



**TRABAJO FIN DE MÁSTER EN BIOESTADÍSTICA**

**ANÁLISIS DE SUPERVIVENCIA  
EN COMPETICIONES HÍPICAS  
DE RESISTENCIA**

**OCTUBRE 2020**

Alumna: Andrea Gaetano Gil

Tutores: Teresa Pérez Pérez, Isabel Cervantes Navarro, Juan Pablo Gutiérrez García



**INDICE:**

INDICE: .....	3
1. RESUMEN .....	5
1.1 ABSTRACT .....	5
2.INTRODUCCIÓN.....	6
2.1 Carreras de resistencia (RAID).....	6
2.2 Categorías de los Raid .....	7
2.3 Programas de mejora .....	8
2. 3.1 Programa de Mejora del Pura Raza árabe .....	11
2.4 Carácter longevidad.....	15
3.OBJETIVOS.....	17
4. MÉTODOS.....	18
4.1 El análisis de supervivencia, uso en mejora genética.....	18
4.2 Conceptos básicos de Análisis de supervivencia.....	19
4.3 Modelos de fragilidad.....	21
4.4 Estimación de los modelos de fragilidad individual.....	23
4.5 Cox vs Weibull en modelos de mejora genética.....	24
4.6 Parámetros genéticos .....	25
4.7 Survival Kit .....	26
5.DESCRIPCIÓN DE LA BASE DE DATOS .....	29
6.APLICACIÓN A UN CASO DE MEJORA GENÉTICA .....	31
6.1 Estadística descriptiva .....	32
6.2 Resultados.....	35
6.2.1. Efectos de los factores de riesgo.....	35
6.2.2 Determinación genética de la longevidad productiva.....	36
7.DISCUSION .....	38
8.CONCLUSIONES.....	42

9.AGRADECIMIENTOS.....	43
10. REFERENCIAS BIBLIOGRÁFICAS .....	44
11.APÉNDICES .....	49
APENDICE 1 .....	49

## 1. RESUMEN

El presente estudio tiene como objetivo realizar un análisis genético de la longevidad deportiva medida como el número de años en competición en la disciplina del Raid, y examinar los factores que le afectan a través del uso del análisis de supervivencia. Se han analizado los resultados de pruebas de Raid nacionales e internacionales en España de 1419 equinos entre 2000 y 2018 ( $n = 6136$  registros) utilizando la metodología de análisis de supervivencia. Se utilizó un modelo de fragilidad individual y paramétrico que incluía el sexo (variable tiempo dependiente), la edad en la primera carrera, si se elimina o no el caballo en la primera carrera, el número de competiciones de cada caballo en cada categoría (1,2,3 y 4) y el caballo como efecto aleatorio. Se observaron mayores hazard ratio para los caballos que habían sido eliminados en la primera carrera (1.32, con un IC al 95% de: 1.15;1.48) con respecto aquellos que no lo habían sido. Una mayor frecuencia de carreras de categoría 1,2 y 3 resultó ser un factor protector para no presentar el evento. Para la categoría 1 se obtuvo un hazard ratio de 0.73 con un IC al 95% de 0.63;0.78, para la categoría 2 fue de 0.68 con un IC al 95% de 0.62;0.74 y finalmente, para la categoría 3, se obtuvo un valor de 0.72 con un IC al 95% de 0.61;0.83. También, se demostró que el riesgo de no seguir en competición aumentaba con la edad de inicio en la misma; se obtuvo un hazard ratio de 1.36 con un IC al 95% de 1.21;1.51. La heredabilidad estimada ( $0,45 \pm 0,01$ ) para la longevidad deportiva, definida como años en competición, indica que debe esperarse una buena respuesta a la selección genética directa. Por lo tanto, utilizar esta variable podría ser una buena opción para maximizar el éxito de los programas de mejora de equinos de resistencia.

**Palabras clave:** Raid, longevidad, heredabilidad, análisis de supervivencia, modelos de fragilidad

### 1.1 ABSTRACT

The present study aims to perform a genetic analysis of sports longevity measured as the number of years in competition in the discipline of Raid and examine the factors that affect it using survival analysis. The results of national and international Raid tests in Spain of 1419 horses between 2000 and 2018 ( $n = 6136$  records) have been analysed using the methodology of survival analysis. An individual and parametric frailty model was used which included sex (time dependent variable), age in the first race, whether or not the horse is eliminated in the first race, the number of competitions of each horse in each category (1,2,3 and 4) and the horse as a random effect. Higher hazard ratios were observed for horses that had been eliminated in the first race (1.32, with a 95% CI of: 1.15;1.48) with respect to those that had not been eliminated. A higher frequency of category 1,2 and 3 races turned out to be a protective factor for not presenting the event. For category 1 a hazard ratio of 0.73 with a 95% CI of 0.63;0.78 was obtained, for category 2 it was 0.68 with a 95% CI of 0.62;0.74 and finally, for category 3, a value of 0.72 with a 95% CI of 0.61;0.83 was obtained. Also, it was demonstrated that the risk of not continuing in competition increased with the age of beginning the competition; a hazard ratio of 1.36 was obtained with a 95% CI of 1.20;1.51. The estimated heritability ( $0.45 \pm 0.01$ ) for sports longevity, defined as years in competition, indicates that a good response to direct genetic selection should be expected. Therefore, using this variable could be a good option to maximize the success of endurance equine breeding programs.

**Keywords:** Raid, longevity, heritability, survival analysis, fragility models.

## 2.INTRODUCCIÓN

### 2.1 Carreras de resistencia (RAID)

Según la Real Federación Hípica Española (1) el Raid se define como disciplina ecuestre, en la cual se pone a prueba la resistencia, condición física y psicológica de caballo y jinete, ya que deben recorrer grandes distancias, en diferentes terrenos, a menudo en diferentes climas y contra el tiempo. Para poder lograr exitosamente esto, el jinete debe tener la capacidad de controlar el esfuerzo de su caballo y llevar así un ritmo adecuado durante la prueba, asegurando al máximo la salud y bienestar del mismo. Si al finalizar una carrera, o durante los controles veterinarios intermedios, el caballo presenta una frecuencia cardiaca más alta de lo permitido o alguna anomalía locomotora o metabólica es eliminado.

De forma general, la competición consiste en una serie de fases de una distancia máxima de 40 Km y una mínima de 20 Km en las cuales se incluyen cambios de terreno, altitud, dirección y longitud. Estas fases pueden ser repartidas en uno o más días, llamándose así cada día etapa. Tras finalizar cada una de estas fases, se debe dar una parada obligatoria para la inspección veterinaria o Vet. Gate (Ilustración 1).



*Ilustración 1. Equino en un control veterinario.*

Finalmente, al tratarse de una competición a contrarreloj, el binomio jinete-caballo que termine la carrera en el menor tiempo será el ganador (superando los controles veterinarios). De cara a los programas de mejora, de los Raid se recogen los tiempos totales en carrera, los tiempos de recuperación y también los puestos clasificatorios. La información de los controles veterinarios también queda registrada. Cabe destacar que en

esta disciplina se otorga también el premio a la mejor condición cuya finalidad es dar un reconocimiento al caballo que, después de haber completado el Raid entre los diez primeros, se encuentre en un estado de salud superior al resto (2).

## 2.2 Categorías de los Raid

Los Raid se dividen en diferentes categorías atendiendo a si se trata de concursos nacionales (CEN) (Tabla 1) o internacionales (CEI) (Tabla 2) y dentro de estos según la distancia de las carreras. En todos, las pulsaciones por minuto máximas permitidas a los 15-20 minutos de cruzar una línea de meta son 64 (2).



Ilustración 2. Recorrido de un Raid

Tabla 1. Categorías de los Raid nacionales (2)

Categoría	Distancia en km
<b>CEN****</b>	más de 120-160 Km
<b>CEN***</b>	140-160 Km
<b>CEN**</b>	120-139 Km
<b>CEN*</b>	80 y 119 Km
<b>0*</b>	80 y 90 Km
<b>Promoción 40</b>	40-45 Km
<b>Promoción 60</b>	57-70

Tabla 2. Categorías de los concursos internacionales (3)

<b>Categoría</b>	<b>Distancia en km</b>
<b>CEI***</b>	140-160 Km
<b>CEI**</b>	120-139 Km
<b>CEI*</b>	100 y 119 Km

### **2.3 Programas de mejora**

A lo largo de la historia, los caballos han desempeñado un papel importante en la sociedad y la economía de la mayoría de los países europeos otorgando una gran importancia al proceso de la cría equina, ya que es, en última instancia, la que permitía a los ganaderos ofrecer animales competitivos (4). Esto provocó una creciente necesidad de mejorar los recursos zoogenéticos que se tenían y a su vez conservarlos, sobre todo, hablando del caballo, considerado en un principio animal para la guerra (5).

En 1978, a través de la Orden del 26 de diciembre (6), el Ministerio de Defensa aprueba el Reglamento del Registro-Matricula de Caballos y Yeguas de Pura Raza. No obstante, no fue hasta el año 1997, con el Real Decreto 1682/1997 (7), cuando el caballo se empezó a considerar especie ganadera ya que se incluyó en el Catálogo Oficial de Razas de Ganado de España. Posteriormente, se ha actualizado esta última normativa a través del Real Decreto 45/2019, de 8 de febrero (8) por el cual se establecen las normas zootécnicas aplicables a los animales reproductores de raza pura, porcinos reproductores híbridos y su material reproductivo, se actualiza el Programa nacional de conservación, mejora y fomento de las razas ganaderas y se modifican los Reales Decretos 558/2001, de 25 de mayo; 1316/1992, de 30 de octubre; 1438/1992, de 27 de noviembre; y 1625/2011, de 14 de noviembre

Así pues, actualmente, se define Programa de Mejora o de cría según el Ministerio de Agricultura, Pesca y Alimentación (9) al conjunto de actuaciones, entre las que se incluye el registro, selección, cría e intercambio de animales reproductores y de su material reproductivo, diseñadas y aplicadas para conservar y/o mejorar las características fenotípicas y/o genotípicas deseadas en la población reproductora objetivo. Su finalidad



podrá ser la conservación, la mejora, la reconstrucción o la creación de una raza, o una combinación de dichas finalidades. Por tanto, deben contener las disposiciones que afectan tanto al libro genealógico como las actividades que realizan para conseguir sus objetivos.

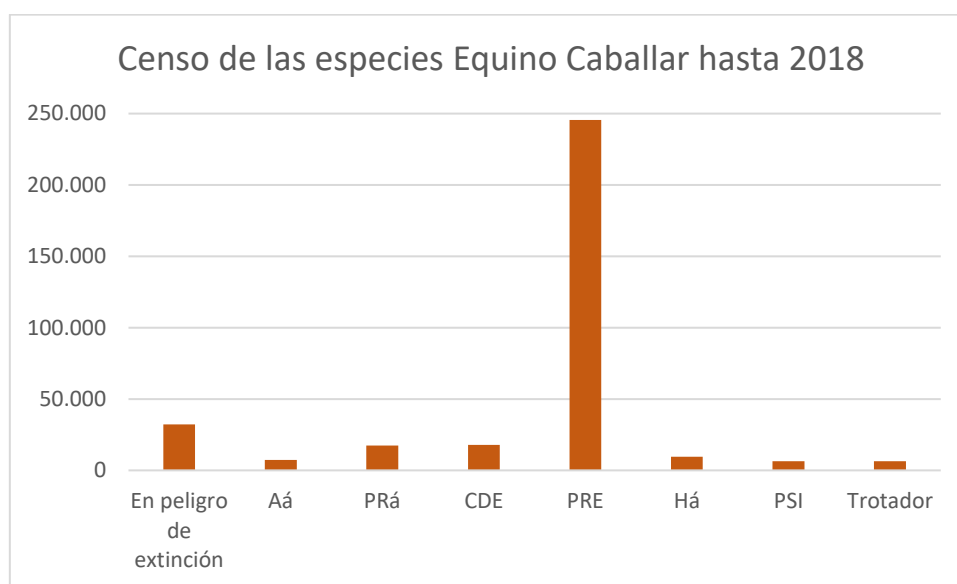
Los programas de cría son desarrollados por las Asociaciones de Criadores oficialmente reconocidas por las autoridades competentes. Estas a su vez, tienen entre sus objetivos aconsejar a los ganaderos sobre los óptimos equinos reproductores obteniendo así caballos económicamente viables y manteniendo en todo momento las características interpoblaciones (10)

Actualmente, el Catálogo Oficial de Razas de Ganado de España, recogido en el Anexo I del Real Decreto 45/2019 (8) contiene la relación oficial y la clasificación de todas las razas ganaderas reconocidas y utilizadas en España por su interés económico, zootécnico, productivo, cultural, medioambiental o social, destinadas a ser objeto de un programa de cría. En concreto, en cuanto a la especie Equino Caballar encontramos: Asturcón, Burguete, Caballo de las Retuertas, Caballo de Monte de País Vasco, Caballo de Pura Raza Galega, Cavall Mallorquí, Cavall Menorquí, Cavall Pirineic Catalá, Hispano Breton, Jaca Navarra, Losina, Marismeña, Monchina y Pottoka todos ellos nacionales y en peligro de extinción por lo que no se usan para los deportes ecuestres. Los caballos que si se usan para los deportes los encontramos en la Tabla 3, cabe destacar el Caballo Hispano- Árabe que, aunque está clasificado en peligro de extinción (9), es también reconocido por su uso en el deporte.

Tabla 3. Especies utilizadas para el deporte ecuestre (MAPA) (9) donde encontramos el Anglo-Árabe (Aá), el Pura Raza árabe (PRá), Caballo de Deporte Español (CDE), Pura Raza Española (PRE), el Hispano Árabe (Há), el Pura Sangre Inglés (PSI) y el Trotador

Especie	Autóctona	No autóctona
Aá		✗
PRá		✗
CDE	✗	
PRE	✗	
Há	✗	
PSI		✗
Trotador	✗	

En la Gráfica 1 se muestran los datos del último censo registrado por el Ministerio de Agricultura Pesca y Alimentación (9). En esta podemos observar que, el PRE es el equino más numeroso hasta la fecha en nuestro país, llegando a alcanzar en número, casi los 250.000 caballos. Seguidamente, encontramos el PRá, el CDE, el Há y el Aá que son los más utilizados en las carreras de resistencia y cuyas características principales se encuentran en la Tabla 4 (9).



Gráfica 1. Censo de las especies Equino Caballar hasta 2018 (9) en ella encontramos los equinos en peligro de extinción, el caballo Anglo-árabe (Aá), el Pura Raza árabe (PRá), el Caballo de Deporte Español (CDE), el Pura raza Española, el Hispano-árabe (Há), el Pura Sangre Inglés (PSI) y el Trotador

Tabla 4. Características principales y para el deporte ecuestre del caballo Pura Raza Árabe (PRá), Anglo-árabe (Aá), el Caballo de Deporte Español (CDE) y el Hispano-árabe (Há) (9).

	<b>Origen</b>	<b>Descripción del origen</b>	<b>Cualidades para el deporte</b>
<b>PRá</b>	Mar Rojo y Golfo Pérsico	Datan del 1500 a.C sometidos a selección durante 350 generaciones	Enorme resistencia y velocidad
<b>Aá</b>	Francia.	cruce de PRá y el PSI.	Resistencia, movimientos elásticos y equilibrados, marcha distinguida, galope amplio y rápido
<b>CDE</b>	España	Determinados como: “caballos cruzados” de puras razas hasta 1999	Según la raza de sus progenitores tendrá unas cualidades u otras
<b>Há</b>	España	Cruce del PRá y del PRE	Docilidad, equilibrio, resistencia, agilidad, temperamento y rápida respuesta y resistencia

Tal y como se puede observar en la Tabla 4 el PRá es el caballo mejor adaptado a las carreras de resistencia, de hecho, el formar parte del resto de razas hace que éstas sean más aptas para este tipo de disciplinas (11). Por esta razón, por su gran influencia en el territorio nacional y por ser la raza que más participa en la base de datos que utilizaremos más tarde (67%) nos centraremos en su Programa de Mejora.

### 2. 3.1 Programa de Mejora del Pura Raza árabe

Para la elaboración del Programa de Mejora del Pura Raza árabe (PRá) se ha respetado, la normativa reciente del 7 de febrero de 2020 por la cual se aprueba el Programa de cría del caballo de PRá y el Programa de difusión de la mejora (12).



*Ilustración 3. Caballos de Pura Raza árabe. A la izquierda Zul-Latz, Foto de la Asociación Española de Criadores de Caballos Árabes (AECCA); a la derecha Licersi; Foto de AECCA*

El objetivo principal de cría y selección de la Asociación Española de Criadores de Caballos Árabes (AECCA) es, además de mantener las características raciales del caballo PRá (Ilustración 3), conseguir que éste sea capaz de destacar en la disciplina ecuestre en la que participe (13).

Este objetivo principal, a su vez se estructura en varios objetivos específicos:

1. Obtención de caballos que mantengan una buena morfología, aptitud y movimientos para el deporte
2. Obtención de caballos que destaquen en la disciplina deportiva Raid
3. Obtención de caballos que puedan destacar también en otros deportes

Para poder llevar a cabo todos estos objetivos se necesita una serie de criterios de selección. Estos criterios vienen dados según unos controles de rendimiento de la morfología y controles de rendimiento en carreras de Raid oficialmente reconocidas.

Para obtener los controles de rendimiento en el Raid se recopilan en cada carrera el tiempo total, el tiempo de marcha, el tiempo de recuperación, el puesto que obtenga cada caballo, si ha sido eliminado o no y las variables ambientales (como por ejemplo el terreno o la lluvia). A partir de estos valores y los registros genealógicos de sus parientes (hayan participado o no en estas pruebas); se predice el valor genético (VG) de los caracteres analizados individuales y combinados (Índice Genético Global, IGG) (14) de cada animal. Dicho de otra forma, con el primer valor se puede seleccionar a cada caballo según cada característica evaluada (tiempo de recuperación, tiempo de marcha...) mientras que con el segundo se ofrece la oportunidad de seleccionar a los caballos de forma global, al combinar los VG de forma ponderada según su importancia para la cría en la disciplina ecuestre de la que se trate. La metodología que hasta ahora se ha ido utilizando para realizar las valoraciones genéticas y por tanto para la obtención de estos parámetros, se detalla en (15).

Tanto a los VG como a los IGG se les aplica un cambio de escala de forma que estén centrados en el valor 100 y tengan una desviación típica de 20. Este cambio de escala se hace para evitar valores genéticos negativos. Las nuevas variables obtenidas ofrecen así valores positivos para todos los individuos, pero con una variabilidad suficiente para poder discernir entre ellos (16). Así pues, a la hora de elegir posibles reproductores en una ganadería los más adecuados serán los que obtengan un valor de IGG por encima de 100. No obstante, si se debe elegir entre caballos con un IGG similar los VG cobran gran importancia, ya que ayudan a seleccionar para la mejora de esa determinada característica (14) como, por ejemplo; tiempo de marcha, haciendo así especial hincapié en la selección para este carácter.

A la valoración genética de cada semental se le otorga también un nivel de precisión que se calcula como la correlación entre el predictor y el verdadero valor (15). Se expresa con valores entre 0 y 1. Cuanto mayor sea, mayor exactitud en la valoración y mayor fiabilidad o seguridad de que ese animal va a repetir el comportamiento deportivo que ha tenido hasta ahora en el futuro y que va a transmitir esas características a su descendencia (16). Según todos estos valores, los animales inscritos en el Libro Genealógico de la raza se pueden clasificar en (13):

- Jóvenes Reproductores Recomendados para Raid: haber participado en una prueba de Raid en el rango de edad establecido por la ley (4-7 años), haber finalizado al menos dos pruebas y haber obtenido un IGG superior a la media (100)
- Reproductores Mejorantes para Raid con prueba de descendencia: IGG superior a la media, precisión igual o mayor de 0,50 y con un mínimo de hijos que hayan obtenido datos de controles de rendimiento
- Reproductor Mejorante para Raid sin prueba de descendencia: IGG superior a la media, precisión igual o mayor de 0,50 y sin control de rendimiento de la descendencia
- Reproductores de Élite: Son animales con categoría de Reproductor Mejorante para más de un objetivo de selección

Así pues, y visto detalladamente, el Programa de Mejora del Caballo de PRá se estructurará en 5 fases (12)

### **1ª Realización de los controles de rendimiento y las actividades relacionadas con la gestión genética de la raza.**

Tal y como se ha indicado anteriormente los controles de rendimiento se obtendrán a partir de pruebas de Raid. En concreto para los equinos jóvenes se obtienen de las llamadas Pruebas de Selección de Caballos Jóvenes organizadas por la AECCA (Ilustración 4). Estas pruebas se consideran un entrenamiento para caballos que se inician en la disciplina del Raid y sirven también para la selección precoz de los individuos. Esto es muy importante ya que, en última instancia provoca una reducción del intervalo generacional y por consiguiente también del tiempo en el que se consiga un progreso.



Ilustración 4. Cartel de la última Prueba de selección para caballos jóvenes AECCA

Además, en cada nacimiento de un nuevo animal se debe realizar un control de paternidad y registrarle al Libro Genealógico de la Raza. A parte de esto, se realizan controles genéticos con objetivo de encontrar y eliminar del Programa de Mejora a aquellos animales portadores de caracteres deletéreos

### **2ª Preparación y envío de la información genealógica y de control de rendimientos disponible para la Raza.**

La información recogida en la fase 1 (rendimientos y genealogía) se remitirá al personal competente de la AECCA para que se incorpore en el Programa de Mejora de la Raza y este así se actualice para la posterior selección

### **3ª Preparación de la información, evaluación genética y envío de resultados**

En esta fase se prepara la información recibida en las fases anteriores para realizar la valoración genética de cada animal (obtención de IGG, VG y precisión). Con los resultados obtenidos, se clasifica a cada animal según las categorías citadas anteriormente y se envían de nuevo a la AECCA

### **4ª Comunicación de los resultados a los propietarios y obtención de la información complementaria para la elaboración del Catálogo de Reproductores**

La AECCA es la encargada de contactar con aquellos propietarios de equinos que hayan obtenido alguna categoría importante para que así estos les envíen más información sobre los animales; como por ejemplo fotografías, si está castrado o no o si sigue vivo.

### **5ª Elaboración, edición, distribución y promoción del Catálogo de Reproductores de la Raza**

Con la información obtenida en el paso anterior se elabora el Catálogo de Reproductores de la Raza. El principal objetivo de este Catálogo es promover y difundir a los buenos reproductores para que se consiga más rápidamente un progreso genético.

Un ejemplo de un equino del Catálogo de Reproductores de 2019 del PRA es el que aparece en la Ilustración 5 (14). Al principio de la ficha aparecen los datos correspondientes a la identificación del caballo.

Tras esto, aparecen los datos de su genealogía, empezando por los padres y seguidamente los abuelos, todos con sus valores de IGG mayores de 100 lo que nos indica que también los antecesores de los abuelos fueron superiores a la media de la población estudiada.

Seguidamente, podemos apreciar los VG para los caracteres tiempo de marcha, tiempo de recuperación y probabilidad de no ser eliminado (caracteres a seleccionar), todos mayores de 100. Un VG alto, como por ejemplo en tiempo de marcha, significa que, el caballo podrá obtener un menor tiempo de recuperación. Los VG vienen acompañados de asteriscos que indican la precisión de la valoración. Finalmente, encontramos el IGG del animal que nos indica que, este equino es globalmente recomendable para estos caracteres.

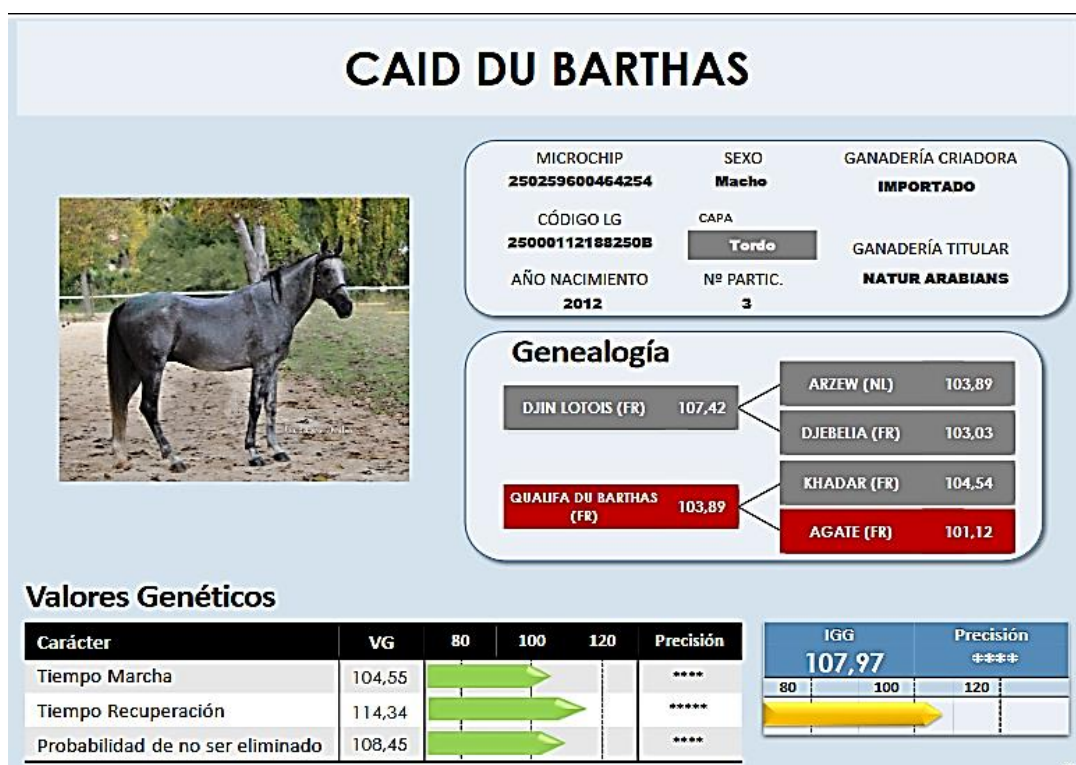


Ilustración 5. Ejemplo de ficha de valoración para Raid del Programa de mejora del PRA. AECCA, Catalogo de reproductores 2019 (14)

### 2.4 Carácter longevidad

En el sector del deporte ecuestre, el rendimiento duradero de los equinos (o longevidad) es una cuestión tanto económica como ética ya que, la mayoría de los caballos llegan a tener una vida deportiva de alrededor de 15 años (5). Por esta razón, en los últimos años, además de la selección para la mejora de caracteres de morfología de los animales, se están incluyendo criterios de selección para alargar la vida de los mismos (17). Por ejemplo, hasta ahora, se ha demostrado que, el año de nacimiento o el ambiente de los caballos tienen efecto sobre la longevidad (18) (19).

Para la selección de cualquier carácter, se debe tener en cuenta su heredabilidad; definida como proporción de variabilidad fenotípica de origen genético. Valores de heredabilidad



iguales a 1 significan que la variabilidad observada sería de origen genético y, por tanto, la longevidad de los individuos dependería directamente de su valor genético. No obstante, si la heredabilidad está en torno a cero no habría grandes diferencias de origen genético en longevidad (15).

La evaluación de la longevidad en la industria del caballo de deporte ha variado a lo largo de los años y ha obtenido valores de heredabilidad muy dispares tal y como se puede observar en la Tabla 5. Esto se puede explicar ya que la heredabilidad depende de la población en estudio que como se ve, es distinta cada vez y se debe estimar cada vez que se incorporan nuevos datos (aunque la disciplina sea la misma) (21).

Tabla 5. Revisión bibliográfica sobre la heredabilidad de la longevidad en equinos

Estudio	Año de recogida de datos	Disciplina	Heredabilidad
Ricard (20) 1997	1972-1991	Salto	0.18
Wallin, L (18) 2003	1968-1982	Doma y salto	0.27 y 0.18
Burns et al (22) 2006	-	Carreras de resistencia	0.15
Ricard et al (21) 2011	1972-2008 en Pura Sangre Inglés	Salto	0.11
Å. Braam a, et al (17) 2011	1971-2008	Doma y salto	0.07 y 0.17
Sabeva I et al. (23) 2012	Yeguas de cría Shagya 1975-2011	-	0.34
Jonsson et al. (24) 2013	Caballos Pura Sangre Suecos 1983-2005	Doma	0.14
Posta et al. (25) 2014	Caballo de Deporte Húngaro de 1996 a 2011	Salto	0.17
Wallin et al. (26) 2016	Pura Sangre Inglés Australiano 1990-2008	Carreras de resistencia	0.12



Seiero et al. (27) 2016	Caballo de salto danés 1981- 1994	Salto	0.11
Velie et al. (28) 2016	Pura Sangre de carreras australiano y de Hong Kong de 2000 a 2011	Carreras de resistencia	0.12
Solé et al. (29) 2016	Pura Raza Español de 2004 a 2014	Doma	0.17

### 3.OBJETIVOS

Así pues, el objetivo principal de este trabajo de Fin de Máster es extender los modelos de supervivencia estándar, que permitan estudiar la longevidad, incorporando la información de parentesco (matriz de parentescos) en datos de competición de caballos de Raid.

En resumen, este trabajo constará principalmente de dos partes: en el Apartado 4 de *Métodos* se explican y detallan los modelos apropiados para conseguir el objetivo descrito anteriormente, además se describen los parámetros genéticos que se van a estimar y el Software específico que se va a utilizar.

Por otro lado, en los Apartados 5 y 6, aplicación a un caso de mejora genética y sus resultados, se ponen en práctica todos los aspectos teóricos descritos anteriormente utilizando como variable tiempo hasta el evento los años en competición.

Por último, en el Apartado 6 se comentarán los resultados y en el Apartado 7 se discutirán y compararán con los obtenidos en otros estudios. Por último, se presentarán las conclusiones pertinentes en el Apartado 8.

## 4. MÉTODOS

### *4.1 El análisis de supervivencia, uso en mejora genética*

Tal y como se ha indicado en apartados anteriores, una mayor longevidad del equino está relacionado con un mejor bienestar del animal y por tanto con un mejor rendimiento y economía (25). No obstante, este carácter no está generalizado como criterio de selección en los programas de cría y por tanto no se realizan evaluaciones genéticas rutinarias para el mismo, aunque en los últimos años está siendo considerado de interés para varias razas y disciplinas (29).

La principal dificultad que se tenía para considerar este carácter era la inclusión de los datos censurados en los estudios. El no hacerlo provocaba resultados sesgados y por tanto conclusiones erradas (22). Una forma común de eludir este problema era utilizar parámetros indirectos de la longevidad, como, por ejemplo, si el animal seguía vivo a cierta edad (30). Este método no resultó ser eficaz debido a la considerable pérdida de información (31).

Por otra parte, se sabe desde hace años que, los métodos de selección de los equinos basados en modelos lineales tradicionales no son apropiados dado el carácter no lineal de la longevidad y su distribución desconocida (32). Además, ésta, en la mayoría de las veces viene influida a su vez por caracteres tiempo dependientes; por ejemplo, el año de competición.

Con objetivo de intentar paliar todos estos problemas se empezó a utilizar el análisis de supervivencia, método estadístico originalmente creado para la medicina que combina la información de datos censurados y no censurados y ofrece un tratamiento estadístico adecuado a la no linealidad de los datos (33). En equinos, se empezó a utilizar en la década de los 80, donde se estimó la longevidad de los sementales, pero a través de la de sus crías (34).

A lo largo de los años, esta técnica en este ámbito se fue desarrollando y ajustando para aplicaciones a mayor escala, lo que finalmente dio lugar a que Ducrocq (35) desarrollase el Software Survival Kit, y que recientemente Meszaros (36) crease la interfaz de R a través de la cual se puede incorporar la información genética de los animales y con la que se ha desarrollado el ejemplo práctico del Apartado 6.

#### 4.2 Conceptos básicos de Análisis de supervivencia

En el análisis de supervivencia se utiliza como variable principal el tiempo transcurrido entre un evento inicial y un evento final que se espera que ocurra, dados unos factores de riesgo. Por ejemplo, en nuestro caso, se utilizará como variable principal el tiempo en años desde que los caballos empiezan a competir hasta que dejan de hacerlo. A esta variable se la denota como  $T$  y uno de los principales intereses reside en el análisis de su función de supervivencia definida como:

$$S(t) = P(T > t) \text{ con } \{t \in [0, \infty)\} \quad [1]$$

La función [1] se refiere a la probabilidad de presentar el evento final después del tiempo  $t$ .

La función de densidad de  $T$  [2] normalmente se la considera como la probabilidad de que en el tiempo  $t$  ocurra el evento final y viene dada por el siguiente límite:

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq t + \Delta t)}{\Delta t} \quad [2]$$

la función de densidad de la variable principal cumple las siguientes relaciones con la función de distribución y la función de supervivencia:

$$f(t) = \frac{d}{dt}F(t) = -\frac{d}{dt}S(t) \quad [3]$$

así que

$$S(t) = \int_t^{\infty} f(y) dy \quad [4]$$

La función de riesgo [5]; está dada por la expresión:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P(t \leq T \leq t + \Delta t | T \geq t) \quad [5]$$

Esta función es no negativa, y se puede reescribir como:

$$h(t) = \frac{1}{P(T \geq t)} \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P(t \leq T \leq t + \Delta t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} \log[S(t)] \quad [6]$$

Así pues, se puede observar la relación que existe entre la función de riesgo y la función de supervivencia, cuando una aumenta, la otra disminuye.

Además, se cumplen las siguientes relaciones entre la función de riesgo, la de supervivencia y de la variable T:

$$S(t) = \exp \left\{ - \int_0^t h(y) dy \right\} \quad [7]$$

La función de riesgo acumulada es no negativa y creciente y se define como

$$H(t) = \int_0^t h(y) dy \quad [8]$$

así que

$$H(t) = -\log[S(t)] \quad [9]$$

$$S(t) = \exp\{-H(t)\} \quad [10]$$

En la gran mayoría de los casos, se desconocen las características exactas de la función de supervivencia o la función de densidad, pero se dispone de cierta información como por ejemplo la forma en la que cambia la tasa de eventos  $h(t)$  con el tiempo. Por lo tanto, los modelos para el análisis de la supervivencia suelen construirse a partir de una función de riesgo, y así se mide el riesgo de evento en un individuo en tiempo  $t$ .

Por tanto, un modelo de supervivencia en su forma general viene definido como:

$$h_i(t, x_i) = \exp(x_i \beta) h_0(t) \quad [11]$$

Donde  $h_i(t, x_i)$  es la función de riesgo para el animal  $i$ ,  $h_0(t)$  es la función de riesgo base para el individuo de referencia,  $x_i$  es el vector de variables explicativas y  $\beta$  es el vector de coeficientes

Los modelos de regresión de riesgos proporcionales son de los más usados en el ámbito de la mejora genética debido a su simplicidad. Se llaman así, porque el riesgo basal puede tomar cualquier forma, pero el riesgo entre los individuos debe ser constante e independiente a  $t$ . Es decir, el riesgo de, por ejemplo, los animales A y B, debe adecuarse a la siguiente expresión:  $\exp\{X_A - X_B\}' \beta$  conocida como la hipótesis de riesgos proporcionales o como la razón de riesgos, constante e independiente del tiempo (37).

Los modelos de riesgos proporcionales ya sean paramétricos o semiparamétricos, pueden ampliarse para incluir efectos aleatorios como el propio individuo. Éste está determinado por su genética, que viene dada por el pedigrí (matriz de parentescos), y por su ambiente (medidas repetidas de cada individuo). Estos modelos mixtos de supervivencia se denominan modelos de fragilidad y se explicarán a continuación, centrándonos en la fragilidad de origen genético (33).

#### **4.3 Modelos de fragilidad**

En aplicaciones reales, es muy común analizar datos de un grupo de individuos que comparten los mismos factores de riesgo pero con distintos tiempos hasta el evento (con distintos valores de  $T$ ). Estas variaciones las intenta explicar el componente de la fragilidad, introducido por Clayton (38).

De manera general, la fragilidad se define como efecto aleatorio inobservable que introduce variabilidad en una población. Fragilidades observables en la población por encima de 0 indican un mayor “predisposición” de presentar el evento, mientras que, valores menores que 0 indican un menor “predisposición”. Dicho con un ejemplo, supongamos que, un equino obtiene un valor de fragilidad en un análisis para años en competición de 8, esto indicaría, de forma general que, este caballo, competiría menos años que el resto; de forma contraria, que otro equino obtenga valores de -2 indicaría así que, sería menos frágil y por tanto competiría mas años. Así pues, es lógico pensar que, en cualquier momento en que se observen los años en competición de una muestra de individuos, los que siguen compitiendo serán menos frágiles que la población correspondiente de la que se ha extraído la muestra.

Estos modelos nos permiten, modelar de forma más exacta el riesgo. Ignorar esto entre la población nos puede llevar a: sesgos en las distribuciones marginales y las covariables observadas, la subestimación del efecto de estas y lo que es más preocupante, una subestimación de la esperanza de vida (39).

En los modelos más comunes, se supone que un efecto aleatorio está compuesto de distintos niveles extraídos de una población de valores posibles, en la que los niveles reales son representativos de esa población. Por esta razón no es realista representarlos como efectos fijos ya que introduciría en el modelo un gran número de parámetros desconocidos. Al tomarlos como efectos aleatorios, sólo hay un parámetro a estimar, la varianza de su supuesta distribución. Por ejemplo, los valores de fragilidad desconocidos que se atribuyen a los individuos en un estudio se extraen esencialmente de un gran número de valores posibles, por eso se le considera un efecto aleatorio.

En concreto, en modelos en los que se estudia para la mejora genética de un carácter, cada individuo posee un valor genético para la fragilidad,  $u$ , que se asume como una realización de la distribución de valores genéticos con una varianza genética  $\sigma_u^2$ . Además, entre parientes existen covarianzas de acuerdo con el porcentaje de genes que comparten (16) de manera que, si el animal 1 y el 2 tienen un hijo 3, entonces la varianza genética de cada uno de los 3 individuos se corresponderá con la varianza de su distribución  $\sigma_u^2$ , y la covarianza entre 3 y cada uno de sus padres sería  $\frac{1}{2}\sigma_u^2$  al compartir con cada uno de ellos la mitad de los genes. Así, la notación para las covarianzas entre todos los individuos se podría representar como el producto entre una matriz de parentescos  $\mathbf{A}$  y  $\sigma_u^2$ , siendo esta matriz para este ejemplo:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0.5 \\ 0 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{bmatrix}$$

Así pues, para cualquier carácter, la fragilidad genética en estos modelos vendrá dada por el efecto aleatorio de la genética del propio individuo que, es heredada de sus padres. De este modo, se podrán hacer inferencias sobre los valores genéticos de los animales haciendo uso de la información de la matriz  $\mathbf{A}$  que se construye a partir de la información del pedigrí. A estos modelos se les denomina modelos de fragilidad individuales y se estimaran a continuación

#### 4.4 Estimación de los modelos de fragilidad individual

La variación entre los tiempos hasta el evento de los individuos se va a modelizar suponiendo que cada individuo tiene su propio valor de fragilidad genética. Esta fragilidad se supone que actúa de manera multiplicativa sobre el riesgo de presentar el evento.

Así pues, considerando un modelo de riesgos proporcionales para el riesgo de que ocurra el evento en el tiempo  $t$ , en un individuo con una función de riesgo proporcional que depende de los valores de  $p$  variables explicativas  $X_1, X_2, \dots, X_p$  y de una función de riesgo base desconocida  $h_0(t)$ . La función de riesgo para el  $i$ -ésimo de los  $n$  individuos es:

$$h_i(t) = \exp(x_i \beta) h_0(t) \quad [12]$$

Donde  $x_i$  es el vector de valores de las variables explicativas para el  $i$ ésimo individuo y  $\beta$  es el vector de coeficientes  $\beta_1, \beta_2, \dots, \beta_p$ .

Seguidamente, a la anterior expresión se le añade el término de fragilidad  $Z_i$  no negativa que oscila entre el 0 y el 1

$$h_i(t) = z_i \exp(x_i \beta) h_0(t) \quad [13]$$

Un valor de  $Z_i = 1$  nos devuelve al modelo estándar de la ecuación [12], que corresponde a la situación en la que no hay efectos de fragilidad.

Generalmente, es más conveniente trabajar con una representación alternativa del efecto de fragilidad obtenida de la siguiente manera  $Z_i = \exp(u_i)$ . Entonces, el modelo de la ecuación [13] puede ser reescrito de nuevo como sigue:

$$h_i(t) = \exp(x_i \beta + u_i) h_0(t) \quad [14]$$

En este modelo  $u_i$  es una variable aleatoria que puede tomar cualquier valor, pero, el caso de  $u_i = 0$  es el caso de  $z_i = 1$  donde no existía fragilidad.

Los  $u_i$  de la ecuación [14] se consideran como los valores observados, de las  $n$  variables aleatorias independientes e idénticamente distribuidas (distribución normal)  $U_i, i = 1, 2, \dots, n$ , donde se supone que cada variable aleatoria,  $U_i$ , tiene una distribución de probabilidad común. Así pues, aunque las fragilidades individuales de un grupo de individuos pueden ser todas diferentes, se extraen de la misma distribución de probabilidad, tal y como se comentó anteriormente.

Para la correcta estimación de los parámetros se debe seleccionar una distribución para  $u_i$ . Tradicionalmente, se ha utilizado más la distribución gamma debido a su flexibilidad y conveniencia matemática, No obstante, se sabe que una gran mayoría de las covariables de la mejora de animales se distribuyen de forma normal o aproximada a una normal. Por lo tanto, y para que se tengan en cuenta las relaciones genéticas entre los animales, en la matriz de parentesco que se va a introducir, se debe suponer una distribución normal multivariante (40).

Además de esto, se debe decidir si se van a utilizar modelos paramétricos o semiparamétricos.

#### ***4.5 Cox vs Weibull en modelos de mejora genética***

El modelo de Cox se ha utilizado ampliamente en el ámbito de la mejora genética animal, ya que no requiere un conocimiento previo sobre la distribución del tiempo hasta evento que, en muchos casos es desconocida. Sin embargo, cuando un modelo de Cox contiene varias covariables dependientes del tiempo con un gran número de niveles, la estimación de los coeficientes  $\beta$  puede conducir a cálculos bastante complicados (41). Además, cuando es el caso de una escala de tiempo discreta, a menudo se producen muchos eventos en el mismo tiempo, lo que da lugar a errores. En ese caso, el modelo de Cox ya no es válido ya que resulta ser inadecuado y exigente desde el punto de vista computacional (36).

Otro problema que presentan los modelos de Cox es que no pueden aplicarse a conjuntos de datos grandes como los que se utilizan en los programas de mejora (32). Pero, cuando la función de riesgo base tiene una forma paramétrica conocida, la estimación de  $\beta$  y  $h_0(t)$  es generalmente más fácil (42). Así pues, si pensamos de forma intuitiva, para nuestra base de datos la distribución paramétrica más apropiada para nuestro modelo es una función de riesgo asociado a una distribución Weibull. En un modelo de regresión de Weibull, la función de riesgo base se define como:

$$h_0(t) = \lambda\rho(\lambda t)^{\rho-1} \quad [15]$$

Tal y como se puede apreciar, viene descrita por los parámetros  $\lambda$  y  $\rho$  que definen la escala y la forma de dicha función.

En la mayoría de los casos la elección de la distribución Weibull viene determinada por la gran simplicidad de su función de supervivencia:



$$S_0(t) = \exp(-(\lambda t)^\rho) \quad [16]$$

Y su flexibilidad; ya que esta distribución permite que la función de riesgo pueda comportarse de manera constante ( $\rho = 1$ ), creciente ( $\rho > 1$ ) o decreciente ( $\rho < 1$ ).

En conclusión, para el ejemplo práctico que nos concierne se ha elegido un modelo de fragilidad individual con distribución de riesgo conocida y dada por la distribución Weibull y una distribución multivariante para la fragilidad.

Para estimar los parámetros del modelo junto con los hiperparámetros de su distribución, se han sugerido tanto técnicas de Monte Carlo (43) como de aproximación de Laplace de la densidad a posteriori de los hiperparámetros (37), Este último enfoque es el que se utilizará en el caso práctico ya que ha demostrado ser muy eficiente en bases de datos de gran tamaño (44). Una información más detallada sobre la estimación de los parámetros a través del método de máxima verosimilitud de los modelos de fragilidad individuales fue descrita por Collet (33)

#### 4.6 Parámetros genéticos

En el contexto de mejora genética, el parámetro de mayor interés es la heredabilidad. Esta viene definida como la proporción de la variabilidad del carácter que es de origen genético. De forma general, viene definida por la siguiente expresión (15)

$$h^2 = \frac{\sigma_u^2}{\sigma_p^2} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_{ambiente}^2} \quad [17]$$

Donde  $h^2$  denota la heredabilidad,  $\sigma_u^2$  es la varianza de origen genético y  $\sigma_p^2$  es la varianza fenotípica. Por su parte, el fenotipo se divide en genotipo ( $\sigma_u^2$ , varianza genotípica) y ambiente ( $\sigma_{ambiente}^2$ , varianza ambiental).

De la fórmula [17] se puede deducir que, los valores de la heredabilidad se encuentran en el rango de cero a uno. Además, si ésta vale 1, toda la variabilidad observada sería de origen genético de forma que lo que se vería fenotípicamente en los individuos sería una medida directa de su valor genético. Por el contrario, si toma el valor de 0 querría decir que la variabilidad es debida exclusivamente al ambiente (16):

Existen valores de heredabilidad en función del tipo de carácter:

- Alta  $h^2 > 0.40$ . Características relacionadas con la morfología del animal, por ejemplo, la alzada

- Moderada  $0.15 < h^2 < 0.40$ . Estos valores son los mas comunes como por ejemplo el carácter producción de leche
- $0,15 > h^2$ . Baja para caracteres reproductivos como por ejemplo el tiempo entre partos.

Yazdi et al. (45) definieron la heredabilidad para el modelo que nos concierne de la siguiente manera:

$$h^2 = \frac{\sigma_a^2}{\frac{1}{P} + \sigma_a^2} \quad [18]$$

En la fórmula,  $h^2$  es la heredabilidad basada en el modelo final,  $\sigma_a^2$  es la varianza genética muestral y P se refiere a la proporción de datos estudiados (sin censura). Otro parámetro importante es la precisión de los valores genéticos obtenidos de los individuos que se calculó de la siguiente manera (29);

$$P = \sqrt{\frac{VEP}{\sigma_a^2}} \quad [19]$$

Donde P es la precisión, VEP funciona como la varianza del error de predicción; y  $\sigma_a^2$  la varianza genética.

#### **4.7 Survival Kit**

*Paquete SurvKit4R(36)*

##### **Función SKit4R**

La función SKit4R llama a las otras funciones del Survival Kit que crean los archivos de parámetros y clasifican los conjuntos de datos para la ejecución del Software. Esta función incluye todas las posibilidades del Survival Kit, pero normalmente sólo se utiliza parte de ellas.

Su sintaxis general es:

```
SKit4R (,inputData = "", inputPedigree = "", time = "life", idRec = "idnum", censName = "cens", censValue = 0, truncate = "", effects = c("", ""), effectType = c("", ""), pedigreeEffect = "", timeDepEffects = c("", ""), timeDepEffectsType = c("", ""), classEffects = c("", ""), outputEffects = c("", ""), program = "", discrete = FALSE, strata = "", origin = "", strataSort = 0, ite_quasi = -1, model = c("", ""), std_error
```

= TRUE, random = "", correlation = c("", "", ""), test = c("", ""), baseline = FALSE, Kaplan = FALSE, moments = FALSE, survivalOptions = "", residual = "",

La descripción de los principales argumentos se muestra a continuación

Tabla 6 Resumen de los argumentos de la función Skit4R

Resumen de los argumentos de la función Skit4R	
inputData	Nombre del archivo a analizar
inputPedigree	Nombre del pedigrí que se utilizara (matriz de parentesco)
time	Nombre de la variable que contiene el tiempo de supervivencia ya calculado, o un conjunto de 3 variables que contengan el nombre del tiempo de supervivencia, el comienzo y el final de la observación dados como fechas. En este último caso, el tiempo de supervivencia total se calcula dentro del programa. Sólo se permiten variables de tiempo enteras. Si se especifica "life" como el nombre de la variable de tiempo, el argumento de la función podría ser omitido.
idRec	Nombre de la variable que tiene el número de identificación único para cada registro. Si el usuario especifica "idnum" como nombre de la variable ID en los efectos, el argumento de la función podría ser omitido.
censName	Nombre de la variable que contiene la censura. Si el usuario especifica "cens", el argumento de la función podría ser omitido.
censValue	Número que identifica las observaciones censuradas en censName, todo lo demás se considera como un registro no censurado. Si el usuario especifica 0 como el valor que denota los registros censurados, el argumento de la función podría ser omitido
effects	Vector con el nombre de las variables que se encuentran en la base de datos
effectType	Vector que determina como son las variables introducidas en effects; es decir, categóricas (class), continuas(continuous) o fechas (date)

pedigreeEffect	Nombre del efecto aleatorio asociado al pedigrí
timeDepEffects	Nombre de la variable dependiente del tiempo
timeDepEffectsType	Tipo de la variable tiempo dependiente (class, continuous, date)
classEffects	Vector que contiene las variables de tipo class
program	Nombre del programa que se utiliza, Cox o Weibull
discrete	Si se ajusta a TRUE, indica que la escala de tiempo se expresa en unas pocas unidades discretas, por lo que se utiliza el modelo de datos agrupados
Itr.quasi	Especifica el número de iteraciones cuasi-newton antes de cambiar a pasos completos de Newton. En caso de 0 iteraciones el programa comenzará con el Newton completo
model	Especifica las variables independientes que afectan al life
random	Nombre de la variable aleatoria. Esta parte es muy importante para poder introducir el efecto aleatorio que nos concierne, a parte del nombre de la variable se debe escribir también el nombre de su distribución, recordamos que para este caso es multinormal
test	Test de los efectos incluidos en el modelo para SEQUENTIAL LAST, comparan el modelo completo con los que excluyen una variable a la vez

Tal y como se puede observar en la Tabla 6 de resumen de los argumentos de la función SKit4R, este software es capaz de manejar grandes cantidades de datos, contabilizar censurados por la derecha o por la izquierda, introducir variables dependientes de tiempo, estratificadas, y estimar componentes de varianza como la varianza genética de los individuos gracias a la introducción del pedigrí (matriz de parentescos) y el argumento de efectos aleatorios

Las covariables dependientes del tiempo las asume como constantes a lo largo del mismo por lo que, los registros con efectos dependientes del tiempo deben incluir el tiempo de cambio y el nuevo valor de la covariable, tantas veces como la misma cambie (Ver Apéndice 1). Así, como se ha indicado en apartados anteriores se respeta la asunción de riesgos proporcionales.

Los efectos aleatorios que se quieran introducir en el modelo, pueden ser términos de interacción con otras covariables y/o dependientes del tiempo. En nuestro caso, el efecto aleatorio introducido es el individuo.

La estimación de hiperparámetros como la varianza genética, teóricamente, se realizan a partir de su densidad posterior marginal (37). No obstante, cuando se asume una distribución normal multivariante, como es en este caso, se utiliza el método de la integración de Laplace. Entonces, suponiendo los hiperparámetros conocidos, las estimaciones de todos los demás parámetros se obtienen a través de la densidad posterior conjunta. Esta maximización se realiza utilizando una aproximación de memoria limitada cuasi-Newton (46).

En modelos y aplicaciones grandes que implican muchos efectos aleatorios correlacionados, el enfoque citado anteriormente converge lentamente y entonces se utiliza un algoritmo completo de Newton-Raphson para garantizar la convergencia en un número menor de iteraciones (36).

A parte de la varianza genética y el valor genético de cada individuo, el Software proporciona también la media global de este factor y la desviación estándar de cada registro, dando una imagen más precisa de la densidad posterior aproximada del efecto aleatorio. Además, devuelve la parte estadística básica de las variables introducidas en el modelo.

## **5.DESCRIPCIÓN DE LA BASE DE DATOS**

La base de datos de la que se dispone para la realización de este trabajo reúne aquellos equinos que participaron en la disciplina ecuestre del Raid. En total, se utilizaron 6136 registros de carreras de resistencia organizadas por la Real Federación Hípica Española, celebradas entre el año 2000 y el 2018 de 1419 caballos distintos, nacidos entre los años 1981 y 2012. Por lo tanto, es un estudio longitudinal de una base de datos recogida retrospectivamente.

La variable que se ha utilizado para determinar la longevidad de la carrera competitiva de los caballos ha sido los años en competición de los mismos.

Los caballos que siguieron compitiendo en el último año de estudio (2018) se consideran censurados por la derecha. El pedigrí usado para el análisis contiene los antecesores de todos los equinos incluidos en la base de datos lo que hace un total de 10869 registros.

Para la recopilación de las características de los animales, se recogieron todas las variables en cada competición ya que, cabe destacar que, normalmente se da más de una al año.

Las variables principales que se utilizarán a lo largo de este TFM son:

- Idnum: Variable que indica el número de identificación del caballo y que será la que se utilizará como variable aleatoria en el modelo
- sexo: variable categórica que se refiere al sexo del animal en la primera carrera; puede tomar valor de 1 si es macho; 2 hembra y 3 castrado. Esta variable es tiempo dependiente ya que algunos de los equinos machos pasan a ser castrados.
- edad: variable continua que representa la edad en la primera carrera en la que los caballos compitieron (de 5 a 24).
- Placing: variable categórica que se refiere a si han completado los equinos o no la primera carrera. Esta variable puede tomar los valores 1 (eliminado) y 2 (carrera completada)
- La variable continua Ncat viene definida según el número de veces que el caballo haya corrido en las diferentes categorías. Según el kilometraje de estas carreras Ncat se divide en Ncat1, Ncat2, Ncat3 y Ncat 4; sus diferencias se ilustran en la Tabla 7

Tabla 7. Definición de la variable Ncat

Ncat	Kilometraje	Categoría
Ncat1	Hasta 102 km	CEN1, CEI1, CEN0, PROMOCIÓN
Ncat2	Hasta 140 km	CEN2, CEI2, ECTOES, PEWEG, ECTOESY/J, CH-M-E-JY, CEI JY**
Ncat3	Hasta 150 km	CEI3, CEN3, ECHEU
Ncat4	Hasta 160 km	ECTOES, CEI4, CEN4, WEG

Variables importantes de la base de datos del pedigrí o información genealógica

- npadre: variable que denota el número de identificación de los padres de los equinos
- nmadre: variable que denota el número de identificación de las madres de los equinos

Hay que mencionar que el pedigrí contiene toda la genealogía del animal hasta el último ancestro conocido en su Libro Genealógico.

## 6.APLICACIÓN A UN CASO DE MEJORA GENÉTICA

A continuación, se muestra el uso y la aplicación del Survival Kit sobre las bases de datos explicadas en el Apartado 4 para posteriormente estudiar la posible interpretación de los resultados obtenidos.

Los factores que afectan a los años en competición de los caballos fueron estudiados según el siguiente modelo:

$$h(t, x) = h_0(t) \exp \left\{ \left( \text{sexo}_i(t) \beta_1 + \text{placing}_j \beta_2 + \text{Ncat1}_k \beta_3 + \text{Ncat2}_l \beta_4 + \text{Ncat3}_m \beta_5 + \text{Ncat4}_n \beta_6 + \text{edad}_o \beta_7 + \text{caballo}_p(t) \beta_8 \right) \right\}$$

Donde:

- Sexo: nivel 1, machos, el nivel 2, hembras y el nivel 3, castrados. Es una variable tiempo dependiente ya que, a lo largo de todo el periodo de observación existen machos que pasan a ser castrados
- Placing: 1 si los equinos han sido eliminados o, 2, si no han sido eliminados
- Ncat1: frecuencia en categorías de menos de 130 km (Tabla 7)
- Ncat2: frecuencia en categorías de entre 130- 139 km (Tabla 7)
- Ncat3: frecuencia en categorías de 150 km (Tabla 7)
- Ncat4: frecuencia en categorías de 160 km (Tabla 7)
- Edad: edad en años del caballo en la primera carrera
- Caballo: efecto aleatorio dependiente del tiempo

Estas covariables se escogieron debido a que, en otros estudios de caballos de Raid las introdujeron en el modelo resultando ser influyentes en el rendimiento general de los equinos (26) (29).

### 6.1 Estadística descriptiva

Antes de ver los resultados obtenidos con el Software utilizado, se realizarán diferentes gráficas para observar cómo se comportan las covariables con respecto a la variable principal y además se aportarán las medidas de centralización y dispersión de las mismas.

Tabla 8. mediana rango intercuartílico (IQR) de Ncat1, Ncat2, Ncat3, Ncat4 y edad. Para sexo y Placing se dan los porcentajes de sus niveles

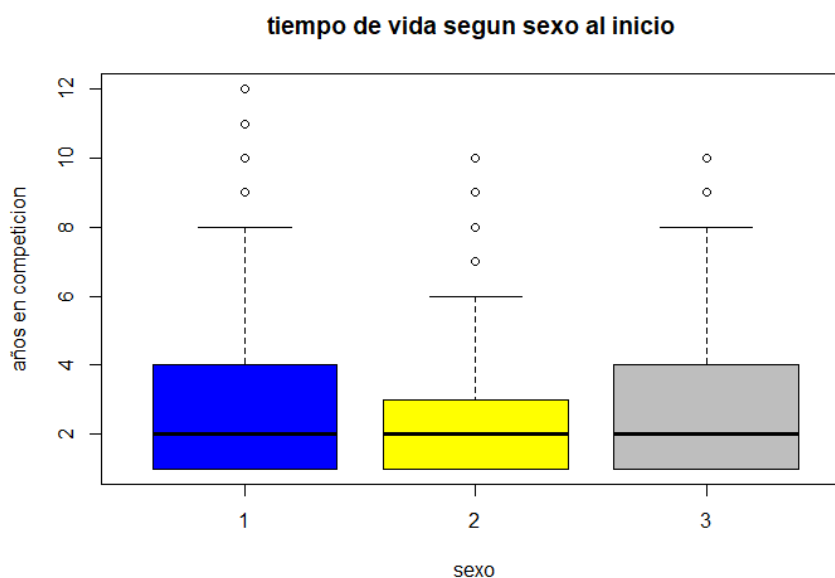
Covariable	
Ncat1	Mediana: 1 IQR:2
Ncat2	Mediana:0 IQR:1
Ncat3	Mediana: 0 IQR:0
Ncat4	Mediana:0 IQR:0
edad	Mediana:8 IQR:2
Sexo	machos:39% hembras:42% castrados:19%
Placing	eliminados:44% no eliminados: 56%

En la Tabla 8 se muestra un resumen de las covariables introducidas en el modelo. Para Ncat1, encontramos una mediana de 1 con un valor de IQR de 2. Por su parte, Ncat2 presenta una mediana situada ya en el 0 con un IQR de 1. Esto nos indica que, los equinos compiten menos veces en Ncat2 que en Ncat1 y que su dispersión también es menor. Seguidamente, para Ncat3 la mediana es de 0 y un IQR del mismo valor. Esta tendencia se mantiene para Ncat4 que presenta la misma mediana y el mismo valor de IQR. Esto nos indica que, a lo largo de los años de competición de los caballos, éstos suelen participar más en las categorías de menos kilómetros mientras que, las categorías más duras (Ncat3, Ncat4) suelen disputarlas pocos equinos.



Si nos centramos en la edad; esta se refiere a la edad en la que el caballo empezó a competir, como vemos, la mediana se encuentra en torno a 8 con un valor de IQR de 2. Esto nos indica que, ya en la mitad de los casos, la edad en la primera carrera es menor que 8 y el valor del IQR nos muestra que el 50% de los datos se encuentran en un rango de 2 indicando así la poca dispersión.

En la Gráfica 2, se puede observar la relación que existe entre los años en competición y el sexo al inicio de la carrera competitiva. Tal y como se puede observar, en media, no parece que exista una diferencia en los años en competición según esta covariable.



Gráfica 2. Boxplot de años en competición con respecto al sexo

Los años en competición según la edad en la primera carrera se pueden observar en la Gráfica 3. En esta podemos observar que, más años en competición, están relacionados con una menor edad en la primera carrera. El punto máximo lo encontramos a la edad de 6 años en la primera carrera ya que está relacionado con 12 años en competición. A partir de edades de 14 años en la primera carrera los valores de años en competición son bajos y cercanos a 1 aunque también cabe destacar que, se tienen menos registros.



## 6.2 Resultados

### 6.2.1. Efectos de los factores de riesgo

Tabla 9. Efectos de la variable categórica Placing, tiempo independiente, en el Riesgo Relativo (RR) para la variable principal años en competición

-	Niveles	Estimates ( $\beta$ )	Err.St	HR	95% CI HR	Total animales
<b>Placing</b>	1	0.28	0.08	1.32	1.15-1.48	445
	2	0	-	1	0	868
<b>Sexo</b>	1	0.13	0.09	1.13	0.96-1.31	510
	2	0	-	1.00	0	558
	3	0	0.11	1.00	0.79-1.21	245
<b>Ncat1</b>	1	-0.31	0.03	0.73	0.68-0.78	1313
<b>Ncat2</b>	1	-0.39	0.03	0.68	0.74-0.62	1313
<b>Ncat3</b>	1	-0.33	0.06	0.72	0.83-0.61	1313
<b>Ncat4</b>	1	-0.05	0.12	0.95	1.19-0.71	1313
<b>edad</b>	1	0.30	0.08	1.36	1.21-1.51	1313

La Tabla 9 muestra el parámetro  $\beta$ , el error estándar (Err.St), el Hazard Ratio (HR), el intervalo de confianza al 95 % para ese HR y el total de los animales de las variables introducidas en el modelo.

La estimación puntual del HR para el nivel 1 del Placing es de 1.32 con un intervalo de confianza (IC) al 95% que va desde 1.15 hasta 1.48. Por tanto, se puede decir que, individuos con nivel 1 (eliminados) en Placing en la primera carrera tienen 1.32 veces más riesgo de finalizar su carrera competitiva que aquellos con nivel 2 (no eliminados), bajo mismas covariables basales. Este efecto ya se podía observar anteriormente, en la parte de la descriptiva de esta covariable (Gráfica 4) resultando finalmente influyente en la longevidad de los equinos de competición.

Tal y como se puede apreciar en los intervalos de confianza de los niveles 1 (machos) y 3 (castrados) no se puede concluir que el sexo sea influyente en la variable años en competición de los caballos. Este efecto, es congruente con lo observado a nivel

descriptivo, que, aunque se referían al sexo al nacimiento, se podía observar que, el comportamiento de los tres grupos era prácticamente idéntico (Gráfica 2).

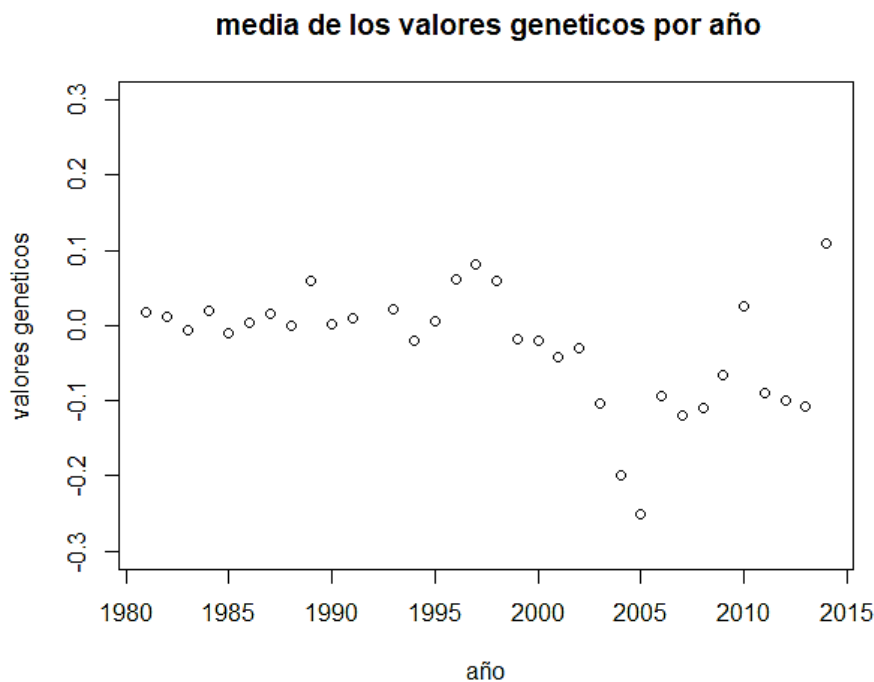
Fijándonos en la variable Ncat1 podemos decir que, si se aumenta en una unidad el participar en este tipo de categorías (Tabla 7) el riesgo de que se termine la carrera competitiva se reduce en un 27%. De la misma forma, para la variable Ncat2 y Ncat3 según se aumenta en una unidad las carreras en estas categorías se reduce en un 32 % y en un 28% respectivamente el riesgo de terminar con la carrera competitiva de los animales.

Por último, de la variable edad de inicio se puede concluir que, aumentar en una unidad la edad de los animales en la primera carrera hace que el riesgo de no competir más años aumente un 36%. Esto se puede observar también en la Gráfica 3 donde a partir del máximo de años se observa esa tendencia decreciente en los años en competición

### **6.2.2 Determinación genética de la longevidad productiva**

El componente de la varianza animal estimada para la longevidad deportiva en este análisis tomó un valor de 0.90 (SD= 0.01). El valor de la heredabilidad de este análisis se ha calculado utilizando la fórmula [18], teniendo en cuenta que hay un 93% de registros no censurados, toma un valor de 0.45.

En la Gráfica 5 se muestra, la evolución de la media de los valores genéticos de los individuos según el año de nacimiento. Tal y como se puede observar, desde los años 1980 a 1995 los valores genéticos de los individuos oscilaban alrededor del 0 (valores genéticos medios de los animales fundadores) mientras que, a partir de este año en adelante aumenta la dispersión de los valores alejándose así cada vez más del cero llegando al punto máximo negativo (-0.2502) en el año 2005 y al punto máximo positivo (0.1090) alrededor del 2014.



Gráfica 5. Media de los valores genéticos de cada individuo según el año de nacimiento

Por su parte, en la Tabla 10, se muestran 5 valores genéticos máximos y 5 mínimos. Tal y como se puede observar, el valor más negativo es de - 2.36 mientras que el valor más positivo es de 1.07. El primero corresponde a un caballo nacido en 2001, donde la media de los registros es ligeramente inferior a cero por lo que se encuentra muy alejado de esta. Esto nos indica que, este equino sería un buen reproductor para el carácter longevidad ya que su fragilidad observada se encuentra muy por debajo del 0 (valor de no fragilidad), indicando así menor predisposición al evento. De la misma forma, el valor mayor data de 2002 donde, la media es muy parecida a la del anterior año (cercana al 0) por lo que, en este caso este equino no sería un buen reproductor para el valor longevidad en este año debido a una fragilidad alta y por encima del 0 lo que nos indica más predisposición al evento. Dado que los fundadores tienen media 0, el hecho de que los valores negativos sean mucho mayores que los positivos en valor absoluto, parece indicar que ha habido un efecto importante de selección a favor de longevidad.

Tabla 10 Valores genéticos máximos negativos (izq) y positivos (drcha)

Nº caballo	V.G	Año de nacimiento	Nº caballo	VG	Año de nacimiento
8976	-2.36	2001	9160	1.07	2002
9719	-2.30	2005	9136	1.06	2002
9648	-2.28	2005	9066	1.03	2001
9106	-2.26	2001	8432	1.02	1998
9587	-2.25	2004	7332	1.01	1991

Seguidamente se calculó la precisión para los valores genéticos obtenidos del análisis según la fórmula [19]. Se obtuvo una precisión mayor del 60% en el 98 % de los equinos analizados.

## 7.DISCUSION

Que un caballo obtenga buenos rendimientos deportivos a lo largo de su vida competitiva está relacionado con su bienestar, pero también con su valor económico (29) por lo que, que no compita, o que lo haga poco, repercute en su longevidad (17). Así pues, la longevidad deportiva, definida como años en competición, asociada a factores de riesgo y protectores, podría ser una herramienta eficaz para la mejora del rendimiento de los equinos.

Una de las variables más comunes que afectan a la longevidad deportiva es la edad en la primera carrera. Esto se explica debido a que, cuanto antes empiece un caballo a competir, más tiempo permanecerá en competición; es decir, tendrá más años potenciales para ello. De hecho, se ha probado que, los caballos que comienzan temprano su carrera tienen más probabilidades de ser más exitosos en los deportes, mientras que, aquellos que lo hacen más tarde, tienden a ver mermada su longitud competitiva (25). Esto también se demostró en disciplinas como la doma, donde el comienzo a una edad muy temprana o por consiguiente demasiado tardía tiene un impacto negativo en la carrera competitiva de los equinos (29)

Un rendimiento bajo en los equinos, como por ejemplo que sean eliminados de las carreras, en la disciplina de doma, está relacionado con una retirada temprana de los mismos de las competiciones (29) y por consiguiente con una vida competitiva menor. En nuestro estudio, el porcentaje de eliminados en la primera carrera fue del 44%, tal y

como se puede observar en la Tabla 8 que fue muy similar a otros publicados anteriormente; como por ejemplo el de Nagy et al. (47). En este estudio se recogieron datos sobre 30.741 caballos de 47 naciones, que participaron en competiciones de Raid, de los cuales el 49 % resultó ser eliminado.

La eliminación de un caballo en una carrera es siempre por su propio bienestar, es decir, son eliminados para que no desarrollen lesiones potencialmente graves (48). Por lo tanto, la disminución del riesgo de eliminación puede aumentar las preocupaciones sobre si los caballos están suficientemente controlados y eliminados en una etapa temprana antes de que su condición sea más grave (47). Una de las razones más comunes de eliminación de caballos en los Raid son las razones metabólicas como por ejemplo la sudoración prolongada que puede causar un déficit de electrolitos que provoque debilidad, calambres musculares, arritmias, agotamiento y por tanto un menor rendimiento deportivo (49). Otra razón es la cojera que, está normalmente producida por lesiones por cargas repetitivas. Es decir, por muchos entrenamientos y más largos o por recorrer distancias más largas de lo adecuado. Esto se agrava también cuando no se eliminan de los Raid a los caballos con lesiones más leves, pero que posteriormente desencadenan daños acumulados en las articulaciones, huesos, tendones y ligamentos que provocan así el final de la vida competitiva (50).

Por el contrario, las variables en nuestro estudio, que han resultado disminuir el riesgo de terminar la vida competitiva han sido Ncat1, Ncat2 y Ncat3 (Tabla 9). Sin embargo, la bibliografía muestra resultados contrarios, (51) siendo esto lógico ya que, como se ha explicado anteriormente, una mayor frecuencia de carreras puede provocar un incremento en las lesiones y por tanto en última instancia en menos años en competición. De hecho, Di Battista et al. (48); demostraron que, es importante para una larga vida competitiva del caballo comenzar por categorías inferiores ya que el riesgo de terminarla aumenta conforme lo hacen también el nivel y la frecuencia en las categorías. Los resultados que se han obtenido han podido ser debidos a una utilización de la variable categoría como un absoluto en la vida del caballo y no como tiempo dependiente como otros estudios (50). Es decir, se ha relacionado que, caballos con mejores rendimientos se inscriban más a menudo en carreras de Raid y que además se les tenga compitiendo más años por las mismas razones.

En la disciplina de salto, se observó que, los machos y las hembras competían menos tiempo que los castrados (21). No obstante, en nuestro estudio encontramos que esta

variable no ha afectado a la vida competitiva de los equinos (Tabla 9). Similares resultados obtuvieron otros estudios (52) que demostraron que, además de que el sexo no tiene ningún efecto en los años en competición en Raid, tampoco en el número total de carreras competidas, por esta razón no ha sido incluido en muchos otros estudios (48).

En general, el número de años en competición se ha utilizado sobre todo como una medida de la longevidad funcional del caballo (29). Los resultados que se obtuvieron para la heredabilidad de la longevidad fueron altos (0.45) en comparación con otros estudios de longevidad que muestran resultados de 0.10 a 0.20. Cabe destacar que, los estudios sobre la longevidad en caballos de raid son limitados; no obstante, si se tienen resultados en otras disciplinas como el salto (25) donde se estimó una heredabilidad de 0.18 y la doma (29) con una heredabilidad de 0.20. Además, Ricard y Blouin (53) llegaron a estimar un valor de la heredabilidad de 0.08 en datos similares. Estos valores tan dispares se pueden explicar gracias al porcentaje de datos censurados que se tienen. Un estudio en 2006 (22) calculó la heredabilidad de la longevidad bajo datos simulados con un 10, un 20 y un 25 % de datos censurados obteniendo resultados desde 0.11, a 0.082. Si comparamos los estudios citados anteriormente, (25) reportaba un 32.7% de datos censurados, (29) reportaban un 18 % y nuestro estudio solo un 7.47 %. Esto demuestra la necesidad de reducir el sesgo en la estimación de la heredabilidad de la longevidad para obtener resultados certeros que posteriormente sean utilizados para la selección.

Otra posible razón de una heredabilidad tan alta reside en que, habitualmente, los análisis genéticos de longevidad se hacen para una sola raza mientras que en el nuestro encontramos hasta seis distintas (17). Además, no hemos introducido en el modelo otros factores importantes como la experiencia del jinete, la interacción de este con el caballo o el ambiente en el que se crían que sí han resultado influir en la heredabilidad en otras disciplinas como la doma (29). Todo esto, junto con los bajos datos censurados puede provocar que se obtengan valores de la heredabilidad tan altos como los obtenidos.

Los resultados de la Gráfica 5 y de la Tabla 10 se explican gracias a la descendencia de los mejores y los peores equinos que sigue compitiendo. Un estudio (19) demostró que el 25 % de la descendencia de los mejores sementales y solo el 16 % de la descendencia de los peores sigue compitiendo después de 6 años en carreras de resistencia. Además, se demostró en doma (29) que, después de 16 pruebas no queda descendencia del peor grupo mientras que el 60 % de la descendencia de los mejores sementales seguía compitiendo después del mismo periodo. En nuestro caso, los valores de los mejores sementales y de



los peores se encuentran alrededor del año 2000 (Tabla 10). Esto quiere decir, que hasta alrededor de 2006 la descendencia de ambos seguirá compitiendo. Así pues, en la Gráfica 11, podemos ver que, la fragilidad genética (valores genéticos) baja alrededor del año 2004 que podría explicarse, por una parte, por el efecto de la descendencia de los mejores sementales (menos frágiles), pero por otra, también podría deberse a un efecto del aumento del número de pruebas por año a lo largo del tiempo que podría indicar esa menor fragilidad (29). No obstante, a partir de 2006 es cuando comienza a crecer, valores que pueden estar influidos por la presencia también de la descendencia de los peores sementales que hacen más altos estos valores de fragilidad, pero alejados todavía de valores positivos. Por lo tanto, todo esto nos indica que, ha habido una selección para la longevidad o al menos indirectamente (29)

Los resultados de precisión obtenidos (98% con precisión de más de 0.60) tampoco coinciden con los de la literatura citada anteriormente donde se obtienen precisiones de más de 0.60 en solo 79 de los 10.394 caballos (29), o 47 de los 8475 (25). Esto, por un lado, puede deberse al tamaño del conjunto de datos, al intervalo de tiempo cubierto, a la disciplina analizada, que es diferente para todos los estudios, y a la gran heredabilidad obtenida que está sobreestimada por las razones mencionadas anteriormente (25). Por otro lado, este tipo de porcentajes de precisión normalmente se dan cuando ya existen relaciones estrechas entre los animales; es decir, que ya se ha seleccionado entre ellos previamente aumentando así la fiabilidad de las estimaciones (29). Por tanto, seleccionar para longevidad puede ser beneficioso en caballos de Raid incrementando su bienestar animal y haciendo rentable la crianza de estos animales.

## **8.CONCLUSIONES**

Aquí se detallan las conclusiones que se han ido obteniendo a lo largo de la aplicación del caso práctico:

1. Un caballo que se elimina en su primera carrera tiene más riesgo de acabar tempranamente su vida deportiva.
2. Aumentar la edad a la que el equino realiza la primera carrera a aumenta el riesgo de dejar de competir.
3. Una mayor frecuencia de carreras en categorías 1, 2 y 3 en este estudio, resultan comportarse como factores protectores incrementando la vida deportiva del animal
4. El valor de heredabilidad obtenida en este análisis se consideró alta (0.43). Esto nos indica que, podemos esperar una respuesta a la selección si se selecciona por esta variable.
5. La tendencia genética encontrada para la longevidad sugiere que, se selecciona para ésta indirectamente seguramente debido a que caballos con mejores rendimientos son más longevos y tienen, por tanto, más oportunidades de dejar descendientes.

## **9.AGRADECIMIENTOS**

Este trabajo ha sido realizado al amparo del Programa de Mejora del Caballo de Pura Raza Árabe desarrollado por la Asociación Española de Criadores de Caballos Árabes (AECCA). Agradecemos la disposición de la Real Federación Hípica Española para ceder los datos de las competiciones hípicas analizadas en este TFM.

## 10. REFERENCIAS BIBLIOGRÁFICAS

- (1) RFHE: Real Federación Hípica Española. 2020. Raid. Disponible en: <http://www.rfhe.com/raid/> Consultado el 01/09/2020
- (2) RFHE: Real Federación Hípica Española. 2020. REGLAMENTO DE RAID. Disponible en: <http://www.rfhe.com/wp-content/uploads/2020/02/Reglamento-de-Raid-hasta-publicacion-del-2020.pdf> Consultado el 01/09/2020
- (3) FEI: Federation Equestre Internationale. Endurance Rules. Disponible en: [https://inside.fei.org/system/files/20.1\\_ANNEX%20GA19\\_ENDURANCE%20RULES%20CLEAN.pdf](https://inside.fei.org/system/files/20.1_ANNEX%20GA19_ENDURANCE%20RULES%20CLEAN.pdf) Consultado: 01/09/2020.
- (4) The European Horse Network. 2017. Horses & Sport. Disponible en: <http://www.europeanhorsenetwork.eu/> Consultado el 01/09/2020.
- (5) Turner, Jacky. Animal Breeding, Welfare and Society, Routledge, 2010. Disponible en: <http://ebookcentral.proquest.com/lib/universidadcomplutenseebooks/detail.action?docID=554810>. Consultado: 02/09/2020
- (6) BOE-A-1979-228. «BOE» núm. 5, 1979, páginas 204 a 208 (5 págs.) Disponible en: <https://www.boe.es/eli/es/o/1978/12/26/> . Consultado 02/09/2020
- (7) BOE-A-1997-24860 «BOE» núm. 279, de 21 de noviembre de 1997, páginas 34205 a 34207 (3 págs.) Disponible en: <https://www.boe.es/eli/es/rd/1997/11/07/1682> . Consultado 02/09/2020
- (8) BOE-A-2019-2859 «BOE» núm. 52, 2019, páginas 19716 a 19748 (33 págs.). Disponible en: <https://www.boe.es/eli/es/rd/2019/02/08/45> . Consulado 02/09/2020
- (9) MAPA: Ministerio de Agricultura Pesca y Alimentación. 2020. Disponible en: <https://www.mapa.gob.es/es/ganaderia/temas/zootecnia/razas-ganaderas/programas-mejora/default.aspx> Consultado 02/09/2020
- (10) E.P.C Koenen, L.I Aldridge, J Philipsson,.An overview of breeding objectives for warmblood sport horses. Livestock Production Science.2004. Volumen 88:77-84
- (11) Navarro, I. C. Estructura genética del caballo de pura raza árabe español y su influencia en razas derivadas: aplicación de nuevas metodologías en el cálculo del tamaño efectivo.2008. (Doctoral dissertation, Universidad Complutense de Madrid).

- (12) BOE-A-2020-2445. «BOE» núm. 43, de 19 de febrero de 2020, páginas 15915 a 15915 (1 pág.). Disponible en: [https://www.mapa.gob.es/es/ganaderia/temas/zootecnia/razasganaderas/razas/catalogo/integrada-espana/equino-caballar/arabe/datos\\_reglamentacion.aspx](https://www.mapa.gob.es/es/ganaderia/temas/zootecnia/razasganaderas/razas/catalogo/integrada-espana/equino-caballar/arabe/datos_reglamentacion.aspx) Consultado: 01/09/2020
- (13) AECCA: Asociación Española de Criadores de Caballos Árabes. Programa de mejora del caballo de pura raza árabe. Disponible: <https://docplayer.es/7817211-Programa-de-mejora-del-caballo-de-pura-raza-arabe.html> Consultado: 01/09/2020
- (14) AECCA: Asociación Española de Criadores de Caballos Árabes. Catálogo de reproductores 2019. Disponible: <https://www.aecca.es/catalogo-reproductores> Consultado: 01/09/2020
- (15) Árnason, T., & Van Vleck, L. D. (2000). Genetic improvement of the horse. The genetics of the horse, 411-438.
- (16) Gutiérrez J. Iniciación a la valoración genética animal. Metodología adaptada al EEES. 1º ed. Editorial Complutense, S.A.2010;
- (17) Å. Braam a,\*, A. Näsholm a, L. Roepstorff b, J. Philipsson a Genetic variation in durability of Swedish Warmblood horses using competition results. Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences. 2011.SE-750 07 Uppsala, Sweden.
- (18) Wallin L., Strandberg E., Philipsson J. Phenotypic relationship between test results of Swedish warmblood horses as 4-year-olds and longevity. 2001.Livest. Prod. Sci., 68, 97–105
- (19) Arnason, T., 2006. Survival analysis of the length of competition life of Standardbred trotters in Sweden. In: Van Der Honing, Y. (Ed.), Book of Abstracts of the 57th Annual Meeting of the European Association for Animal Production. Wageningen Academic Publishers, Antalya, Turkey, pp. 350
- (20) Ricard, A., Fournet-Hanocq, F.. Analysis of factors affecting length of competitive life of jumping horses. Genetics Selection Evolution 1997 29,251–267

(21) Anne Ricard, Christine Blouin. Genetic analysis of the longevity of French sport horses in jumping competition. *Journal of Animal Science*, American Society of Animal Science, 2011, 89 (10), pp.2988-2994.

(22) Burns E, Enns R, Garrick D. The effect of simulated censored data on estimates of heritability of longevity in the Thoroughbred racing industry. 2006. *Genetics and Molecular Research*. 5 (1): 7-15.

(23) Sabeva, I. Apostolov, A. Longevity of shagya broodmares. *Bulgarian journal of agricultural science*. 2012. 18: 272-277

(24) Jonsson, L Nasholm, A Roepstorff, L Egenvall, A Dalin, G Philipsson, J, Jonsson, Lina, Nasholm, Anna, Roepstorff, Lars, Egenvall, Agneta, Dalin, Goran Philipsson, Jan. Genetic analysis of clinical findings at health examinations of young Swedish warmblood riding horses. 2013. *ACTA VETERINARIA SCANDINAVICA*. 55

(25) Posta J, Rudin A, Mezei, Mihók S, Mészáros G. Evaluation of the length of competitive life in Hungarian sport horses. *Journal of Animal Breeding and Genetics*.; 2014. 131:529-535.

(26) Wallin L, Strandberg E, Philipsson J. Genetic correlations between field test results of Swedish Warmblood Riding Horses as 4-year-olds and lifetime performance results in dressage and show jumping. *Livestock Science*; 2016. 184:126-133.

(27) Seiero, T., Mark, T, Jonsson, L. Genetic parameters for longevity and informative value of early indicator traits in Danish show jumping horses. 2016. Volumen: 184 p: 126-133

(28) Velie, B. D, Hamilton, N. A, Wade, C. M. Heritability of racing durability traits in the Australian and Hong Kong Thoroughbred racing populations

(29) M. Solé, M.J Sánchez, M. Valera, A. Molina, P.J. Azor, J. Sölkner, G. Mészáros. Assessment of sportive longevity in Pura Raza Español dressage horses. 2017.

(30) Everett, R. W., J. F. Keown, and E. E. Clapp: Production and stayability trends in dairy cattle. *J. Dairy Sci.* 1976. 59:1532-1539.

(31) Ducrocq, V. Survival analysis, a statistical tool for longevity data. 1997. 48 Annual Meeting of the EAAP, Vienna, Austria

- (32) Natascha Vukasinovic. Application of survival analysis in breeding for longevity. Animal Breeding Group, Swiss Federal Institute of Technology.1999
- (33) Collett, D. Modelling survival data in medical research.2015. CRC press
- (34) Smith, S.P., Quaas, R.L. Productive lifespan of bull progeny groups: failure time analysis. J. Dairy. 1984. Sci 67:2999-307.
- (35) Ducrocq, V. Statistical análisis of length of productive life for dairy cows of the Normande breed J.Dairy Sci 1994. 77: 855-866.
- (36) Meszaros,G., Sölkner,J., Ducrocq,V. The Survival Kit: Software to analyze survival data including possibly correlated random effects. Computer Methods and Programs in Biomedicine. 2013. 110(3): 503-510.
- (37) Ducrocq, V., Casella, G. A Bayesian analysis of mixed survival models. Genetics Selection Evolution.1996. 28: 505-529.
- (38) Clayton, D.GA model for association in bivariate life tables and its application inepidemiological studies of familial tendency in chronic disease incidence. Biometrika . 1978. 65, 141-151
- (39) Hanagal, D. Modeling survival data using frailty models. Chapman & Hall. 2011
- (40) Ducrocq, V., Sölkner, J: Implementation of a routine breeding value evaluation for longevity of dairy cows using survival analysis techniques. Proceedings of the 6th World Congress on Genetics Applied to Livestock Production, Armidale, Australia, Vol. 23: 1988.359-362
- (41) J.P.Sánchez. Modelos semiparamétricos en el análisis de supervivencia aplicado a la mejora genética. Dep. de Ciencia Animal, UPV. 2005
- (42) Cox, D. R., Oakes, D. (1984): Analysis of survival data. Chapman and Hall, London, England
- (43) Ripatti, S., Larsen, K. & Palmgren, J. Maximum Likelihood Inference for Multivariate Frailty Models Using an Automated Monte Carlo EM Algorithm. Lifetime Data Anal. 2002., 349–360

- (44) E. Albertsdóttir S. Eriksson Á. Sigurdsson T. Árnason Genetic analysis of ‘breeding field test status’ in Icelandic horses. 2011. Disponible en: <https://doi.org/10.1111/j.1439-0388.2010.00902.x> Consultado 02/09/2020
- (45) Yazdi, M.H., Visscher, P.M., Ducrocq, V., Thompson, R., Heritability, reliability of genetic evaluations and response to selection in proportional hazard models. *J. Dairy Sci.* 85. 2002, 1563–1577
- (46) D.C. Liu, J. Nocedal On the limited memory BFGS method for large scale optimization *Mathematical Programming*, 45 .1989, pp. 503-528
- (47) Nagy, J. K. Murray and S. J. Dyson. Descriptive epidemiology and risk factors for eliminations from Fédération Equestre Internationale endurance rides due to lameness and metabolic reasons (2008–2011) *Equine Veterinary Journal*.2014. ISSN 0425-1644DOI: 10.1111/evj.12069
- (48) Carlo Di Battista, Marcello Conte, Marco Pepe, Lucio Petrizzi, Francesca Beccati, Epidemiology and risk factors for eliminations from Fédération Equestre Internationale endurance rides between 2004–2015 in Italy, *Preventive Veterinary Medicine*, Volume 170.2019.104737, ISSN 0167-5877,
- (49) Foreman, J.H. The exhausted horse syndrome. *Vet. Clin. N. Am.: Equine Pract.* 1998.14, 205- 219.
- (50) Lannoy, A.B. de Risk factors and elimination rates due to lameness in New Zealand endurance races. 2015. Faculty of Veterinary Medicine Theses.
- (51) Muñoz, A., Riber C., Trigo P., Castejon, Riber, C. and Castejón, F.M. Dehydration, electrolyte imbalances and renin-angiotensin-aldosterone-vasopressin axis in successful and unsuccessful endurance horses. *Equine Veterinary Journal*, 2010. 42: 83-90.
- (52) Árpád Bokor\*, Helga Lukacs, Julianna Bokor, István Nagy and Virág acs. 917 Examining the racing performance and longevity in the Hungarian Thoroughbred population. *Journal of Central European Agriculture*, 2018, 19(4)
- (53) Ricard A., Blouin C. Genetic analysis of the longevity of French sport horses in jumping competition. *J. Anim. 2011.Sci.*, 89, 2988–2994



## 11. APÉNDICES

### APÉNDICE 1

Tal y como se puede observar en el apartado 3 de la descripción de la base de datos, esta no tenía la columna de censurados por lo que se procede a crearla considerando aquellos caballos que corrieron en 2018 como censurados por la derecha (Box1).

```
datos1$cens<-ifelse(datos1$añocar==2018,0,1)
```

Box 1: Sintaxis Definición variable censura (cens)

Con objetivo de generar más tarde las variables principales del estudio comentadas anteriormente, se crea una variable llamada intervalo que va contando el número de registros que tiene cada caballo y posteriormente se selecciona las covariables a utilizar.

```
datos2<-datos1[order(datos1$idnum,datos1$añocar),]
t<-table(datos2$idnum)
freq<-as.data.frame(t)
Freq<-freq$Freq
interval<-c()
for(i in 1:nrow(freq)){ interval<-
c(interval,seq(Freq[i]))}
datos2$interval<-interval
```

Box 2: Sintaxis Definición de la variable intervalo (interval)

Gracias a esta variable, ya sabemos cuál es el primer registro de cada animal y a partir de aquí creamos las variables como: año de nacimiento (Box3) de cada caballo (para crear edad posteriormente), año de cambio de sexo (Box4)...

```
nacimiento<-vector(mode="numeric", length=nrow(datos2))
for(i in 1:nrow(datos2)){nacimiento[i]<-datos2$añocar[i]-
datos2$edad[i]}
datos2$nacimiento<-nacimiento
```

Box 3: Sintaxis Definición de la variable nacimiento (nacimiento)

```
yearcambioL<-vector(mode="numeric", length=nrow(datos2))
for(i in 2:nrow(datos2)){yearcambioL[i]<-
ifelse(datos2$interval[i]>1 &
datos2$sexoL[i]!=datos2$sexoL[i-1],datos2$añocar[i],0) }
yearcambioL[1]<-0
datos2$yearcambioL<-yearcambioL
```

Box 4: Sintaxis Definición de la variable año de cambio de sexo (yearcambioL)

Tal y como se ha venido indicando en anteriores apartados, se va a utilizar la variable sexo como tiempo dependiente ya que muchos equinos pasan a ser castrados (sexo=3) a lo largo de su tiempo de vida. Para ello necesitamos que el fichero pase de ser largo a ancho (Box5);

```
temp<-datosTD[!duplicated(datosTD$idnum),]
```

Box 5: Sintaxis creación fichero de una observación por caballo

Con este comando conseguimos un fichero ancho con la primera observación por caballo que más tarde será importante para covariables como el Placing que se introducirá en el modelo sin más procesamiento. No obstante, otras covariables como las tiempo-dependientes precisan de más cambios.

A partir de aquí, se explica el procesamiento de la variable sexo como tiempo dependiente. La covariable viene descrita por 4 columnas; número de cambios totales de cada caballo (Box7), variable columna que cambia (Box 8), tiempo de cambio (Box 10) y por último el nuevo valor (Box 11):

```
cambios<-vector(mode="numeric", length=nrow(datosTD))
for(i in 1:nrow(datosTD)){cambios[i]<-
ifelse(datosTD$yearcambioL[i]>0,cambios[i]+1,0)}
datosTD$cambios<-cambios
```

Box 6: Sintaxis definición de la variable cambios (cambios)

```
numcambios<-aggregate(datosTD$cambios,
by=list(Category=datosTD$idnum), FUN=sum, drop=TRUE)
datos_ancho$numcambios<-numcambios$x
```

Box 7: Sintaxis definición de la variable número de cambios (cambios)

```
datos_ancho$colcambia<ifelse(datos_ancho$numcambios>0,4,A)
```

Box 8: Sintaxis definición de la variable columna que cambia (colcambia)

También se crea la variable principal años en competición llamada life2 (Box9)

```
empiece<-aggregate(datosTD$añocar,
by=list(Category=datosTD$idnum), FUN=min, drop=TRUE)
datos_ancho$empiece<-empiece$x
final<-aggregate(datosTD$añocar,
by=list(Category=datosTD$idnum), FUN=max)
datos_ancho$final<-final$x+1
datos_ancho$life2<-datos_ancho$final-datos_ancho$empiece
```

Box 9: Sintaxis definición de la variable principal años en competición (life2)

```
yearcambioA<-aggregate(datosTD$yearcambioL,
by=list(Category=datosTD$idnum), FUN=max, drop=TRUE)
```

```
datos_ancho$timecambio2<-
ifelse(datos_ancho$numcambios>0,yearcambioA$x-
datos_ancho$empiece,NA)
```

Box 10: Sintaxis definición de la variable tiempo de cambio en años en competicion (timecambio2)

```
newsexo<-datosTD[!duplicated(datosTD$idnum, fromLast=T),]
datos_ancho$newsexo<ifelse(datos_ancho$numcambios>0,newse
xo$sexoL,NA)
```

Box 11: Sintaxis definición de la variable nuevo valor de sexo (newsexo)

Además, se crean variables como la edad al cuadrado (Box 12) y ncat1 (Box 13), ncat2 (Box 14), ncat3 (Box 15) y ncat 4 (Box 16)

```
datos_ancho$edpored<-datos_ancho$edad*datos_ancho$edad
```

Box 12: Sintaxis definición de la variable edad cuadrática (edpored)

```
cat1<-vector(mode="numeric", length=nrow(datosTD))
for(i in 1:nrow(datosTD)){ cat1[i]<-
ifelse(datosTD$ncategoria[i]==1 ,cat1[i]+1,0)}
datosTD$cat1<-cat1
ncat1<-aggregate(datosTD$cat1,
by=list(Category=datosTD$idnum), FUN=sum, drop=TRUE)
datos_ancho$ncat1<-ncat1$x
```

Box 13: Sintaxis definición de la variable categoria1 (ncat1)

```
cat2<-vector(mode="numeric", length=nrow(datosTD))
for(i in 1:nrow(datosTD)){cat2[i]<-
ifelse(datosTD$ncategoria[i]==2,cat2[i]+1,0)}
datosTD$cat2<-cat2
ncat2<aggregate(datosTD$cat2,by=list(Category=datosTD$idn
um), FUN=sum, drop=TRUE)
datos_ancho$ncat2<-ncat2$x
```

Box 14: Sintaxis definición de la variable categoria2 (ncat2)

```
cat3<-vector(mode="numeric", length=nrow(datosTD))
for(i in 1:nrow(datosTD)){ cat3[i]<-
ifelse(datosTD$ncategoria[i]==3 ,cat3[i]+1,0) }
datosTD$cat3<-cat3
ncat3<-aggregate(datosTD$cat3,
by=list(Category=datosTD$idnum), FUN=sum, drop=TRUE)
datos_ancho$ncat3<-ncat3$x
```

Box 15: Sintaxis definición de la variable categoria3 (ncat3)

```

cat4<-vector(mode="numeric", length=nrow(datosTD))
for(i in 1:nrow(datosTD)){
cat4[i]<- ifelse(datosTD$ncategoria[i]==4 ,cat4[i]+1,0) }
datosTD$cat4<-cat4
##VARIABLE que suma las ncat
ncat4<-aggregate(datosTD$cat4,
by=list(Category=datosTD$idnum), FUN=sum, drop=TRUE)
datos_ancho$ncat4<-ncat4$x

```

Box 16: Sintaxis definición de la variable categoria4 (ncat4)

En el Box 17 se encuentra la sintaxis del uso de la función SKit4R

```

SKit4R(inputData="datosdef1.dat",inputPedigree="familia.p
ed",effects=c("idnum","life","cens","sexoL","Placing","nc
at1","ncat2","ncat3","ncat4","edad","edpored"),
effectType=c("class","class","class","class","class","con
tinuous","continuous","continuous","continuous","continuo
us","continuous"),classeffects=c("idnum","sexoL","Placing
"),random=("idnum ESTIMATE MULTINORMAL USUAL_RULES 0.01
0.90 0.01"),pedigreeEffect = c("idnum"),timeDepEffects =
c("sexoL"),timeDepEffectsType =
c("class"),program="weibull",title="MODELOFINAL",model=c(
"idnum","sexoL","Placing","ncat1","ncat2","ncat3","ncat4"
,"edad","edpored"),ite_quasi = "10000",nze_hessian =
"500000",nrecmax = "50000",ndimax = "50000",test =
"SEQUENTIAL LAST)

```

Box 17: Sintaxis de la función Skit4R