNAYNa1[A]=NA
```

```
BNAY2<-BNAY$PAAj
```

```
BNAYNa2=BNAY2
```

```
BNAYNa2[A]=NA
```

```
BNAY<-cbind(BNAYNa1,BNAYNa2)
```

```
fBANA <- MTM(Y = BNAY,K = list(list(K = G2,COV = list(type = 'UN',df0 = 4,S0 = diag(2)))),  
            resCov = list(type = 'UN',S0 = diag(2),df0 = 4),nlter = 10000,burnIn = 500,thin =  
5,saveAt = 'ex1_')
```

```
## B ##
```

```
BNAY1<-BNAY$PNAj
```

```
BNAYNa1=BNAY1
```

```
#BNAYNa1[B]=NA
```

```
BNAY2<-BNAY$PAAj
```

```
BNAYNa2=BNAY2
```

```
BNAYNa2[B]=NA
```

```
BNAY<-cbind(BNAYNa1,BNAYNa2)
```

```
fBANB <- MTM(Y = BNAY,K = list(list(K = G2,COV = list(type = 'UN',df0 = 4,S0 = diag(2)))),  
            resCov = list(type = 'UN',S0 = diag(2),df0 = 4),nlter = 10000,burnIn = 500,thin =  
5,saveAt = 'ex1_')
```

```
## C ##
```

```
BNAY1<-BNAY$PNAj
```

```
BNAYNa1=BNAY1
```

```
#BNAYNa1[C]=NA
```

```
BNAY2<-BNAY$PAAj
```

```

BNAYna2=BNAY2
BNAYna2[C]=NA

BNAY<-cbind(BNAYna1,BNAYna2)

fBANC <- MTM(Y = BNAY,K = list(list(K = G2,COV = list(type = 'UN',df0 = 4,S0 =
diag(2)))),
      resCov = list(type = 'UN',S0 = diag(2),df0 = 4),nlter = 10000,burnIn = 500,thin =
5,saveAt = 'ex1_')

## D ##
BNAY1<-BNAY$PNAj
BNAYna1=BNAY1
#BNAYna1[D]=NA
BNAY2<-BNAY$PAAj
BNAYna2=BNAY2
BNAYna2[D]=NA

BNAY<-cbind(BNAYna1,BNAYna2)

fBAND <- MTM(Y = BNAY,K = list(list(K = G2,COV = list(type = 'UN',df0 = 4,S0 =
diag(2)))),
      resCov = list(type = 'UN',S0 = diag(2),df0 = 4),nlter = 10000,burnIn = 500,thin =
5,saveAt = 'ex1_')

## E ##
BNAY1<-BNAY$PNAj
BNAYna1=BNAY1
#BNAYna1[E]=NA

BNAY2<-BNAY$PAAj
BNAYna2=BNAY2
BNAYna2[E]=NA

BNAY<-cbind(BNAYna1,BNAYna2)

fBANE <- MTM(Y = BNAY,K = list(list(K = G2,COV = list(type = 'UN',df0 = 4,S0 = diag(2)))),
      resCov = list(type = 'UN',S0 = diag(2),df0 = 4),nlter = 10000,burnIn = 500,thin =
5,saveAt = 'ex1_')

CorVCPAPN<-c(cor(fBANAS$YHat[A],[,2],PNPA[A],[,2]),
      cor(fBANB$YHat[B],[,2],PNPA[B],[,2]),
      cor(fBANC$YHat[C],[,2],PNPA[C],[,2]),
      cor(fBAND$YHat[D],[,2],PNPA[D],[,2]),
      cor(fBANE$YHat[E],[,2],PNPA[E],[,2]))

CorVCPAPN

```

## Apéndice 9. Código en R para análisis bivariado de PD y PA y su validación cruzada

```
rm(list=ls())
library(MTM)
library(BGLR)

#Cargar carpeta de trabajo
setwd("C:/Users/jev_h_/Documents/Maestría/Genética/Análisis de
datos/Bivariados/BivPD_PA")
dir()

BDAY=read.csv(file="FenPD_PA.csv",header=TRUE)
PDPA<-cbind(BDAY$PDAj,BDAY$PAAj)

MatrizX<-load("C:/Users/jev_h_/Documents/Maestría/Genética/Análisis de
datos/Bivariados/BivPD_PA/Matriz X.RData")

#Genomic Relationship matrix
Z=scale(X, center=TRUE,scale=TRUE)
G=tcrossprod(Z)/ncol(Z)
dim(G)

ids <- as.character(BDAY$ID)
idsall<-colnames(G)
idsall<-rownames(G)
indice=idsall%in%ids
G1<-G[indice,]
G2<-G1[,indice]
dim(G2)
rownames(G2)

##### Sin validación
fBDA <- MTM(Y = PDPA,K = list(list(K = G2,COV = list(type = 'UN',df0 = 4,S0 = diag(2)))),
          resCov = list(type = 'UN',S0 = diag(2),df0 = 4),nlter = 10000,burnIn = 500,thin =
5,saveAt = 'ex1_')

cor(fBDA$YHat,PDPA)

# Recuperando estimaciones
fBDA$YHat # predicitons
fBDA$resCov$R # Matriz de covarianzas residual
fBDA$K[[1]]$G # Matriz de covarianza genética
fBDA$K[[1]]$U # Efectos aleatorios

cor(fBDA$YHat[,1],fBDA$YHat[,2])
```

```
##### Validación cruzada PD CON PA #####
```

```
### Creación de poblaciones de entrenamiento y prueba ###
```

```
Ind.sets=sample(rep(seq(1,5,by=1),19),95,replace=FALSE)
```

```
Ind.sets
```

```
A=which(Ind.sets==1)
```

```
B=which(Ind.sets==2)
```

```
C=which(Ind.sets==3)
```

```
D=which(Ind.sets==4)
```

```
E=which(Ind.sets==5)
```

```
sets<-Ind.sets
```

```
## A ##
```

```
BDAy1<-BDAY$PDAj
```

```
BDAyNa1=BDAy1
```

```
BDAyNa1[A]=NA
```

```
BDAy2<-BDAY$PAAj
```

```
BDAyNa2=BDAy2
```

```
#BDAyNa2[A]=NA
```

```
BDAy<-cbind(BDAyNa1,BDAyNa2)
```

```
fBDAA <- MTM(Y = BDAy,K = list(list(K = G2,COV = list(type = 'UN',df0 = 4,S0 = diag(2)))),  
            resCov = list(type = 'UN',S0 = diag(2),df0 = 4),nlter = 10000,burnIn = 500,thin =  
5,saveAt = 'ex1_')
```

```
## B ##
```

```
BDAy1<-BDAY$PDAj
```

```
BDAyNa1=BDAy1
```

```
BDAyNa1[B]=NA
```

```
BDAy2<-BDAY$PAAj
```

```
BDAyNa2=BDAy2
```

```
#BDAyNa2[B]=NA
```

```
BDAy<-cbind(BDAyNa1,BDAyNa2)
```

```
fBDAB <- MTM(Y = BDAy,K = list(list(K = G2,COV = list(type = 'UN',df0 = 4,S0 = diag(2)))),  
            resCov = list(type = 'UN',S0 = diag(2),df0 = 4),nlter = 10000,burnIn = 500,thin =  
5,saveAt = 'ex1_')
```

```
## C ##
```

```
BDAy1<-BDAY$PDAj
```

```
BDAyNa1=BDAy1
```

```
BDAyNa1[C]=NA
```

```
BDAy2<-BDAY$PAAj
```

```
BDAyNa2=BDAy2
```

```
#BDAyNa2[C]=NA
```

```
BDAy<-cbind(BDAyNa1,BDAyNa2)
```

```
fBDAC <- MTM(Y = BDAy,K = list(list(K = G2,COV = list(type = 'UN',df0 = 4,S0 =  
diag(2)))),  
resCov = list(type = 'UN',S0 = diag(2),df0 = 4),nlter = 10000,burnIn = 500,thin =  
5,saveAt = 'ex1_')
```

```
## D ##
```

```
BDAy1<-BDAY$PDAj  
BDAyNa1=BDAy1  
BDAyNa1[D]=NA  
BDAy2<-BDAY$PAAj  
BDAyNa2=BDAy2  
#BDAyNa2[D]=NA
```

```
BDAy<-cbind(BDAyNa1,BDAyNa2)
```

```
fBDAD <- MTM(Y = BDAy,K = list(list(K = G2,COV = list(type = 'UN',df0 = 4,S0 =  
diag(2)))),  
resCov = list(type = 'UN',S0 = diag(2),df0 = 4),nlter = 10000,burnIn = 500,thin =  
5,saveAt = 'ex1_')
```

```
## E ##
```

```
BDAy1<-BDAY$PDAj  
BDAyNa1=BDAy1  
BDAyNa1[E]=NA  
BDAy2<-BDAY$PAAj  
BDAyNa2=BDAy2  
#BDAyNa2[E]=NA
```

```
BDAy<-cbind(BDAyNa1,BDAyNa2)
```

```
fBDAE <- MTM(Y = BDAy,K = list(list(K = G2,COV = list(type = 'UN',df0 = 4,S0 = diag(2)))),  
resCov = list(type = 'UN',S0 = diag(2),df0 = 4),nlter = 10000,burnIn = 500,thin =  
5,saveAt = 'ex1_')
```

```
CorVCPDPA<-c(cor(fBDAA$YHat[A],[,1],PDPA[A],[,1]),  
cor(fBDAB$YHat[B],[,1],PDPA[B],[,1]),  
cor(fBDAC$YHat[C],[,1],PDPA[C],[,1]),  
cor(fBDAD$YHat[D],[,1],PDPA[D],[,1]),  
cor(fBDAE$YHat[E],[,1],PDPA[E],[,1]))
```

```
CorVCPDPA
```

```
##### Validación cruzada PA CON PD #####
```

```
### Creación de poblaciones de entrenamiento y prueba ###
```

```
Ind.sets=sample(rep(seq(1,5,by=1),19),95,replace=FALSE)
```

```
Ind.sets
```

```
A=which(Ind.sets==1)
```

```
B=which(Ind.sets==2)
```

```
C=which(Ind.sets==3)
```

```
D=which(Ind.sets==4)
```

```
E=which(Ind.sets==5)
```

```
sets<-Ind.sets
```

```
## A ##
```

```
BDAy1<-BDAY$PDAj
```

```
BDAyNa1=BDAy1
```

```
#BNAyNa1[A]=NA
```

```
BDAy2<-BDAY$PAAj
```

```
BDAyNa2=BDAy2
```

```
BDAyNa2[A]=NA
```

```
BDAy<-cbind(BDAyNa1,BDAyNa2)
```

```
fBADA <- MTM(Y = BDAy,K = list(list(K = G2,COV = list(type = 'UN',df0 = 4,S0 = diag(2)))),  
            resCov = list(type = 'UN',S0 = diag(2),df0 = 4),nlter = 10000,burnIn = 500,thin =  
5,saveAt = 'ex1_')
```

```
## B ##
```

```
BDAy1<-BDAY$PDAj
```

```
BDAyNa1=BDAy1
```

```
#BDAyNa1[B]=NA
```

```
BDAy2<-BDAY$PAAj
```

```
BDAyNa2=BDAy2
```

```
BDAyNa2[B]=NA
```

```
BDAy<-cbind(BDAyNa1,BDAyNa2)
```

```
fBADB <- MTM(Y = BDAy,K = list(list(K = G2,COV = list(type = 'UN',df0 = 4,S0 = diag(2)))),  
            resCov = list(type = 'UN',S0 = diag(2),df0 = 4),nlter = 10000,burnIn = 500,thin =  
5,saveAt = 'ex1_')
```

```
## C ##
```

```
BDAy1<-BDAY$PDAj
```

```
BDAyNa1=BDAy1
```

```
#BDAyNa1[C]=NA
```

```
BDAy2<-BDAY$PAAj
```

```
BDAyNa2=BDAy2
```

```
BDAyNa2[C]=NA
```

```
BDAy<-cbind(BDAyNa1,BDAyNa2)
```

```
fBADC <- MTM(Y = BDAy,K = list(list(K = G2,COV = list(type = 'UN',df0 = 4,S0 =  
diag(2)))),  
resCov = list(type = 'UN',S0 = diag(2),df0 = 4),nlter = 10000,burnIn = 500,thin =  
5,saveAt = 'ex1_')
```

```
## D ##
```

```
BDAy1<-BDAY$PDAj
```

```
BDAyNa1=BDAy1
```

```
#BDAyNa1[D]=NA
```

```
BDAy2<-BDAY$PAAj
```

```
BDAyNa2=BDAy2
```

```
BDAyNa2[D]=NA
```

```
BDAy<-cbind(BDAyNa1,BDAyNa2)
```

```
fBADD <- MTM(Y = BDAy,K = list(list(K = G2,COV = list(type = 'UN',df0 = 4,S0 =  
diag(2)))),  
resCov = list(type = 'UN',S0 = diag(2),df0 = 4),nlter = 10000,burnIn = 500,thin =  
5,saveAt = 'ex1_')
```

```
## E ##
```

```
BDAy1<-BDAY$PDAj
```

```
BDAyNa1=BDAy1
```

```
#BDAyNa1[E]=NA
```

```
BDAy2<-BDAY$PAAj
```

```
BDAyNa2=BDAy2
```

```
BDAyNa2[E]=NA
```

```
BDAy<-cbind(BDAyNa1,BDAyNa2)
```

```
fBADE <- MTM(Y = BDAy,K = list(list(K = G2,COV = list(type = 'UN',df0 = 4,S0 = diag(2)))),  
resCov = list(type = 'UN',S0 = diag(2),df0 = 4),nlter = 10000,burnIn = 500,thin =  
5,saveAt = 'ex1_')
```

```
CorVCPAPD<-c(cor(fBADA$YHat[A],[,2],PDPA[A],[,2]),  
cor(fBADB$YHat[B],[,2],PDPA[B],[,2]),  
cor(fBADDC$YHat[C],[,2],PDPA[C],[,2]),  
cor(fBADD$YHat[D],[,2],PDPA[D],[,2]),  
cor(fBADE$YHat[E],[,2],PDPA[E],[,2]))
```

```
CorVCPAPD
```